Article

A Joint Specification Test for Response Probabilities in Unordered Multinomial Choice Models

Masamune Iwasawa [1,2]

[1] Graduate School of Economics, Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto 606-8501, Japan; E-Mail: iwasawa.masamune.88m@st.kyoto-u.ac.jp; Tel.: +81-75-753-3400

[2] Japan Society for the Promotion of Science, Kojimachi Business Center Building, 5-3-1 Kojimachi, Chiyoda-ku, Tokyo 102-0083, Japan

Academic Editor: Kerry Patterson

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 a test that jointly tests the specifications of multiple response probabilities in unordered multinomial choice models. The test statistic is 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.

Keywords: specification test; multinomial choice models; parametric bootstrap; nonparametric methods

JEL classifications: C12; C25; C15

1. Introduction

Not infrequently, variables of interest in economic research are discrete and unordered, as we often find the variables that indicate the 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 [1,2], respectively, are widely employed, for example, in structural econometric analysis (e.g., the economic models of automobile sales in [3,4]) and as part of econometric methods (e.g., selection bias correction of [5,6]). 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 a new specification test that is 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 test proposed here can be utilized to justify the choice of parametric models and to avoid misspecification problems.

The novelty of the test provided in this study is that it allows 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 has 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.

The test proposed here is based on moment conditions. We show that the test statistic is asymptotically chi-square distributed, consistent against a fixed alternative and able to detect a local alternative approaching the null at the rate of $1/\sqrt{nh^{q/2}}$, where q is the number of independent variables.

One eminent feature of our test is that a parametric bootstrap procedure works well to calculate the rejection region for the test statistic. Since the testing method involves 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 the test statistic. 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 proven by [7]. 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 [8–18], among others. Smoothing-based tests are proposed by [7,19–32], 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. Empirical process-based tests can detect Pitman local alternatives approaching the null at the parametric rate $n^{-1/2}$, whereas 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 the drastic changes of alternatives. The work in [33] shows that smoothing-based tests can detect singular local alternatives at a rate faster than $n^{-1/2}$.

The test proposed in this study is most related to [30], which proposes a smoothing-based test 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 a single response probability that is equal to the conditional expectation of the outcome. For example, [34] 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 size and power. However, applying them to unordered multinomial choice models, as done in this study, is not a trivial task. 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 the problems of parametric specification. The new test statistic is proposed in Section 3. The assumptions and asymptotic behavior are provided in Section 4. Section 5 shows how to bootstrap parametrically. We investigate the size and power of the test 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 are 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 a parameter in a subset of a finite dimensional space Θ. Then, the response probability that i chooses j can be formulated as follows:

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

---

[1] Rate optimal tests are proposed by [35–40], among others.

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 the distributions of ϵ leads 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 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 [2] 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).

Specification of the distribution of ϵ in (1) is less restrictive than specifying a distribution of a random variable. This is because specification of ϵ is true if ϵ is in a family of distributions. The strict inequality in (1) holds after any transformations on both sides of the inequality with any strictly increasing functions. For example, the distribution of $\epsilon_{i,j} - \epsilon_{i,m}$ in (1) could be transformed into a well-known one as the normal or Type I extreme distribution. In these special cases, the distributions of ϵ's may not be an essential specification issue, provided we can specify the right-hand side of the inequality correctly. In other words, distributional assumptions of error terms that help us simplify the estimation of parametric models could be justified by specifying the functional forms of $g_j(\cdot)$ prudently.

In empirical studies, however, functional forms of $g_j(\cdot)$ and distributions of $\epsilon_{i,j}$ are generally unknown for all j. Moreover, in unordered multinomial choice models, the functional forms of $g_j(\cdot)$ and the distributions of $\epsilon_{i,j}$ 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 Statistic

The test statistic proposed in this study is built on the features of response probabilities, that is the 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. Rejection of the null hypothesis, thus, indicates that at least one of the parametric specifications of $g_j(\cdot)$ and $\epsilon_{i,j}$ is misspecified.

---

[2] To be accurate, the MNL model 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 a model with alternative-invariant coefficients.

Before presenting the test statistic, we introduce some notations. Let $f_h(x)$ be the non-parametric density estimator for a continuous point of X_i as follows:

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

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 that of $K^{(2)}$.

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 $u_{\theta,i,j} = Y_{i,j} - m_{\theta,j}(X_i)$ and $f(\cdot)$ is the marginal density of X_i. Under the null hypothesis, $Z_j = 0$, since $\mathbb{E}(u_{\theta,i,j}|X_i) = 0$. Under the alternative hypothesis, $\mathbb{E}[u_{\theta,i,j}\mathbb{E}(u_{\theta,i,j}|X_i)f(X_i)] = \mathbb{E}[\mathbb{E}(u_{\theta,i,j}|X_i)^2 f(X_i)] \geq c\, \mathbb{E}\{[P(Y_{i,j} = 1|X_i) - m_{\theta,j}(X_i)]^2\} > 0$, for some positive constant c, provided that $f(\cdot)$ is bounded away from zero.

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

We 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 J-th response probability from our test statistic. Let $Z_n \equiv (Z_{n,1}, \ldots, Z_{n,J-1})'$ be a $(J-1) \times 1$ vector and $\hat{V}$ a $(J-1) \times (J-1)$ variance-covariance matrix whose (j, m) elements are estimates of $V_{j,m}$. Then, the test statistic is

$$C_n = n^2 h^q Z_n' \hat{V}^{-1} Z_n,$$

where

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

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

for all $j = 1, \ldots, J-1$ and $j \neq m$. $\hat{\sigma}_j^2(\cdot)$ 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 $m_j(x) \equiv P(Y_{i,j} = 1|X_i = x) = \mathbb{E}(Y_{i,j}|X_i = x)$, 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, consistent parametric estimators of $\sigma_j^2(X_i)$ and $\sigma_{j,m}(X_i)$ 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.

4. The Asymptotic Behavior

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

4.1. Assumptions

The following are sufficient assumptions to show the test statistic's 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 zero.

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 any other.

Assumption 4: $m_{\theta,j}(X)$ is differentiable with respect to θ, the derivative $\frac{\partial}{\partial\theta}m_{\theta,j}(X)$ is continuous with respect to X and θ, and $\sