KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

Discussion Paper No.919

“JOINT SPECIFICATION TESTS FOR RESPONSE PROBABILITIES IN UNORDERED MULTINOMIAL CHOICE MODELS”

Masamune Iwasawa

March 2015

KYOTO UNIVERSITY

KYOTO, JAPAN

JOINT SPECIFICATION TESTS FOR RESPONSE PROBABILITIES IN UNORDERED MULTINOMIAL CHOICE MODELS*

Masamune Iwasawa †

Graduate School of Economics, Kyoto University

Research Fellow of Japan Society for the Promotion of Science

Abstract

Estimation results obtained by parametric models may be seriously misleading when the model is misspecified or poorly approximates the true model. This study proposes two tests that jointly test the specifications of multiple response probabilities in unordered multinomial choice models. Both test statistics are asymptotically chi-square distributed, consistent against a fixed alternative, and able to detect a local alternative approaching to the null at a rate slower than the parametric rate. We show that rejection regions can be calculated by a simple parametric bootstrap procedure, when the sample size is small. The size and power of the tests are investigated by Monte Carlo experiments.

*The author is grateful to Yoshihiko Nishiyama, Ryo Okui, and Naoya Sueishi for their comments, guidance, and encouragement. I also would like to thank Kohtaro Hitomi and Yoon-Jae Whang for helpful comments. This work was supported by JSPS KAKENHI Grant Number 13J06130.

†Yoshida Honmachi, Sakyo-ku, Kyoto, Japan, 606-8501; Email: iwasawa.masamune.88m@st.kyoto-u.ac.jp

1 INTRODUCTION

Not infrequently, variables of interest in economic research are discrete and unordered as we often find the variables that indicate behavior or state of economic agents. Some econometric models have been developed to deal with these discrete and unordered outcomes. Above all, parametric models such as the multinomial logit (MNL) and probit (MNP) models proposed by McFadden (1974) and Hausman and Wise (1978), respectively, are widely employed, for example, in structural econometric analysis (e.g., the economic models of automobile sales in Berry, Levinsohn, & Pakes, 1995; Goldberg, 1995) and as part of econometric methods (e.g., selection bias correction of Heckman, 1976; Dubin & McFadden, 1984). However, results obtained by such parametric models may be seriously misleading when the model is misspecified or poorly approximates the true model. Thus, researchers need to examine the validity of parametric assumptions as long as the assumptions are refutable from data alone.

This study proposes two new specification tests that are directly applicable to any multinomial choice models with unordered outcome variables. These models set parametric assumptions on response probabilities that an option is chosen from multiple alternatives, and identical assumptions are often set for all response probabilities. Problems occur when these models do not mimic the true models because the response probabilities and partial effects of some variables on the probabilities cannot be properly predicted. Moreover, the parameter estimation results may be misleading and their interpretation confusing. The specification tests proposed here can be utilized to justify the choice of parametric models and avoid misspecification problems.

The novelty of the tests provided in this study is that they allow us to test the specifications of response probabilities jointly for all choice alternatives. Multinomial choice models with unordered outcomes consist of multiple response probabilities, each of which may be parameterized differently. This implies that one needs to test multiple null hypotheses to justify the parametric assumptions of these models. A substantial number of specification tests have been developed to test a single null hypothesis. To our knowledge, however, no joint specification tests have so far been theoretically suggested for multiple null hypotheses.

One test proposed here is based on the L_2-distances between parametric and nonparametric fits of response probabilities, and the other is based on moment conditions. We show that both test statistics are asymptotically chi-square distributed, consistent against a fixed alternative, and able to detect a local alternative approaching the null at the rate $1/\sqrt{nh^{q/2}}$, where q is the number of independent variables.

One eminent feature of our tests is that a parametric bootstrap procedure works well to calculate the rejection region for the test statistic. Since both testing methods involve nonparametric estimation, a sufficiently large sample size could be required to establish that the chi-squared distribution is a proper approximation for the distribution of test statistics. Thus, a simple parametric bootstrap procedure to calculate rejection regions is a practical need.

A crucial point that makes parametric bootstrap work is that the orthogonality condition holds with bootstrap sampling under both the null and alternative hypotheses. This is different from the specification test for the regression function that requires the wild bootstrapping procedure to calculate the rejection region proved by Härdle and Mammen (1993). It is also noteworthy that the parametric nature of the model leads to substantial savings in the computational cost of bootstrapping.

Methodologically, two different approaches have been developed to construct specification tests. One uses an empirical process, and the other a smoothing technique. We call the first type empirical process-based tests and the second type smoothing-based tests. Most of the literature on specification tests can be categorized into one of these two types. Empirical process-based tests are proposed by Bierens (1982), Bierens (1984), Bierens (1990), Delgado (1993), De Jong (1996), Andrews (1997), Bierens and Ploberger (1997), Stute (1997), Stinchcombe and White (1998), Chen and Fan (1999), and Whang (2000), among others. Smoothing-based tests are proposed by Eubank and Spiegelman (1990), Le Cessie and van Houwelingen (1991), Wooldridge (1992), Yatchew (1992), Gozalo (1993), Härdle and Mammen (1993), Aït-Sahalia, Bickel, and Stoker (1994), Delgado and Stengos (1994), Horowitz and Härdle (1994), Hong and

White (1995), Y. Fan and Li (1996), Lavergne and Vuong (1996), Zheng (1996), Li and Wang (1998), and Lavergne and Vuong (2000) to mention only a few.

These two types of tests are complementary to each other, rather than substitutional, in terms of the power property. For Pitman local alternatives, empirical process-based tests are more powerful than the smoothing-based tests. The empirical process-based tests can detect Pitman local alternatives approaching the null at the parametric rate $n^{-1/2}$, whereas the smoothing-based tests can detect them at a rate slower than the parametric rate. Smoothing-based tests are, however, more powerful for a singular local alternative that changes drastically or is of high frequency. Empirical process-based tests can be represented by a kernel-like weight function with a fixed smoothing parameter. Thus, it can be intuitively understood that empirical process-based tests oversmooth the true function and obscure drastic changes of alternatives. Y. Fan and Li (1996) show that smoothing-based tests can detect singular local alternatives at a rate faster than $n^{-1/2}$.

The two tests proposed in this study are most related to Härdle and Mammen (1993) and Zheng (1996), both of which propose smoothing-based tests for functional forms of the regression function. Most of the specification tests developed for functional forms of the regression function can be directly applied to test the parametric specifications of ordered choice models such as the parametric binary choice models because ordered choice models have only single response probability that is equal to conditional expectation of outcome. For example, Mora and Moro-Egido (2008) applied several specification tests, originally developed for regression functions, to some ordered discrete choice models for a comparison of their relative merits based on their asymptotic sizes and powers. However, extending their application to unordered multinomial choice models is not trivial, which is carried out in this paper. Extending empirical process-based tests and rate-optimal tests[1] to unordered multinomial choice models is a task left for future research.

This paper is organized as follows. Section 2 introduces unordered multinomial choice models and reveals problems of parametric specification. The two new test statistics are proposed in Section 3. The assumptions and asymptotic behaviors are provided in Section 4. Section 5 shows how to bootstrap parametrically. We investigate the size and power of the tests by conducting Monte Carlo experiments in Section 6. We conclude with Section 7. The proofs of the lemmas and propositions are provided in the appendix.

2 UNORDERED MULTINOMIAL CHOICE MODELS

We have the observations $\{\{Y_{i,j}, X_{i,j}\}_{i=1}^n\}_{j=1}^J$, where $Y_{i,j} \in \{0, 1\}$ is a binary response variable that takes one if individual i chooses alternative j and zero otherwise. Each individual chooses one of J alternatives, which implies $Y_{i,m} = 0$ for all $m \neq j$ if $Y_{i,j} = 1$. $X_{i,j} \in \mathbb{R}^{k_j}$ is a vector of independent variables that affect the choice decision made by individual i. Throughout this paper, we assume that $\{X_{i,j}, Y_{i,j}\}_{i=1}^n$ is independent and identically distributed for each $j = 1, \ldots, J$. With i remaining fixed, however, $\{X_{i,j}, Y_{i,j}\}_{j=1}^J$ is not necessarily independent or identical.

Multinomial choice models with unordered response variables is constructed by introducing latent variables $y_{i,j}^*$, which may be interpreted as the utility or satisfaction that i can obtain by choosing alternative j. We assume each individual chooses an alternative that maximizes personal utility; that is, $Y_{i,j} = 1$ if $y_{i,j}^* > y_{i,m}^*$ for all $m \neq j$. Further, $y_{i,j}^*$ depends on a function $g_j(X_{i,j}, \theta)$ and unobserved error $\epsilon_{i,j}$: $y_{i,j}^* = g_j(X_{i,j}, \theta) + \epsilon_{i,j}$, where $\epsilon_{i,j}$ is independent of $X_{i,j}$ and $\theta \in \Theta$ is the parameter in a subset of a finite dimensional space Θ. Then, the response probability that i chooses j can be formulated as follows:

$$\begin{aligned} P(Y_{i,j} = 1|X_i) &= P(y_{i,j}^* > y_{i,m}^* \ \forall m \neq j \ |X_i) \\ &= P(\epsilon_{i,j} - \epsilon_{i,m} > g_m(X_{i,m}, \theta) - g_j(X_{i,j}, \theta) \ \forall m \neq j \ |X_i), \end{aligned} \tag{1}$$

where $X_i \in \mathbb{R}^q$ is a vector consisting of all independent variables. The dimension q of X_i is equal to $\sum_{j=1}^J k_j$ when all variables in $X_{i,j}$ are alternative-specific for all j. This occurs when no variable in $X_{i,j}$ is identical to any of those in $X_{i,m}$ as long as $j \neq m$.

A specification of the functional forms of $g(\cdot)$ and distributions of ϵ lead to full parameterization of the model in the sense that parameters and response probabilities can be estimated parametrically. For example, if we assume linearity, $g_j(X_{i,j}, \theta) = X'_{i,j}\beta$, and the type I extreme-value distribution for $\epsilon_{i,j}$ for all j, we have McFadden's (1974) MNL model in which $P(Y_{i,j} = 1|X_i) = \exp(X'_{i,j}\beta)/\sum_{j=1}^{J}\exp(X'_{i,j}\beta)$.[2] An alternative model suggested by Hausman and Wise (1978) is the MNP model, in which $\epsilon_{i,j}$ is assumed to be normally distributed. In both cases, the parameters can be inferred by maximum likelihood estimation, and the choice probabilities are obtained by plugging the estimated values into (1).

In empirical studies, however, functional forms of $g_j(\cdot)$ and distributions of $\epsilon_{i,j}$ are generally unknown for all j. Moreover, functional forms of $g_j(\cdot)$ and distributions of $\epsilon_{i,j}$ in unordered multinomial choice models may be nonidentical across j. Thus, we need joint specification tests that indicate whether parametric specifications provide a good approximation to the true models. The appropriate null and alternative hypotheses are as follows:

$$H_0 : P[m_{\theta,j}(X_i) = P(Y_{i,j} = 1|X_i)] = 1, \text{for some } \theta \in \Theta \text{ and for all } j$$

$$H_1 : P[m_{\theta,j}(X_i) = P(Y_{i,j} = 1|X_i)] < 1, \text{for any } \theta \in \Theta \text{ and for some } j,$$

where $m_j(X_i)$ denotes the true response probabilities and $m_{\theta,j}(X_i)$ their parameterized variants.

3 TEST STATISTICS

Both test statistics proposed in this study are built on the features of response probabilities. One uses the L_2-distance between the parametric and nonparametric fits of response probabilities, and the other uses moment conditions that are satisfied when the parametric response probability is true. This implies that we test the specifications of the functional forms of $g_j(\cdot)$ and the distributions of $\epsilon_{i.j}$ simultaneously for all j.

One may think that the rejection of the null hypothesis reveals nothing about what is misspecified because the tests reject the null hypothesis if any combination of the functional forms of $g_j(\cdot)$ and the distributions of $\epsilon_{i,m}$ for all j and m is misspecified. Nonetheless, rejection of the null hypothesis could imply something more. Note that strict inequality holds after any transformations on both sides of inequalities with any strictly increasing functions. This implies that the distribution of the conditional response probabilities given in equation (1) could be transformed into what is well-known as the normal or type-I extreme distribution. In this case, the distributions of $\epsilon_{i.j}$ are not an essential specification issue, provided we can specify the functional forms of $g_j(\cdot)$ correctly. In other words, distributional assumptions of error terms could help us simplify the estimation of parametric models by specifying the functional forms of $g_j(\cdot)$ prudently.

Before presenting the test statistics, we introduce some notations. Let $f_h(x)$ be the nonparametric density estimator for a continuous point of X_i and $m_{h,j}(x)$ the Nadaraya–Watson kernel estimator for $m_j(x) \equiv P(Y_{i,j} = 1|X_i = x) = \mathbb{E}(Y_{i,j}|X_i = x)$, as follows:

$$f_h(x) = \frac{1}{nh^q} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)$$
$$m_{h,j}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) Y_{i,j}}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)},$$

where $K(\cdot)$ is a kernel function and h is a bandwidth depending on n. In addition, we define $K^{(2)}$ as the two-times convolution product of the kernel and $K^{(4)}$ as the two-times convolution product of $K^{(2)}$.

3.1 Test Statistic Based on L_2-distance

We consider the weighted L_2-distance between the parametric and nonparametric fits of response probabilities for each j:

$$T_{n,j} = nh^{q/2} \int [m_{h,j}(x) - \mathfrak{K}_{h,n} m_{\hat{\theta},j}(X_i)]^2 \pi(x) dx,$$

where $\mathfrak{K}_{h,n} m(X_i) \equiv \sum_{i=1}^{n} K[(X_i - x)/h] m(X_i) / \sum_{i=1}^{n} K[(X_i - x)/h]$ is a smoothing operator, $m_{\hat{\theta},j}(X_i)$ is the estimate of $m_{\theta,j}(X_i)$, and $\pi(\cdot)$ is an arbitrary weight function. We denote the expectation of $T_{n,j}$ by $\mu_j^{(HM)}$, its asymptotic variance by $V_{j,j}^{(HM)}$, and the covariance between $T_{n,j}$ and $T_{n,m}$ by $V_{j,m}^{(HM)}$.

Let us introduce some further notations to provide the test statistic. Note that testing the specification of an arbitrary pair of $J-1$ response probabilities is a sufficient test for the null hypothesis subject to $\sum_{j=1}^{J} P(Y_{i,j} = 1|X_i) = 1$ for all i. For notational simplicity, we omit the L_2-distances of the Jth response probability from our test statistic. Let $T_n \equiv (T_{n,1}, T_{n,2}, \ldots, T_{n,J-1})'$ and $\hat{N}^{(HM)} \equiv (\hat{\mu}_1^{(HM)}, \hat{\mu}_2^{(HM)}, \ldots, \hat{\mu}_{J-1}^{(HM)})'$ be a $(J-1) \times 1$ vector of weighted L_2-distances and the estimates of $\mu_j^{(HM)}$, respectively. $\hat{V}^{(HM)}$ is defined as a $(J-1) \times (J-1)$ variance-covariance matrix whose (j,m) elements are estimates of $V_{j,m}^{(HM)}$. Then, the test statistic is

$$C_n^{(HM)} \equiv [T_n - \hat{N}^{(HM)}]' [\hat{V}^{(HM)}]^{-1} [T_n - \hat{N}^{(HM)}],$$

where

$$\begin{aligned}
\hat{\mu}_j^{(HM)} &= n h^{q/2} \int \left\{ \mathfrak{K}_{h,n} \left[m_{h,j,-i}(X_i) - m_{\hat{\theta},j}(X_i) \right] \right\}^2 \pi(x) dx \\
&\quad + h^{-q/2} K^{(2)}(0) \int \frac{\hat{\sigma}_j^2(x) \pi(x)}{f_h(x)} dx \\
\hat{V}_{j,j}^{(HM)} &= 2K^{(4)}(0) \int \frac{[\hat{\sigma}_j^2(x)]^2 \pi(x)^2}{f_h^2(x)} dx \\
\hat{V}_{j,m}^{(HM)} &= 2K^{(4)}(0) \int \frac{[\hat{\sigma}_{j,m}(x)]^2 \pi(x)^2}{f_h^2(x)} dx
\end{aligned}$$

for $j = 1, \ldots, J-1$ and $m \neq j$. $m_{h,j,-i}(X_i)$ is the leave-one-out Nadaraya–Watson kernel estimator for $m_j(X_i)$; that is, $m_{h,j,-i}(X_i) = \sum_{l \neq i}^{n} K[(X_l - X_i)/h] Y_{l,j} / \sum_{l \neq i}^{n} K[(X_l - X_i)/h]$. $\hat{\sigma}_j^2(\cdot)$ is the estimated conditional variance of $u_{i,j} \equiv Y_{i,j} - m_j(X_i)$, where $\mathbb{E}(u_{i,j}|X_i) = 0$, and $\hat{\sigma}_{j,m}(\cdot)$ is the estimated covariance between $u_{i,j}$ and $u_{i,m}$.

Considering the nature of the model, $\hat{\sigma}_j^2(\cdot)$ and $\hat{\sigma}_{j,m}(\cdot)$ can be easily obtained. Since $Y_{i,j}$ is a binary variable taking zero or one, $u_{i,j} = [1 - m_j(X_i)]\mathbb{1}(Y_{i,j} = 1) - m_j(X_i)\mathbb{1}(Y_{i,j} = 0)$, where $\mathbb{1}(\cdot)$ is an indicator function. The conditional variance of $u_{i,j}$ and the covariance between $u_{i,j}$ and $u_{i,m}$ can then be written straightforwardly as follows:

$$\sigma_j^2(X_i) \equiv \mathbb{E}(u_{i,j}^2|X_i) = m_j(X_i)[1 - m_j(X_i)] \tag{2}$$

$$\sigma_{j,m}(X_i) \equiv \mathbb{E}(u_{i,j}u_{i,m}|X_i) = -m_j(X_i)m_m(X_i). \tag{3}$$

Thus, their consistent parametric estimators under the null hypothesis are $\hat{\sigma}_j^2(x) = m_{\hat{\theta},j}(x)[1 - m_{\hat{\theta},j}(x)]$ and $\hat{\sigma}_{j,m}(x) = -m_{\hat{\theta},j}(x)m_{\hat{\theta},m}(x)$, respectively.

3.2 Test Statistic Based on Moment Conditions

The test statistic is based on

$$Z_j \equiv \mathbb{E}[u_{\theta,i,j}\mathbb{E}(u_{\theta,i,j}|X_i)f(X_i)],$$

where $u_{\theta,i,j} = Y_{i,j} - m_{j,\theta}(X_i)$. Under the null hypothesis, $Z_j = 0$, since $\mathbb{E}(u_{\theta,i,j}|X_i) = 0$. Under the alternative hypothesis, $\mathbb{E}[u_{\theta,i,j}\mathbb{E}(u_{\theta,i,j}|X_i)f(X_i)] = \mathbb{E}[\mathbb{E}(u_{\theta,i,j}|X_i)^2 f(X_i)] = \mathbb{E}\{[P(Y_{i,j} = 1|X_i) - m_{\theta,j}(X_i)]^2 f(X_i)\} > 0$.

The nonparametric estimates of Z_j, denoted as $Z_{n,j}$, can be obtained as follows:

$$Z_{n,j} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i}^{n} \frac{1}{h^q} K\left(\frac{X_i - X_l}{h}\right) \hat{u}_{\theta,i,j}\hat{u}_{\theta,l,j},$$

where $\hat{u}_{\theta,i,j} = Y_{i,j} - m_{j,\hat{\theta}}(X_i)$. We denote the asymptotic variance of $Z_{n,j}$ and the covariance between $Z_{n,j}$ and $Z_{n,m}$ by $V_{j,j}^{(Zh)}$ and $V_{j,m}^{(Zh)}$, respectively.

We introduce some further notations to provide the test statistic. Note that we omit the Jth alternative for the reason discussed above. Let $Z_n \equiv (Z_{n,1}, \ldots, Z_{n,J-1})'$ be a $(J-1) \times 1$ vector and $\hat{V}^{(Zh)}$ be a $(J-1) \times (J-1)$ variance-covariance matrix whose (j, m) elements are estimates of $V_{j,m}^{(Zh)}$. Then, the test statistic is

$$C_n^{(Zh)} = n^2 h^q Z_n' [\hat{V}^{(Zh)}]^{-1} Z_n,$$

where

$$\hat{V}_{j,j}^{(Zh)} = K^{(2)}(0) \frac{2}{n} \sum_{i=1}^{n} [\hat{\sigma}_j^2(X_i)]^2 f_h(X_i)$$
$$\hat{V}_{j,m}^{(Zh)} = K^{(2)}(0) \frac{2}{n} \sum_{i=1}^{n} [\hat{\sigma}_{j,m}(X_i)]^2 f_h(X_i)$$

for all $j = 1, \ldots, J-1$ and $j \neq m$. Note that $\hat{\sigma}_j^2(X_i)$ and $\hat{\sigma}_{j,m}(X_i)$ are consistent parametric estimators under the null hypothesis for (2) and (3), respectively.

4 THE ASYMPTOTIC BEHAVIOR

First, we provide sufficient assumptions to show the asymptotic behavior of the test statistics. Asymptotic distributions under the null hypothesis and alternative hypothesis are then given. Finally, we show the asymptotic behaviors of the test statistics under Pitman local alternatives.

4.1 Assumptions

The following are sufficient assumptions to show the test statistics' asymptotic behavior.

Assumption 1: X lies on a compact set. The marginal density of X_i, denoted as $f(\cdot)$, is continuously differentiable and bounded away from 0.

Assumption 2: $m(\cdot)$ is continuously differentiable on the support of X.

Assumption 3: $P(Y_{i,j} = 1|X_i) \neq 0$ and $P(Y_{i,j} = 1|X_i) \neq 1$, for all i and j. None of the alternatives is a perfect substitute for another.

Assumption 4: $\pi(\cdot)$ is continuously differentiable.

Assumption 5: $m_{j,\theta}(\cdot)$ is continuously differentiable, and $m_{j,\hat{\theta}}(\cdot) - m_{j,\theta}(\cdot) = o_p(1/\sqrt{n})$ for all j.

Assumption 1 establishes that the first-order derivative of $f(\cdot)$ is bounded. The assumption that X lies on a compact set may be considered a strong one because it excludes X to follow some tractable distributions such as the normal. However, it does not confine applications of the test to empirical study because, in general, observations rarely take an infinite value. The assumption that $f(\cdot)$ is bounded from 0 avoids the random denominator problem associated with a nonparametric kernel estimation. It is also straightforward to see that the first-order derivative of $m(\cdot)$ is also bounded under Assumptions 1 and 2.

Assumption 3 guarantees that $\sigma_j^2(X_i) \neq 0$ and $\sigma_{j,l}(X_i) \neq 0$ for any j and $l \neq j$ because $\sigma_j^2(X_i) = P(Y_{i,j} = 1|X_i)P(Y_{i,j} = 0|X_i)$ and $\sigma_{j,l}(X_i) = -P(Y_{i,j} = 1|X_i)P(Y_{i,l} = 1|X_i)$. It is also clear that $\sigma_j^2(X_i)$ and $\sigma_{j,l}(X_i)$ never tend to infinity owing to the nature of the model. The fact that no alternatives are perfect substitutes for each other ensures that the variance-covariance matrices $V^{(HM)}$ and $V^{(Zh)}$ are invertible.

We need Assumptions 4 and 5 to show the asymptotic behavior of $C_n^{(HM)}$ and $C_n^{(Zh)}$, respectively. The $\sqrt{n}$-consistency of the parametric estimation given in Assumption 5 can be obtained, for example, by maximal likelihood estimation of a multinomial probit or logit model.

The kernel function assumption is as follows:

Assumption 6: The kernel K is a symmetric function and satisfies $\int K(u)du = 1$, $\int |K(u)|du < \infty$, $\sup |K(u)| < \infty$, and $|uK(u)| \to 0$ as $|u| \to \infty$.

Assumption 6 is satisfied by commonly used second-order kernels, such as the Epanechnikov, Gaussian, and quartic kernels, and the two-times convolution product of the kernel is bounded under this assumption. Furthermore, the nonparametric density estimator and the Nadaraya–Watson kernel estimator are consistent under Assumption 1, 2, and 6 (see, for example, Theorem 4.1 of Härdle, Müller, Sperlich, & Werwatz, 2004).

4.2 Asymptotic Distribution under the Null Hypothesis

We provide propositions about the asymptotic distributions of $C_n^{(HM)}$ and $C_n^{(Zh)}$ under the null hypothesis. The proofs of the propositions are provided in the appendix.

Proposition 1. *Let Assumptions 1–4, and 6 hold. Then, under the null hypothesis,*

$$C_n^{(HM)} \xrightarrow{d} \chi^2_{(J-1)}$$

as $h \to 0$ and $nh^q \to \infty$.

Proposition 2. *Let Assumptions 1–3, 5, and 6 hold. Then, under the null hypothesis,*

$$C_n^{(Zh)} \xrightarrow{d} \chi^2_{(J-1)}$$

as $h \to 0$ and $nh^q \to \infty$.

Propositions 1 and 2 indicate that the asymptotic distributions of the test statistic $C_n^{(HM)}$ and $C_n^{(Zh)}$ under the null hypothesis are both chi-squared distributions with $J-1$ degrees of freedom. Therefore, we reject the null hypothesis that the parametric specification of the response probability is identical to the true one with a probability of one if the test statistic is larger than the $(1-\alpha)$ quantile of the chi-squared distribution with $J-1$ degrees of freedom, where α is the significance level.

4.3 Asymptotic Distribution under the Alternative Hypothesis

We show that both the test statistics are consistent, that is, their asymptotic power is equal to one. The proofs of the lemmas are provided in the appendix.

Lemma 1. *Let Assumptions 1–4, and 6 hold. Then, under the alternative hypothesis,*

$$\frac{1}{(n-1)h^{q/2}} \frac{T_{n,j} - \hat{\mu}_j^{(HM)}}{\sqrt{\hat{V}_{j,j}^{(HM)}}} \xrightarrow{p} \frac{\int [m_{\theta,j}(u) - m_j(u)]^2 \pi(u) du + O(h^{q/2})}{\{2K^{(4)}(0) \int \pi(x)^2 \{m_{\theta,j}(x)[1-m_{\theta,j}(x)]\}^2 f^{-2}(x) dx\}^{1/2}} > 0$$

for all j *as* $n \to \infty$ *and* $h \to 0$.

Lemma 2. *Let Assumptions 1–3, 5, and 6 hold. Then, under the alternative hypothesis,*

$$\frac{1}{nh^{2/q}} \frac{nh^{2/q} Z_{n,j}}{\sqrt{\hat{V}_{j,j}^{(Zh)}}} \xrightarrow{p} \frac{\mathbb{E}\{[m_{\theta,j}(X_i) - m_j(X_i)]^2 f(X_i)\}}{\sqrt{2K^{(2)}(0)\,\mathbb{E}\{m_{\theta,j}(X_i)^2[1 - m_{\theta,j}(X_i)]^2 f(X_i)\}}} > 0$$

for all j *as* $n \to \infty$ *and* $h \to 0$.

The proofs of Lemmas 1 and 2 provided in the appendix imply that $T_{n,j} - \hat{\mu}_j^{HM}$ and $nh^{2/q} Z_{n,j}$ diverge for all j as the sample size n increases and $\hat{V}_{j,j}^{(HM)}$ and $\hat{V}_{j,j}^{(Zh)}$ converges to constants, both strictly larger than zero. In addition, it is straightforward to see that the probability limits of $\hat{V}_{j,m}^{(HM)}$ and $\hat{V}_{j,m}^{(Zh)}$ under the alternative hypothesis are

$$\begin{aligned}
&2K^{(4)}(0) \int \pi(x)^2 m_{\theta,j}(x)^2 m_{\theta,m}(x)^2 f^{-2}(x)dx, \\
&2K^{(2)}(0)\,\mathbb{E}[m_{\theta,j}(X_i)^2 m_{\theta,m}(X_i)^2 f(X_i)],
\end{aligned}$$

respectively, both bounded above by Assumptions 1–4 and 6 for any $j \neq m$. Thus, the following propositions follow immediately.

Proposition 3. *Let Assumptions 1–4, and 6 hold. Then, under the alternative hypothesis,* $C_n^{(HM)}$ *diverges in probability, and thus the asymptotic power of the test is 1.*

Proposition 4. *Let Assumptions 1–3, 5, and 6 hold. Then, under the alternative hypothesis,* $C_n^{(Zh)}$ *diverges in probability, and thus the asymptotic power of the test is 1.*

The proofs of Propositions 3 and 4 are apparent from Lemmas 1 and 2 and the discussion on the probability limits of $\hat{V}_{j,m}^{(HM)}$ and $\hat{V}_{j,m}^{(Zh)}$ under the alternative hypothesis mentioned above.

4.4 Asymptotic Distribution under Pitman Local Alternative

We show that both the test statistics $C_n^{(HM)}$ and $C_n^{(Zh)}$ have nontrivial power against Pitman local alternatives approaching the null at the rate $1/\sqrt{nh^{q/2}}$. Proofs of the lemmas are provided in the appendix. Let us consider a sequence of local alternatives:

$$H_{1n} : P(Y_{i,j} = 1|X_i) = m_{\theta,j}(X_i) + \delta_n l_j(X_i),$$

where $l(\cdot)$ is a known continuous function with $\mathbb{E}[l(\cdot)^2] < \infty$ and $\delta_n \to 0$ at the rate $1/\sqrt{nh^{q/2}}$.

Lemma 3. *Let Assumptions 1–4, and 6 hold. Then, under the local alternative hypothesis,*

$$T_{n,j} - \hat{\mu}_j^{(HM)} \xrightarrow{d} N(M_j^{(HM)}, V_{j,j}^{(HM)}) \text{ for all } j,$$

where $M_j^{(HM)} \equiv \int [l_j(x) + 2m_j(x) - 2m_{\hat{\theta},j}(x)] l_j(x)\pi(x)dx.$

Lemma 4. *Let Assumptions 1–3, 5, and 6 hold. Then, under the local alternative hypothesis,*

$$nh^{q/2} Z_{n,j} \xrightarrow{d} N(M_j^{(Zh)}, V_{j,j}^{(Zh)}) \text{ for all } j,$$

where $M_j^{(Zh)} \equiv \mathbb{E}[l_j(x)^2 f(x)].$

Lemma 3 indicates that the limiting distribution of $[T_{n,j} - \hat{\mu}_j^{(HM)}]/V_{j,j}^{(HM)}$ is the normal distribution with mean $M_j^{(HM)}[V_{j,j}^{(HM)}]^{-1/2}$ and variance one. Similarly, Lemma 4 indicates that the limiting distribution of $nh^{q/2}Z_{n,j}/V_{j,j}^{(Zh)}$ is the normal distribution with mean $M_j^{(Zh)}[V_{j,j}^{(Zh)}]^{-1/2}$ and variance one. The following propositions show that both the test statistics can detect the local alternative with nontrivial powers.

Proposition 5. *Let Assumptions 1–4, and 6 hold. Then, under the local alternative hypothesis, the test statistic $C_n^{(HM)}$ converges to a non-central chi-squared distribution with $J-1$ degrees of freedom:*

$$C_n^{(HM)} \xrightarrow{d} \chi^2_{(J-1)}(\lambda)$$

where $\lambda = [M^{(HM)}]'[V^{(HM)}]^{-1}M^{(HM)}$ *is a noncentrality parameter.*

Proposition 6. *Let Assumptions 1–3, 5, and 6 hold. Then, under the local alternative hypothesis, the test statistic* $C_n^{(Zh)}$ *converges to a non-central chi-squared distribution with* $J-1$ *degrees of freedom:*

$$C_n^{(Zh)} \xrightarrow{d} \chi^2_{(J-1)}(\tilde{\lambda})$$

where $\tilde{\lambda} \equiv [M^{(Zh)}]'[V^{(Zh)}]^{-1}M^{(Zh)}$ *is a noncentrality parameter.*

The proofs of Propositions 5 and 6 are straightforward from Lemmas 3 and 4 and the discussion on the probability limits of $\hat{V}_{j,m}^{(HM)}$ and $\hat{V}_{j,m}^{(Zh)}$ for $j \neq m$ in the proofs of Propositions 1 and 2.

5 BOOTSTRAP METHODS

This section presents bootstrapping methods that are useful in approximating the distribution of test statistics when the sample size is small. We show that the parametric bootstrap procedure works well to calculate the rejection region for the test statistic. Proofs of the propositions in this section are provided in the appendix.

The response probability that person i chooses alternative j can be parametrically estimated under the null hypothesis for all i and j by using the observations $\{\{X_{i,j}, Y_{i,j}\}_{i=1}^n\}_{j=1}^J$. We randomly choose one of J alternatives (say, alternative m_i) for each person with the probabilities equal to the estimated response probabilities. Then, we derive bootstrap observations $Y_i^* \equiv \{Y_{i,1}^*, Y_{i,2}^*, \ldots, Y_{i,m_i}^*, \ldots, Y_{i,J}^*\}$ for each $i = 1, \cdots, n$, where $Y_{i,m_i}^* = 1$ and $Y_{i,j}^* = 0$ for $j \neq m_i$. We use $\{\{X_{i,j}, Y_{i,j}^*\}_{i=1}^n\}_{j=1}^J$ as the bootstrap observations.

Assumptions 3 and 5 can be rewritten by using the bootstrap observations as follows:

Assumption 3': $P(Y_{i,j}^* = 1|X_i) \neq 0$ and $P(Y_{i,j}^* = 1|X_i) \neq 1$, for all i and j. None of the alternatives is a perfect substitute for another.

Assumption 5': $m_{j,\theta}(\cdot)$ is continuously differentiable, and $m_{j,\hat{\theta}^*}(\cdot) - m_{j,\theta}(\cdot) = o_p(1/\sqrt{n})$ for all j,

where $\hat{\theta}^*$ is the estimate of θ obtained by using the bootstrap observations $\{\{X_{i,j}, Y^*_{i,j}\}_{i=1}^n\}_{j=1}^J$.

Since the bootstrap sample $Y^*_{i,j}$ is derived in accordance with the parametrically estimated response probabilities $m_{\hat{\theta},j}(X_i)$, Assumption 3' implies that these probabilities do not take the values zero and one; that is, $m_{\hat{\theta},j}(X_i) \neq 0$ and $m_{\hat{\theta},j}(X_i) \neq 1$, for all i and j.

Assumption 5' requires that $\hat{\theta}^*$ be a consistent estimator of θ. Clearly, Assumption 5' is satisfied whenever Assumption 5 holds. Moreover, Assumption 3' is also satisfied with a probability of one whenever Assumptions 3 and 5 hold.

5.1 Bootstrap Methods for $C_n^{(HM)}$

With the bootstrap observations $\{\{X_{i,j}, Y^*_{i,j}\}_{i=1}^n\}_{j=1}^J$, we construct $C_n^{*(HM)}$ similarly to $C_n^{(HM)}$. By the Monte Carlo approximation for the distribution of $C_n^{*(HM)}$, we can obtain the $(1-\alpha)$ quantile $t_\alpha^{*(HM)}$. The null hypothesis is rejected if $C_n^{(HM)} > t_\alpha^{*(HM)}$. We show in the following proposition that this parametric bootstrap procedure works: under the null hypothesis, $C_n^{*(HM)}$ converges to the asymptotic distribution of $C_n^{(HM)}$; under the alternative hypothesis, $C_n^{*(HM)}$ converges to the asymptotic distribution of test statistics under the null hypothesis. A crucial point that makes the parametric bootstrap work is that the orthogonality condition, $\mathbb{E}(u^*_{i,j}|X_i) = 0$, holds under both the null and alternative hypotheses with a probability of one, where $u^*_{i,j} \equiv Y^*_{i,j} - m^*_j(X_i)$, and $m^*_j(X_i)$ is the true response probability under the bootstrap sample. This is because the model deriving the bootstrap sampling is the parametric model that we attempt to test; that is, $m^*_j(X_i) = m_{\theta,j}(X_i)$.

Proposition 7. *Let Assumptions 1–4, and 6 hold. Then, the test statistic obtained with the bootstrap observation converges to a chi-squared distribution with $J-1$ degrees of freedom:*

$$C_n^{*(HM)} \xrightarrow{p} \chi^2_{(J-1)}$$

as $n \to \infty$ and $h \to 0$.

5.2 Bootstrap Methods for $C_n^{(Zh)}$

The test statistic $C_n^{*(Zh)}$ is constructed similarly to $C_n^{(Zh)}$ by using the bootstrap observations $\{\{X_{i,j}, Y_{i,j}^*\}_{i=1}^n\}_{j=1}^J$. By Monte Carlo approximation for the distribution of $C_n^{*(Zh)}$, we can obtain the $(1-\alpha)$ quantile $t_\alpha^{*(Zh)}$. The null hypothesis is rejected if $C_n^{(Zh)} > t_\alpha^{*(Zh)}$. We show in the following proposition that this parametric bootstrap procedure works: under the null hypothesis, $C_n^{*(Zh)}$ converges to the asymptotic distribution of $C_n^{(Zh)}$; under the alternative hypothesis, $C_n^{*(Zh)}$ converges to the asymptotic distribution of test statistics under the null hypothesis.

Proposition 8. *Let Assumptions 1–3, 5, and 6 hold. Then, the test statistic obtained with the bootstrap observation converges to a chi-squared distribution with $J-1$ degrees of freedom:*

$$C_n^{*(Zh)} \xrightarrow{p} \chi^2_{(J-1)}$$

as $n \to \infty$ and $h \to 0$.

6 MONTE CARLO EXPERIMENTS

The size and power of the tests are examined by Monte Carlo experiments. We consider a simple case in which each individual chooses one of three alternatives. To explore the power properties of the tests, we consider three different true models.

The null hypothesis to be tested is the following:

$$H_0 : P\left[m_{\theta,j}(X_i) = \frac{\exp(\beta_0 + \beta_1 X_{i,j})}{\sum_{j=1}^J \exp(\beta_0 + \beta_1 X_{i.j})}\right] = 1$$

for some $\beta_0, \beta_1 \in \mathbb{R}$ and for all $j = 1, 2, 3$. The null hypothesis is based on the assumptions that the function $g_j(X_{i,j}, \theta)$ is linear, specifically, $\beta_0 + \beta_1 X_{i,j}$, and that $\epsilon_{i,j}$ follows the type I extreme-value distribution for all j. For simplicity of calculation, $X_{i,j}$ is assumed to be one-dimensional.

We consider three different true models. Each of these true models has a specific form of $g_j(\cdot)$, which can be generally written as $g_j(X_{i,j}, \theta) = \gamma_j X_{i,j} + c_j(X_{i,j} - 1/2)^2 + d_j(2X_{i,j} - 2/3)^3$. By applying specific values in $\gamma \equiv \{\gamma_1, \gamma_2, \gamma_3\}$, $c \equiv (c_1, c_2, c_3)$, and $d \equiv (d_1, d_2, d_3)$, we propose three kinds of true models; Model 1: $\gamma = \{1, 1, 1\}$, $c = (0, 0, 1)$, and $d = (0, 0, 0)$, Model 2: $\gamma = \{1, 1, 5\}$, $c = (0, 3, 5)$, and $d = (0, 0, 0)$, and Model 3: $\gamma = \{1, 1, 1\}$, $c = (0, 3, 5)$, and $d = (0, 3, 5)$. The true distribution of $\epsilon_{i,j}$ is a type I extreme-value distribution for all j.

These true models allow us to investigate power properties of the tests in the case of misspecification due to nonlinearity and choice-specific coefficients. We can add nonlinearity to the true function of $g_j(\cdot)$ by setting c_j and/or d_j at a nonzero value, which is imposed on all true models. Choice-specific coefficients can be inserted by setting γ_j at different values across j, which is placed on Model 2. In this experiment, we do not consider the misspecification originating in the distribution of $\epsilon_{i,j}$ and the omitted variables.

We derive $\{\{X_{i,j}\}_{j=1}^3\}_{i=1}^n$ uniformly from [0,1] and $\{\{\epsilon_{i,j}\}_{j=1}^3\}_{i=1}^n$ randomly from the type I extreme-value distribution. Then, the latent variable y^* is generated by each true model: $y^*_{i,j} = g_j(X_{i,j}, \theta) + \epsilon_{i,j}$. The binary outcome $Y_{i,j}$ is chosen to be 1, if $y^*_{i,j} > y^*_{i,m}$ for all $m \neq j$, and 0 otherwise. Sample sizes are $n = 100$ and $n = 250$. The critical value for each test statistic is computed by $B = 100$ repetitions of the parametric bootstrap, and all results are based on $M = 1,000$ simulation runs.

For the nonparametric parts of the test statistics, $X_{i,j}$ are considered to be specific to each alternative, namely, $q = 3$. The quartic kernel $K(z) = (15/16)(1 - z^2)^2 \mathbb{1}(|z| < 1)$ is used for nonparametric estimation. Bandwidths for the kernel estimator are chosen to be $h \in \{0.30, 0.35, 0.40, 0.45\}$. We use the nonparametric density estimator of X_i as the weight function for $C_n^{(HM)}$; that is, $\pi(x) = f_h(x)$.

Table 1 illustrates the size of the tests at the 5% significance level. The first and second rows of the table show the size of the test statistics $C_n^{(HM)}$ and $C_n^{(Zh)}$, respectively. The first to fourth columns of the table illustrate the results obtained with a sample size of $n = 100$ and bandwidths

Table 1: *Monte Carlo estimates for the size of the test statistics $C_n^{(HM)}$ and $C_n^{(Zh)}$.*

	$n = 100$				$n = 250$			
Test\h	0.30	0.35	0.40	0.45	0.30	0.35	0.40	0.45
$C_n^{(HM)}$	0.062	0.046	0.059	0.055	0.047	0.054	0.055	0.054
$C_n^{(Zh)}$	0.058	0.064	0.063	0.077	0.065	0.074	0.064	0.049

Note: The significance level is 0.05.

h of 0.30, 0.35, 0.40, and 0.45, respectively. Similarly, the fifth to eighth columns show the result with $n = 250$. Overall, the probabilities of rejection by $C_n^{(HM)}$ vary with the bandwidths but stay around the nominal size. The size of the test statistic is close to its nominal value when $h = 0.35$ for $n = 100$ and $h = 0.30$ for $n = 250$. In contrast, $C_n^{(Zh)}$ tends to overreject the null hypothesis. The probability of rejection is close to its nominal size when $h = 0.45$ and $n = 250$.

In comparing the power performance of the tests, it is possible to correct size distortion by using the bandwidths corresponding to the nominal size of the tests. In practice, however, this procedure cannot be employed because we do not know the true model. Thus, we do not correct the size distortion in this experiment. We rather show the power performance with each bandwidth level, since choosing an appropriate bandwidth in practice is outside the scope of this paper.

Before beginning to show the simulation results of the power performance of the test statistics, we illustrate the discrepancy between the true and parametric null models. The response probabilities in this simulation are mappings of the unit cube to the unit interval. For illustration simplicity, however, we focus on the domain of the response probabilities, being $\{X_i = (X_{i,1}, X_{i,2}, X_{i,3}) : X_{i,j} \in [0, 1]$ for all j and $X_{i,1} = X_{i,2} = X_{i,3}\}$. In this setting, the fitted values for the response probabilities of the parametric model under the null hypothesis are always $1/3$ for all j because the model does not have any alternative-variant coefficients.

Figure 1 shows how the true and null response probabilities react to the covariates. The larger distance between the true and null models with x fixed indicates that the parametric null

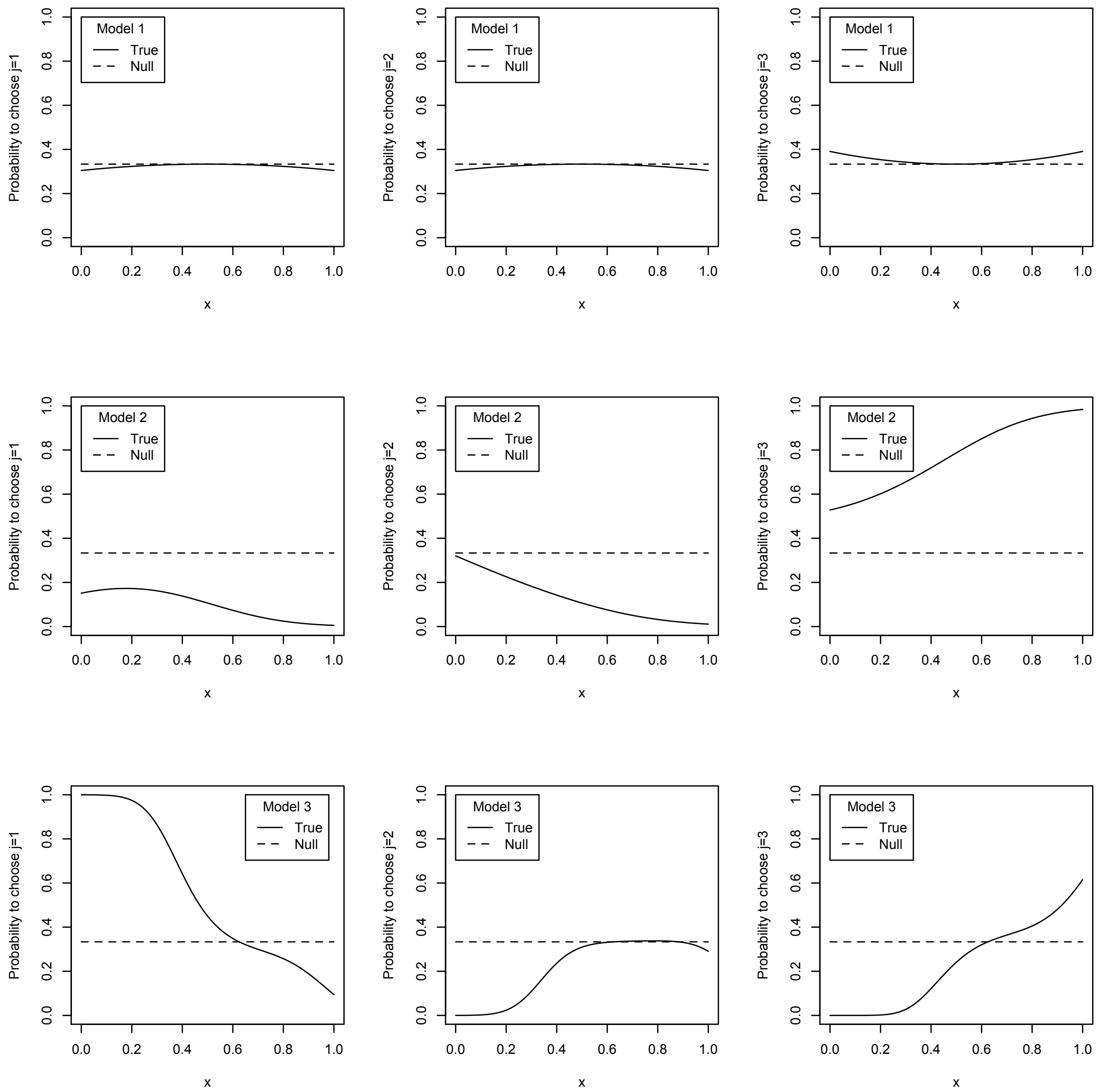

Figure 1: Discrepancy between true and estimated parametric response probabilities for Models 1, 2, and 3.

Table 2: *Monte Carlo results for the proportion of rejection of the null hypothesis by employing the test statistics* $C_n^{(HM)}$ *and* $C_n^{(Zh)}$.

		$n = 100$				$n = 250$			
Test	Model\h	0.30	0.35	0.40	0.45	0.30	0.35	0.40	0.45
$C_n^{(HM)}$	Model 1	0.056	0.052	0.056	0.077	0.053	0.075	0.064	0.063
	Model 2	0.377	0.952	0.985	0.931	0.932	0.988	0.864	0.575
	Model 3	0.149	0.298	0.190	0.109	0.721	0.616	0.418	0.334
$C_n^{(Zh)}$	Model 1	0.065	0.053	0.063	0.076	0.063	0.054	0.064	0.083
	Model 2	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Model 3	0.672	0.709	0.791	0.838	0.995	0.999	0.999	1.000

Note: The significance level is 0.05.

model does not approximate the true model well. The parametric predictions of the response probabilities lie close to the true response probability of Model 1 relative to Models 2 and 3 for all j. For the second and third alternatives, the parametric null response probability appears to lie closer to true one of Model 3 than that of Model 2. For the first alternative, however, the distance between the true and null models seems to closer for Models 3. Summing it up, the null model gives the best predictions of response probabilities for Model 1 and the predictions are less accurate for Models 2 and 3. The prediction precision of the null model could reflect in the power performance of the test statistics.

Table 2 reports the proportion of rejections of the null hypothesis at 5% significance by test statistics $C_n^{(HM)}$ and $C_n^{(Zh)}$. Both the test statistics have almost no nontrivial power when the true model is Model 1. Non-rejection of the null hypothesis does not imply that the null model is true. However, in fact, as the top three figures in Figure 1 exhibit, the parametric model under the null hypothesis may provide proper approximation for the response probabilities of Model 1. Therefore, the low power of the test statistics may be the acceptable result. In contrast, both the test statistics have more non-trivial power when Model 2 or 3 is true. The power performance improves along with the increase of sample size and depends on the choice of bandwidth.

By comparing the power performances of $C_n^{(HM)}$ and $C_n^{(Zh)}$ in Table 2, we find that $C_n^{(Zh)}$

tends to outperform $C_n^{(HM)}$, especially when the true model is Model 2 or 3. However, this finding may be specific to sample sizes and types of misspecifications set in this experiment. First, $C_n^{(HM)}$ incorporates more nonparametric components than $C_n^{(Zh)}$ because the asymptotic mean of $T_{n,j}$ is non-zero. Thus, $C_n^{(HM)}$ could be more scattered in a small sample, which impairs its power performance. Second, the relative power performances of these two tests may differ when we set other functional forms for $g(\cdot)$ and distributions for ϵ, because these test statistics are constructed on different bases, L_2-distance and moment conditions.

7 CONCLUSION

This study proposes two consistent specification tests for unordered multinomial choice models. They test the specifications of multiple response probabilities jointly for all choice alternatives. Both test statistics are asymptotically chi-square distributed with $J-1$ degrees of freedom, consistent against a fixed alternative, and have nontrivial power against local alternatives approaching the null at the rate $1/\sqrt{nh^{q/2}}$. The rejection region for the test statistic can be calculated through a simple parametric bootstrap procedure, when the sample size is small. In Monte Carlo experiments, we test the specification of the MNL model under three true models to examine the power performance of the tests. We found that both the test statistics have almost no nontrivial power when the parametric model under the null hypothesis provides a proper approximation for the response probabilities of the true model. Both the test statistics have more non-trivial power when the approximation of the null model is less successful. The test performance depends on the choice of bandwidth. We can reduce size distortion by choosing an appropriate bandwidth, but this issue remains for future research.

The tests proposed in this study can be applied to testing the parametric specifications of response probabilities for any unordered multinomial choice models, including MNL and MNP models. However, these tests are not able to detect local alternatives approaching the null hypothesis at the parametric rate, nor are they rate-optimal. Extending the testing procedure to incorporate such features is left for future research.

NOTES

[1] Rate optimal tests are proposed by J. Fan and Huang (2001), Horowitz and Spokoiny (2001), Spokoiny (2001), Baraud, Huet, and Laurent (2003), Zhang (2003), and Guerre and Lavergne (2005), among others.

[2] To be accurate, the MNL model proposed by McFadden (1974) consists of alternative-variant coefficients, whose response probabilities are indicated by $P(Y_j = 1|X) = \exp(X'\beta_j)/[1 + \sum_{j=1}^{J} \exp(X'\beta_j)]$. However, the models represented by alternative-variant coefficients are able to transform into a model with alternative-invariant coefficients without loss of generality, which is sometimes called a conditional logit model. In this paper, we describe only the model with alternative-invariant coefficients.

References

Aït-Sahalia, Y., Bickel, P. J., & Stoker, T. M. (1994). *Goodness-of-fit tests for regression using kernel methods*. Sloan School of Management, Massachusetts Institute of Technology.

Andrews, D. W. (1997). A conditional kolmogorov test. *Econometrica*, *65(5)*, 1097–1128.

Baraud, Y., Huet, S., & Laurent, B. (2003). Adaptive tests of linear hypotheses by model selection. *Annals of Statistics*, *31(1)*, 225–251.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63(4)*, 841–890.

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics*, *20(1)*, 105–134.

Bierens, H. J. (1984). Model specification testing of time series regressions. *Journal of Econometrics*, *26(3)*, 323–353.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica*, *58(6)*, 1443–1458.

Bierens, H. J., & Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica*, *65(5)*, 1129–1151.

Chen, X., & Fan, Y. (1999). Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. *Journal of Econometrics*, *91(2)*, 373–401.

De Jong, R. M. (1996). The bierens test under data dependence. *Journal of Econometrics*, *72(1)*, 1–32.

Delgado, M. A. (1993). Testing the equality of nonparametric regression curves. *Statistics & Probability Letters*, *17(3)*, 199–204.

Delgado, M. A., & Stengos, T. (1994). Semiparametric specification testing of non-nested econometric models. *The Review of Economic Studies*, *61(2)*, 291–303.

Dubin, J. A., & McFadden, D. L. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica*, *52(2)*, 345–362.

Eubank, R. L., & Spiegelman, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association*, *85(410)*, 387–392.

Fan, J., & Huang, L.-S. (2001). Goodness-of-fit tests for parametric regression models. *Journal of the American Statistical Association*, *96(454)*, 640–652.

Fan, Y., & Li, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica*, *64(4)*, 865–890.

Goldberg, P. K. (1995). Product differentiation and oligopoly in international markets: The case of the US automobile industry. *Econometrica*, *63(4)*, 891–951.

Gozalo, P. L. (1993). A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory*, *9(3)*, 451–451.

Guerre, E., & Lavergne, P. (2005). Data-driven rate-optimal specification testing in regression models. *Annals of Statistics*, *33(2)*, 840–870.

Härdle, W., & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, *21(4)*, 1926–1947.

Härdle, W., Müller, M., Sperlich, S., & Werwatz, A. (2004). *Nonparametric and semiparametric models*. Springer.

Hausman, J. A., & Wise, D. A. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*, *46(2)*, 403–426.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, *5(4)*, 475–492.

Hong, Y., & White, H. (1995). Consistent specification testing via nonparametric series regression. *Econometrica*, *63(5)*, 1133–1159.

Horowitz, J. L., & Härdle, W. (1994). Testing a parametric model against a semiparametric alternative. *Econometric Theory*, *10(5)*, 821–848.

Horowitz, J. L., & Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, *69(3)*, 599–631.

Lavergne, P., & Vuong, Q. (2000). Nonparametric significance testing. *Econometric Theory*, *16(4)*, 576–601.

Lavergne, P., & Vuong, Q. H. (1996). Nonparametric selection of regressors: The nonnested case. *Econometrica*, *64(1)*, 207–219.

Le Cessie, S., & van Houwelingen, J. C. (1991). A goodness-of-fit test for binary regression models, based on smoothing methods. *Biometrics*, *47(4)*, 1267–1282.

Li, Q., & Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics*, *87(1)*, 145–165.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (p. 105-142). New York: Academic Press.

Mora, J., & Moro-Egido, A. I. (2008). On specification testing of ordered discrete choice models.

Journal of Econometrics, *143(1)*, 191–205.

Powell, J. L., Stock, J. H., & Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, *57(6)*, 1403–1430.

Spokoiny, V. (2001). Data-driven testing the fit of linear models. *Mathematical Methods of Statistics*, *10(4)*, 465–497.

Stinchcombe, M. B., & White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory*, *14(3)*, 295–325.

Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, *25(2)*, 613–641.

Whang, Y.-J. (2000). Consistent bootstrap tests of parametric regression functions. *Journal of Econometrics*, *98(1)*, 27–46.

Wooldridge, J. M. (1992). A test for functional form against nonparametric alternatives. *Econometric Theory*, *8(4)*, 452–475.

Yatchew, A. J. (1992). Nonparametric regression tests based on an infinite dimensional least squares procedure. *Econometric Theory*, *8(4)*, 435–451.

Zhang, C. M. (2003). Adaptive tests of regression functions via multiscale generalized likelihood ratios. *Canadian Journal of Statistics*, *31(2)*, 151–171.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, *75(2)*, 263–289.

APPENDIX

Proof of Proposition 1. It suffices to prove the following:

$$T_{n,j} - \hat{\mu}_j^{(HM)} \xrightarrow{d} N(0, V_{j,j}^{(HM)}), \tag{A.1}$$

$$V_{j,j}^{(HM)} - \hat{V}_{j,j}^{(HM)} = o_p(1), \tag{A.2}$$

$$V_{j,m}^{(HM)} - \hat{V}_{j,m}^{(HM)} = o_p(1), \tag{A.3}$$

where $V_{j,j}^{(HM)}$ and $V_{j,m}^{(HM)}$ are the asymptotic variance of $T_{n,j}$ and covariance between $T_{n,j}$ and $T_{n,m}$, respectively. We show that they can be written as follows:

$$V_{j,j}^{(HM)} \equiv 2K^{(4)}(0)\int \frac{[\sigma_j^2(x)]^2\pi(x)^2}{f^2(x)}dx,$$
$$V_{j,m}^{(HM)} \equiv 2K^{(4)}(0)\int \frac{[\sigma_{j,m}(x)]^2\pi(x)^2}{f^2(x)}dx.$$

Proof of (A.1). Since $\mathbb{E}(u_{i,j}|X_i) = 0$, we obtain the following:

$$\begin{aligned}
T_{n,j} &= nh^{q/2}\int [m_{h,j}(x) - \mathfrak{K}_{h,n} m_{\hat{\theta},j}(X_i)]^2\pi(x)dx \\
&= nh^{q/2}\int \left\{\mathfrak{K}_{h,n}[m_j(X_i) - m_{\hat{\theta}.j}(X_i)] + \mathfrak{K}_{h,n}u_{i,j}\right\}^2 \pi(x)dx \\
&= nh^{q/2}\int \left\{\mathfrak{K}_{h,n}[m_j(X_i) - m_{\hat{\theta},j}(X_i)]\right\}^2 \pi(x)dx \\
&+ \frac{1}{nh^{3q/2}}\int \frac{\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)^2 u_{i,j}^2}{f_h^2(x)}\pi(x)dx \\
&+ \frac{1}{nh^{3q/2}}\int \frac{\sum_{i=1}^n \sum_{l\neq i} K\left(\frac{X_i-x}{h}\right) K\left(\frac{X_l-x}{h}\right) u_{i,j}u_{l,j}}{f_h^2(x)}\pi(x)dx \\
&\equiv T_{1,j} + T_{2,j} + T_{3,j}.
\end{aligned}$$

We will show that

$$T_{1,j} + T_{2,j} - \hat{\mu}_j^{(HM)} = o_p(1), \tag{A.4}$$
$$T_{3,j} \xrightarrow{d} N(0, V_{j,j}^{(HM)}). \tag{A.5}$$

Proof of (A.4). It follows that

$$T_{2,j} = \frac{1}{nh^{3q/2}}\int \frac{\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)^2 u_{i,j}^2}{f_h^2(x)}\pi(x)dx$$

$$
\begin{aligned}
&= \frac{1}{nh^{q/2}} K^{(2)}(0) \sum_{i=1}^{n} u_{i,j}^2 \frac{\pi(X_i)}{f_h^2(X_i)} + o(1) \\
&= \frac{1}{h^{q/2}} K^{(2)}(0) \int \frac{\sigma_j^2(x)\pi(x)}{f_h(x)} dx + o_p(1).
\end{aligned}
$$

Thus, the proof of (A.4) is straightforward because the leave-one-out Nadaraya–Watson kernel estimator for $m_j(X_i)$ is consistent under Assumption 1, 2, and 6, and the parametric estimator for $\sigma_j^2(x)$ is consistent under the null hypothesis.

Proof of (A.5). It is clear that $T_{3,j}$ can be treated as a second-order degenerate U-statistic:

$$
\frac{1}{n-1} T_{3,j} = \frac{1}{n(n-1)h^{3q/2}} \sum_{i=1}^{n} \sum_{l \neq i} u_{i,j} u_{l,j} \bar{K}(X_i, X_l),
$$

where $\bar{K}(X_i, X_l) \equiv \int \frac{K[(X_i - x)/h] K[(X_l - x)/h]}{f_h^2(x)} \pi(x) dx$. Letting $Z_i = \{X_i, u_i\}$, we define $G_n(Z_1, Z_2) = \mathbb{E}_{Z_i}\{[u_{1,j} u_{i,j} \bar{K}(X_1, X_i)][u_{2,j} u_{i,j} \bar{K}(X_2, X_i)]\}$. According to the central limit theorem for degenerate U-statistics proposed by Hall (1984),

$$
\frac{T_{3,j}}{\frac{n-1}{nh^{3q/2}} \sqrt{2\mathbb{E}\{[u_{1,j} u_{2,j} \bar{K}(X_1, X_2)]^2\}}} \xrightarrow{d} N(0,1)
$$

if

$$
\frac{\mathbb{E}[G_n^2(Z_1, Z_2)] + n^{-1}\mathbb{E}\{[u_{1,j} u_{2,j} \bar{K}(X_1, X_2)]^4\}}{\mathbb{E}\{[u_{1,j} u_{2,j} \bar{K}(X_1, X_2)]^2\}^2} \to 0 \quad \text{as } n \to \infty. \tag{A.6}
$$

Thus, it is enough to show that (A.6) and the following hold:

$$
2h^{-3q} \mathbb{E}\{[u_{1,j} u_{2,j} \bar{K}(X_1, X_2)]^2\} \xrightarrow{p} V_{j,j}^{(HM)}. \tag{A.7}
$$

Proof of (A.6). First, straightforward calculation gives

$\mathbb{E}[G_n^2(Z_1, Z_2)]$

$$= \mathbb{E}\left\{\sigma_j^2(X_1)\sigma_j^2(X_2)\left[\int \sigma_j^2(x)\bar{K}(X_1,x)\bar{K}(X_2,x)f(x)dx\right]^2\right\}$$
$$= h^{4q}\int \sigma_j^2(x)\sigma_j^2(y)\left[h^q\frac{\pi(x)\pi(y)}{f_h^2(x)f_h^2(y)}\sigma_j^2(x)f(x)K^{(4)}\left(\frac{x-y}{h}\right)+O(h^q)\right]^2 f(x)f(y)dxdy$$
$$= h^{7q}K^{(6)}(0)\int \frac{[\pi(x)]^3[\sigma_j^2(x)]^3 f(x)^4}{[f_h^2(x)]^3}dx + O(h^{8q})$$
$$= O(h^{7q}), \tag{A.8}$$

where $K^{(6)}$ is defined as the two-times convolution product of $K^{(4)}$.

Second, in the same way as above, we obtain

$$\begin{aligned}
&n^{-1}\mathbb{E}\{[u_{1,j}u_{2,j}\bar{K}(X_1,X_2)]^4\}\\
&\quad= \frac{1}{n}\mathbb{E}\left\{\sigma^4(x_1)\sigma^4(x_2)\left[\int \frac{K(\frac{x_1-x}{h})K(\frac{x_2-x}{h})}{f_h^2(x)}\pi(x)dx\right]^4\right\}\\
&\quad= \frac{h^{4q}}{n}\mathbb{E}\left\{\frac{\sigma^4(x_1)\sigma^4(x_2)\pi^4(x_1)}{[f_h^2(x_1)]^4}\left[K^{(2)}\left(\frac{x_1-x_2}{h}\right)\right]^4\right\}+O\left(\frac{h^{9q}}{n}\right)\\
&\quad= \frac{h^{5q}}{n}\int\left[K^{(2)}(u)\right]^4du\int\frac{[\sigma^4(x)]^2\pi^4(x)}{[f_h^2(x)]^4}f^2(x)dx+O\left(\frac{h^{6q}}{n}\right)\\
&\quad= O\left(\frac{h^{5q}}{n}\right).
\end{aligned} \tag{A.9}$$

The last equation holds because $\sigma_j^4(x) \equiv \mathbb{E}[u_{i,j}^4|X_i = x]$ is bounded by Assumption 1 and the fact that $Y_{i,j}$ is a binary variable taking the values zero and one.

Next, we obtain the following:

$$\begin{aligned}
\mathbb{E}\{[u_{1,j}u_{2,j}\bar{K}(X_1,X_2)]^2\}^2 &= \mathbb{E}\{\mathbb{E}\{[u_{1,j}u_{2,j}\bar{K}(X_1,X_2)]^2|X_1,X_2\}\}^2\\
&= \mathbb{E}\left\{\sigma_j^2(x_1)\sigma_j^2(x_2)\left[\int\frac{K(\frac{x_1-x}{h})K(\frac{x_2-x}{h})}{f_h^2(x)}\pi(x)dx\right]^2\right\}^2\\
&= O(h^{5q}).
\end{aligned} \tag{A.10}$$

Finally, (A.8)–(A.10) indicate that (A.6) holds because $O(h^{7q})O(h^{5q}/n)/O(h^{5q}) \to 0$ as $n \to \infty$ and $h \to 0$.

Proof of (A.7). It can be shown by straightforward calculation that

$$
\begin{aligned}
&2h^{-3q}\mathbb{E}\{[u_{1,j}u_{2,j}\bar{K}(X_1, X_2)]^2\} \\
&= 2h^{-3q}\mathbb{E}\left\{\sigma_j^2(x_1)\sigma_j^2(x_2)\left[\int \frac{K(\frac{x_1-x}{h})K(\frac{x_2-x}{h})}{f_h^2(x)}\pi(x)dx\right]^2\right\} \\
&= 2K^{(4)}(0)\int \frac{[\sigma_j^2(x)]^2\pi(x)^2}{[f_h^2(x)]^2}f(x)^2dx + O(h). \\
&\xrightarrow{p} V_{j,j}^{(HM)}
\end{aligned} \tag{A.11}
$$

because the nonparametric density estimator f_h is consistent under Assumptions 1 and 6.

Proof of (A.2) and (A.3). Since the asymptotic variance $V_{j,j}^{(HM)}$ is shown above, we derive the asymptotic covariance $V_{j,m}^{(HM)}$. According to the result of (A.1), it is clear that $\mathbb{E}(T_{3,j}T_{3,m}) \xrightarrow{p} V_{j,m}^{(HM)}$ as $n \to \infty$. Because $\mathbb{E}(u_{i,j}u_{l,j}) = 0$ if $i \neq l$, and $\mathbb{E}(u_{i,j}u_{i,m}|X_i) = \sigma_{j,m}(X_i)$ if $j \neq m$, it follows that

$$
\begin{aligned}
\mathbb{E}(T_{3,j}T_{3,m}) &= \frac{1}{n^2h^{3q}}\mathbb{E}\left[\int \frac{\sum_i^n \sum_{l\neq i} K\left(\frac{X_i-x}{h}\right)K\left(\frac{X_l-x}{h}\right)u_{i,j}u_{l,j}}{f_h^2(x)}\pi(x)dx\right] \\
&\left[\int \frac{\sum_s^n \sum_{t\neq s} K\left(\frac{X_s-y}{h}\right)K\left(\frac{X_t-y}{h}\right)u_{s,m}u_{t,m}}{f_h^2(y)}\pi(y)dy\right] \\
&= 2n^{-2}h^{-q}\mathbb{E}\left[\sum_i^n\sum_{l\neq i}u_{i,j}u_{l,j}u_{i,m}u_{l,m}\int \frac{K\left(\frac{X_i-x}{h}\right)K\left(\frac{X_l-x}{h}\right)}{f_h^2(x)}\pi(x)dx\right. \\
&\left.\int \frac{K\left(\frac{X_s-y}{h}\right)K\left(\frac{X_t-y}{h}\right)}{f_h^2(y)}\pi(y)dy\right] \\
&= 2K^{(4)}(0)\int \frac{[\sigma_{j,m}(x)]^2\pi(x)^2}{[f_h^2(x)]^2}f(x)^2dx + O(h) \\
&\xrightarrow{p} V_{j,m}^{(HM)}.
\end{aligned} \tag{A.12}
$$

Thus, the proofs of (A.2) and (A.3) are straightforward from (A.11) and (A.12). ■

Proof of Proposition 2. It suffices to prove the following:

$$nh^{q/2}\hat{Z}_{n,j} \xrightarrow{d} N(0, V_{j,j}^{(Zh)}), \tag{A.13}$$

$$V_{j,j}^{(Zh)} - \hat{V}_{j,j}^{(Zh)} = o_p(1), \tag{A.14}$$

$$V_{j,m}^{(Zh)} - \hat{V}_{j,m}^{(Zh)} = o_p(1), \tag{A.15}$$

where $V_{j,j}^{(Zh)}$ and $V_{j,m}^{(Zh)}$ are the asymptotic variance of $nh^{q/2}\hat{Z}_{n,j}$ and the covariance between $nh^{q/2}\hat{Z}_{n,j}$ and $nh^{q/2}\hat{Z}_{n,m}$, respectively. We show that they can be written as follows:

$$V_{j,j}^{(Zh)} \equiv 2K^{(2)}(0)\mathbb{E}\{[\sigma^2(x)]^2 f(x)\},$$
$$V_{j,m}^{(Zh)} \equiv 2K^{(2)}(0)\mathbb{E}\{[\sigma_{j,m}(x)]^2 f(x)\}.$$

Proof of (A.13). Under the null hypothesis, we have $m_j(\cdot) = m_{\theta,j}(\cdot)$. Thus, it follows that

$$\begin{aligned}
nh^{q/2}\hat{Z}_{n,j} &= \frac{1}{h^{q/2}(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}^{n} K\left(\frac{X_i - X_l}{h}\right) u_{\hat{\theta},i,j}u_{\hat{\theta},l,j} \\
&= \frac{1}{h^{q/2}(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}^{n} K\left(\frac{X_i - X_l}{h}\right)[m_{\theta,j}(X_i) - m_{\hat{\theta},j}(X_i) + u_{i,j}] \\
&\quad [m_{\theta,j}(X_l) - m_{\hat{\theta},j}(X_l) + u_{l,j}] \\
&= \frac{1}{h^{q/2}(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}^{n} K\left(\frac{X_i - X_l}{h}\right) \\
&\quad [m_{\theta,j}(X_i) - m_{\hat{\theta},j}(X_i)][m_{\theta,j}(X_l) - m_{\hat{\theta},j}(X_l)] \\
&\quad + \frac{1}{h^{q/2}(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}^{n} K\left(\frac{X_i - X_l}{h}\right) u_{i,j}u_{l,j} \\
&\quad + \frac{1}{h^{q/2}(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}^{n} K\left(\frac{X_i - X_l}{h}\right)[m_{\theta,j}(X_i) - m_{\hat{\theta},j}(X_i)]u_{l,j}
\end{aligned}$$

$$+\frac{1}{h^{q/2}(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}^{n}K\left(\frac{X_i-X_l}{h}\right)[m_{\theta,j}(X_l)-m_{\hat{\theta},j}(X_l)]u_{i,j}$$

$$\equiv Z_{1,j}+Z_{2,j}+Z_{3,j}+Z_{4,j}.$$

We will prove the following:

$$Z_{1,j}=o_p(1), \tag{A.16}$$

$$Z_{2,j}\xrightarrow{d}N\left(0,V_{j,j}^{(Zh)}\right), \tag{A.17}$$

$$Z_{3,j}=o_p(1), \tag{A.18}$$

$$Z_{4,j}=o_p(1). \tag{A.19}$$

Proof of (A.16). Assumptions 1, 5, and 6 along with straightforward calculation show that $Z_{1,j}=o_p(1)$.

Proof of (A.17). Note that $Z_{2,j}$ can be treated as a second-order degenerate U-statistic:

$$\frac{h^{q/2}}{n}Z_{2,j}=\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}^{n}K\left(\frac{X_i-X_l}{h}\right)u_{i,j}u_{l,j}.$$

Define $G_n(Z_1,Z_2)=\mathbb{E}_{Z_i}[\{K[(X_1-X_i)/h]\,u_{1,j}u_{i,j}\}\{K[(X_2-X_i)/h]\,u_{2,j}u_{i,j}\}]$, where $Z_i=\{X_i,u_i\}$. According to the central limit theorem for degenerat U-statistics proposed by Hall (1984),

$$\frac{Z_{2,j}}{h^{-q/2}\sqrt{2\mathbb{E}\{[u_{1,j}u_{2,j}K\left(\frac{X_1-X_2}{h}\right)]^2\}}}\xrightarrow{d}N(0,1) \tag{A.20}$$

if

$$\frac{\mathbb{E}[G_n^2(Z_1,Z_2)]+n^{-1}\mathbb{E}\{[u_{1,j}u_{2,j}K\left(\frac{X_1-X_2}{h}\right)]^4\}}{\mathbb{E}\{[u_{1,j}u_{2,j}K\left(\frac{X_1-X_2}{h}\right)]^2\}^2}\to 0 \quad \text{as } n\to\infty. \tag{A.21}$$

Thus, it is enough to show that (A.21) and the following hold:

$$\frac{2}{h^q}\mathbb{E}\left\{\left[u_{1,j}u_{2,j}K\left(\frac{X_1-X_2}{h}\right)\right]^2\right\} \to V_{j,j}^{(Zh)}. \tag{A.22}$$

Proof of (A.21). First, straightforward calculation gives

$$\begin{aligned}
\mathbb{E}[G_n^2(Z_1,Z_2)] &= \mathbb{E}\left[\left\{\mathbb{E}_{Z_i}\left[u_{1,j}u_{2,j}u_{i,j}^2K\left(\frac{X_1-X_i}{h}\right)K\left(\frac{X_2-X_i}{h}\right)\right]\right\}^2\right] \\
&= \mathbb{E}\left\{\sigma_j^2(X_1)\sigma_j^2(X_2)\left[\int \sigma_j^2(z)K\left(\frac{X_1-z}{h}\right)K\left(\frac{X_2-z}{h}\right)f(z)dz\right]^2\right\} \\
&= h^{3q}K^{(4)}(0)\int[\sigma_j^2(x)]^4 f(x)^4 dx + O(h^{3q+1}) + o(h^{3q+1}) \\
&= O(h^{3q}).
\end{aligned} \tag{A.23}$$

In the same way, it can be shown that

$$\begin{aligned}
\frac{1}{n}\mathbb{E}\left\{\left[u_{1,j}u_{2,j}K\left(\frac{X_1-X_2}{h}\right)\right]^4\right\} &= \frac{1}{n}\int \sigma_j^4(x)\sigma_j^4(y)\left[K\left(\frac{x-y}{h}\right)\right]^4 f(x)f(y)dxdy \\
&= \frac{h^q}{n}\int[\sigma_j^4(x)]^2 f^2(x)dx\int[K(u)]^4du + O\left(\frac{h^{2q}}{n}\right) \\
&= O\left(\frac{h^q}{n}\right).
\end{aligned} \tag{A.24}$$

Next, after some calculation, we obtain

$$\begin{aligned}
\mathbb{E}\left\{\left[u_{1,j}u_{2,j}K\left(\frac{X_1-X_2}{h}\right)\right]^2\right\}^2 &= \mathbb{E}\left\{\sigma^2(X_1)\sigma^2(X_2)\left[K\left(\frac{X_1-X_2}{h}\right)\right]^2\right\}^2 \\
&= h^{2q}\left\{K^{(2)}(0)\int[\sigma^2(x)]^2 f^2(x)dx + O(h)\right\}^2 \\
&= O(h^{2q}).
\end{aligned} \tag{A.25}$$

Finally, (A.23)–(A.25) indicate that (A.21) holds because $\frac{O(h^{3q})+O\left(\frac{h^q}{n}\right)}{O(h^{2q})} \to 0$ as $h \to 0$ and $nh^q \to \infty$.

Proof of (A.22). From equation (A.25), it is clear that

$$
\begin{aligned}
\frac{2}{h^q}\mathbb{E}\left\{\left[u_{1,j}u_{2,j}K\left(\frac{X_1-X_2}{h}\right)\right]^2\right\} &= 2K^{(2)}(0)\int[\sigma^2(x)]^2f^2(x)dx+O(h)\\
&= 2K^{(2)}(0)\mathbb{E}\{[\sigma^2(x)]^2f(x)\}+O(h)\\
&\to V_{j,j}^{(Zh)}. \qquad\qquad (A.26)
\end{aligned}
$$

Proof of (A.18) and (A.19). (A.18) and (A.19) are straightforward because $\mathbb{E}(u_{i,j}|X_i)=0$ and $\mathbb{E}(u_{l,j}|X_l)=0$ from the definition. First, we show that $Z_{3,j}=o_p(1)$:

$$
\begin{aligned}
Z_{3,j} &= \frac{1}{(n-1)h^{q/2}}\sum_{i=1}^{n}\sum_{l\neq i}^{n}K\left(\frac{X_i-X_l}{h}\right)[m_{\theta,j}(X_i)-m_{\hat{\theta},j}(X_i)]u_{l,j}\\
&= \frac{1}{h^{q/2}}\sum_{i=1}^{n}\left\{\mathbb{E}_{X_l}\left[K\left(\frac{X_i-X_l}{h}\right)o_p(1/\sqrt{n})u_{l,j}\right]+o_p(1/\sqrt{n})\right\}\\
&= \frac{1}{h^{q/2}}o_p(1/\sqrt{n})=o_p(1/\sqrt{nh^q})=o_p(1).
\end{aligned}
$$

We can also prove (A.19) by similar calculation.

Proof of (A.14) and (A.15). Since the asymptotic variance is shown above, we derive the asymptotic covariance between $nh^{q/2}\hat{Z}_{n,j}$ and $nh^{q/2}\hat{Z}_{n,m}$, which we denote as $V_{j,m}^{(Zh)}$. From the results of (A.16)–(A.19), it is clear that $\mathbb{E}(Z_{2,j}Z_{2,m})\to V_{j,m}^{(Zh)}$ as $n\to\infty$. Because $\mathbb{E}(u_{i,j}u_{l,j})=0$ if $i\neq l$, and $\mathbb{E}(u_{i,j}u_{i,m}|X_i)=\sigma_{j,m}(X_i)$ if $j\neq m$, it follows that

$$
\begin{aligned}
&\mathbb{E}(Z_{2,j}Z_{2,m})\\
&= \frac{1}{(n-1)^2h^q}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{l\neq 1}^{n}K\left(\frac{X_i-X_l}{h}\right)u_{i,j}u_{l,j}\sum_{s=1}^{n}\sum_{t\neq s}^{n}K\left(\frac{X_s-X_t}{h}\right)u_{s,m}u_{t,m}\right]\\
&= \frac{2}{(n-1)^2h^q}\mathbb{E}\left\{\sum_{i=1}^{n}\sum_{l\neq 1}^{n}u_{i,j}u_{l,j}u_{i,m}u_{l,m}\left[K\left(\frac{X_i-X_l}{h}\right)\right]^2\right\}\\
&= \frac{2n}{(n-1)h^q}\int\sigma_{j,m}(x)\sigma_{j,m}(y)\left[K\left(\frac{x-y}{h}\right)\right]^2f(x)f(y)dxdy
\end{aligned}
$$

$$
\begin{aligned}
&= 2K^{(2)}(0) \int [\sigma_{j,m}(x)]^2 f^2(x) dx + O(h) \\
&\to V_{j,m}^{(Zh)}. \qquad \text{(A.27)}
\end{aligned}
$$

Thus, the proofs of (A.14) and (A.15) are straightforward from (A.26) and (A.27). ■

Proof of Lemma 1. According to the proof of Proposition 1, we can write $T_{n,j} = T_{1,j} + T_{2,j} + T_{3,j}$. Thus, it is enough to show the following:

$$
\begin{aligned}
T_{1,j} + T_{2,j} - \hat{\mu}_j^{(HM)} &= \frac{1}{h^{q/2}} K^{(2)}(0) \int \frac{\{\sigma_j^2(x) - m_{\hat{\theta},j}(x)[1 - m_{\hat{\theta},j}(x)]\}\pi(x)}{f_h(x)} dx + o_p(1) \\
&= O(h^{q/2}), \qquad \text{(A.28)} \\
\frac{1}{(n-1)h^{q/2}} T_{3,j} &= \int [m_{\theta,j}(u) - m_j(u)]^2 \pi(u) du + o_p(1), \qquad \text{(A.29)} \\
\hat{V}_{j,j}^{(HM)} &= 2K^{(4)}(0) \int \pi(x)^2 \{m_{\theta,j}(x)[1 - m_{\theta,j}(x)]\}^2 f^{-2}(x) dx + o_p(1). \\
& \qquad \text{(A.30)}
\end{aligned}
$$

Since $\hat{\sigma}_j^2(x) = m_{\hat{\theta},j}(x)[1 - m_{\hat{\theta},j}(x)]$ converges to $m_{\theta,j}(x)[1 - m_{\theta,j}(x)]$ in probability under the alternative hypothesis, the proofs of (A.28) and (A.30) are straightforward.

Proof of (A.29). To show the probability limit, we apply Lemma 3.1 of Powell, Stock, and Stoker (1989), which shows that $(n-1)^{-1} h^{-q/2} T_{3,j} = \bar{r}_n + o_p(1)$ if $\mathbb{E}[\|h^{-2q} u_{i,j} u_{l,j} \bar{K}(X_i, X_l)\|^2] = o(n)$, where $\bar{r}_n = \mathbb{E}[h^{-2q} u_{i,j} u_{l,j} \bar{K}(X_i, X_l)]$.

The condition for the application of of Lemma 3.1 holds as follows:

$$
\begin{aligned}
&\mathbb{E}[\|h^{-2q} u_{i,j} u_{l,j} \bar{K}(X_i, X_l)\|^2] \\
&= \frac{1}{h^{4q}} \mathbb{E}\left\{ \left[\int \frac{K\left(\frac{X_i - x}{h}\right) K\left(\frac{X_l - x}{h}\right)}{f_h^2(x)} \pi(x) dx \right]^2 \sigma_j^2(X_i) \sigma_j^2(X_l) \right\} \\
&= h^{-q} K^{(4)}(0) \int \frac{\pi(x)^2 [\sigma_j^2(x)]^2}{[f_h^2(x)]^2} f(x)^2 dx + O(1) \\
&= O(h^{-q}) = O(n(nh^q)^{-1}) = o(n),
\end{aligned}
$$

since $nh^q \to \infty$. Applying Lemma 3.1, we get $\frac{1}{(n-1)h^{q/2}} T_{3,j} = \bar{r}_n + o_p(1)$, where

$$
\begin{aligned}
\bar{r}_n &= \mathbb{E}[h^{-2q} u_{i,j} u_{l,j} \bar{K}(X_i, X_l)] \\
&= h^{-2q} \mathbb{E}[\mathbb{E}(u_{i,j}|X_i)\mathbb{E}(u_{l,j}|X_l)\bar{K}(X_i, X_l)] \\
&= h^{-2q} \int\int \frac{K(\frac{u-x}{h})K(\frac{v-x}{h})}{f_h^2(x)} \pi(x)dx[m_{\theta,j}(u) - m_j(u)][m_{\theta,j}(v) - m_j(v)]f(u)f(v)dudv \\
&= \int K^{(2)}(s)ds \int \frac{\pi(u)}{f_h^2(u)}[m_{\theta,j}(u) - m_j(u)]^2 f(u)^2 du + O(h) \\
&= \int [m_{\theta,j}(u) - m_j(u)]^2 \pi(u)du + o_p(1).
\end{aligned}
$$

The last equality holds because $\int K^{(2)}(s)ds = 1$. ■

Proof of Lemma 2. According to the proof of Proposition 2, we can write $nh^{q/2} Z_{n,j} = Z_{2,j} + o_p(1)$, where $\frac{1}{nh^{q/2}} Z_{2,j}$ is a second-order U-statistic. It is enough to show the following:

$$
\frac{1}{nh^{2/q}} Z_{2,j} = \mathbb{E}\{[m_{\theta,j}(X_i) - m_j(X_i)]^2 f(X_i)\} + o_p(1), \tag{A.31}
$$

$$
\hat{V}_{j,j}^{Zh} = 2K^{(2)}(0)\,\mathbb{E}\{m_{\theta,j}(X_i)^2[1 - m_{\theta,j}(X_i)]^2 f(X_i)\} + o_p(1). \tag{A.32}
$$

Since $\hat{\sigma}_j^2(x) = m_{\hat{\theta},j}(x)[1 - m_{\hat{\theta},j}(x)]$ converges to $m_{\theta,j}(x)[1 - m_{\theta,j}(x)]$ in probability under the alternative hypothesis, the proofs of (A.32) is straightforward.

Proof of (A.32). We show that $n^{-1}h^{-2/q} Z_{2,j}$ satisfies the condition for the application of Lemma 3.1 of Powell et al. (1989):

$$
\begin{aligned}
&\mathbb{E}[\|h^{-q} u_{i,j} u_{l,j} K\left(\frac{X_i - X_l}{h}\right)\|^2] \\
&\quad = h^{-2q} \int K\left(\frac{x-y}{h}\right)^2 \sigma_j^2(x)\sigma_j^2(y)f(x)f(y)dxdy \\
&\quad = h^{-q} \int K(u)^2[\sigma_j^2(x)]^2 f(x)^2 dxdu + O(1) \\
&\quad = O(h^{-q}) = O(n(nh^q)^{-1}) = o(n) \ \text{ since } \ nh^q \to \infty.
\end{aligned}
$$

Applying Lemma 3.1, we obtain $\frac{1}{nh^{q/2}} Z_{2,j} = \bar{r}_n + o_p(1)$, where

$$\begin{aligned}
\bar{r}_n &= \mathbb{E}\left[h^{-q} u_{i,j} u_{l,j} K\left(\frac{X_i - X_l}{h}\right)\right] \\
&= h^{-q} \int K\left(\frac{x-y}{h}\right)[m_{\theta,j}(x) - m_j(x)][m_{\theta,j}(y) - m_j(y)] f(x) f(y) dx dy \\
&= \int K(u)[m_{\theta,j}(x) - m_j(x)]^2 f(x)^2 dx du + O(h) \\
&= \mathbb{E}\{[m_{\theta,j}(X_i) - m_j(X_i)]^2 f(X_i)\} + O(h).
\end{aligned}$$

■

Proof of Lemma 3. Under the local alternative hypothesis, $T_{n,j}$ can be written as follows:

$$\begin{aligned}
T_{n,j} &= nh^{q/2} \int [m_{h,j}(x) - \mathfrak{K}_{h,n} m_{\hat{\theta},j}(X_i)]^2 \pi(x) dx \\
&= nh^{q/2} \int \left\{ \mathfrak{K}_{h,n}[m_j(X_i) - m_{\hat{\theta},j}(X_i) + \delta_n l_j(X_i) + u_{i,j}] \right\}^2 \pi(x) dx \\
&= nh^{q/2} \int \left\{ \mathfrak{K}_{h,n}[m_j(X_i) - m_{\hat{\theta},j}(X_i)] \right\}^2 \pi(x) dx \\
&+ \frac{1}{nh^{3q/2}} \int \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)^2 [u_{i,j} + \delta_n l_j(X_i)]^2 f_h^{-2}(x) \pi(x) dx \\
&+ \frac{1}{nh^{3q/2}} \int \sum_{i=1}^{n} \sum_{l \neq i} K\left(\frac{X_i - x}{h}\right) K\left(\frac{X_l - x}{h}\right) [u_{i,j} + \delta_n l_j(X_i)] \\
&\quad [u_{l,j} + \delta_n l_j(X_l)] f_h^{-2}(x) \pi(x) dx \\
&+ \frac{2}{nh^{3q/2}} \int \sum_{i=1}^{n} \sum_{l=1}^{n} K\left(\frac{X_i - x}{h}\right) K\left(\frac{X_l - x}{h}\right) [m_j(X_i) - m_{\hat{\theta},j}(X_i)] \\
&\quad \delta_n l_j(X_l) f_h^{-2}(x) \pi(x) dx + o_p(1) \\
&= T_{1,j} + T_{2,j} + \frac{1}{nh^{3q/2}} \int \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)^2 [\delta_n l_j(X_i)]^2 f_h^{-2}(x) \pi(x) dx \\
&+ T_{3,j} + \frac{1}{nh^{3q/2}} \int \sum_{i=1}^{n} \sum_{l \neq i} K\left(\frac{X_i - x}{h}\right) K\left(\frac{X_l - x}{h}\right) \delta_n^2 l_j(X_i) l_j(X_l) f_h^{-2}(x) \pi(x) dx
\end{aligned}$$

$$
+\frac{2}{nh^{3q/2}}\int\sum_{i=1}^{n}\sum_{l=1}^{n}K\left(\frac{X_i-x}{h}\right)K\left(\frac{X_l-x}{h}\right)[m_j(X_i)-m_{\hat{\theta},j}(X_i)]
$$

$$
\delta_n l_j(X_l)f_h^{-2}(x)\pi(x)dx+o_p(1)
$$

$$
\equiv T_{1,j}+T_{2,j}+T'_{2,j}+T_{3,j}+T'_{3,j}+T'_{4,j}+o_p(1).
$$

From (A.4) and (A.5), we have $T_{1,j}+T_{2,j}-\hat{\mu}_j^{(HM)}=o_p(1)$ and $T_{3,j}\xrightarrow{d}N(0,V_{j,j}^{(HM)})$. $T'_{2,j}+T'_{3,j}$ can be written as follows:

$$
\begin{aligned}
T'_{2,j}+T'_{3,j}&=nh^{q/2}\delta_n^2\int\frac{\{\sum_{i=1}^{n}K\left(\frac{X_i-x}{h}\right)l_j(X_i)\}^2}{\{\sum_{i=1}^{n}K\left(\frac{X_i-x}{h}\right)\}^2}\pi(x)dx\\
&=c\int[\mathfrak{K}_{h,n}l_j(X_i)]^2\pi(x)dx,
\end{aligned}
$$

where $c\equiv nh^{q/2}\delta_n^2$ is a constant that converges to 1 as $n\to\infty$. Thus,

$$
T'_{2,j}+T'_{3,j}\xrightarrow{p}\int l_j(x)^2\pi(x)dx. \tag{A.33}
$$

Next, we show the asymptotic behavior of $T'_{4,j}$.

$$
\begin{aligned}
T'_{4,j}&=\frac{2}{nh^{3q/2}}\int\sum_{i=1}^{n}\sum_{l=1}^{n}K\left(\frac{X_i-x}{h}\right)K\left(\frac{X_l-x}{h}\right)[m_j(X_i)-m_{\hat{\theta},j}(X_i)]\\
&\quad\delta_n l_j(X_l)f_h^{-2}(x)\pi(x)dx\\
&=2nh^{q/2}\delta_n\int\left\{\sum_{i=1}^{n}K\left(\frac{X_i-x}{h}\right)[m_j(X_i)-m_{\hat{\theta},j}(X_i)]\right\}\\
&\quad\left\{\sum_{l=1}^{n}K\left(\frac{X_l-x}{h}\right)l_j(X_l)\right\}\left[\sum_{i=1}^{n}K\left(\frac{X_i-x}{h}\right)\right]^{-2}\pi(x)dx\\
&=2c\int\mathfrak{K}_{h,n}[m_j(X_i)-m_{\hat{\theta},j}(X_i)]\mathfrak{K}_{h,n}l_j(X_l)\pi(x)dx.
\end{aligned}
$$

Thus,

$$T'_{4,j} \xrightarrow{p} 2\int [m_j(x) - m_{\hat{\theta},j}(x)] l_j(x)\pi(x)dx. \tag{A.34}$$

Finally, it follows from (A.4), (A.5), (A.33), and (A.34) that

$$T_{n,j} - \hat{\mu}_j^{(HM)} \xrightarrow{d} N(M_j, V_{j,j}^{(HM)}) \quad \text{for all } j.$$

■

Proof of Lemma 4. Under the local alternative hypothesis, $nh^{q/2}Z_{n,j}$ can be written as follows:

$$nh^{q/2}Z_{n,j} = \frac{1}{(n-1)h^{q/2}} \sum_{i=1}^{n}\sum_{l\neq i}^{n} K\left(\frac{X_i - X_l}{h}\right) \hat{u}_{\theta,i,j}\hat{u}_{\theta,l,j}$$

$$= \frac{1}{(n-1)h^{q/2}} \sum_{i=1}^{n}\sum_{l\neq 1}^{n} K\left(\frac{X_i - X_l}{h}\right) [m_j(X_i) - m_{\hat{\theta},j}(X_i) + \delta_n l_j(X_i) + u_{i,j}]$$

$$[m_j(X_l) - m_{\hat{\theta},j}(X_l) + \delta_n l_j(X_l) + u_{l,j}]$$

$$= Z_{1,j} + \frac{1}{(n-1)h^{q/2}} \sum_{i=1}^{n}\sum_{l\neq 1}^{n} K\left(\frac{X_i - X_l}{h}\right) [m_j(X_i) - m_{\hat{\theta},j}(X_i)][\delta_n l_j(X_l)]$$

$$+ \frac{1}{(n-1)h^{q/2}} \sum_{i=1}^{n}\sum_{l\neq 1}^{n} K\left(\frac{X_i - X_l}{h}\right) [m_j(X_l) - m_{\hat{\theta},j}(X_l)][\delta_n l_j(X_i)]$$

$$+ \frac{1}{(n-1)h^{q/2}} \sum_{i=1}^{n}\sum_{l\neq 1}^{n} K\left(\frac{X_i - X_l}{h}\right) \delta_n^2 l_j(X_i) l_j(X_l) + Z_{2,j} + o_p(1)$$

$$\equiv Z_{1,j} + Z'_{3,j} + Z''_{3,j} + Z'_{2,j} + Z_{2,j} + o_p(1).$$

From (A.16) and (A.17), we have $Z_{1,j} = o_p(1)$ $Z_{2,j} \xrightarrow{d} N(0, V_{j,j}^{(Zh)})$. Similar to the proof for (A.18) in Proposition 2, it can be straightforwardly shown that $Z'_{3,j} = o_p(1)$ and $Z''_{3,j} = o_p(1)$. $\mathbb{E}(Z'_{2,j})$ can be written as follows:

$$
\begin{aligned}
Z'_{2,j} &= \frac{1}{(n-1)h^{q/2}} \sum_{i=1}^{n} \sum_{l \neq 1}^{n} K\left(\frac{X_i - X_l}{h}\right) \delta_n^2 l_j(X_i) l_j(X_l) \\
&= \frac{1}{h^{q/2}} \sum_{i=1}^{n} \mathbb{E}\left[K\left(\frac{X_i - X_l}{h}\right) \delta_n^2 l_j(X_i) l_j(X_l)\right] + o_p(1) \\
&= \frac{n}{h^{q/2}} \mathbb{E}\left[K\left(\frac{X_1 - X_l}{h}\right) \delta_n^2 l_j(X_1) l_j(X_l)\right] + o_p(1) \\
&= \frac{n}{h^{q/2}} \int K\left(\frac{x-y}{h}\right) \delta_n^2 l_j(x) l_j(y) f(x) f(y) dx dy + o_p(1) \\
&= n h^{q/2} \delta_n^2 \int K(u) l_j(x) l_j(x - uh) f(x) f(x - uh) dx du + o_p(1) \\
&= c \int l_j(x)^2 f(x)^2 dx + O(h) + o_p(1).
\end{aligned}
$$

Therefore, $Z'_{2,j}$ converges to $\mathbb{E}[l_j(x)^2 f(x)] = M'_j$ as $n \to \infty$. ■

Proofs of Propositions 7 and 8. The proofs of Propositions 7 and 8 are on the same lines as Propositions 1 and 2, respectively. The boundedness of $\sigma_j^{*4}(x) \equiv \mathbb{E}[u_{i,j}^{*4} | X_i = x]$ corresponding to (A.9) and (A.24) can be shown straightforwardly because $Y_{i,j}^*$ is a binary variable taking the values zero and one and X lies on a compact set by Assumption 1. ■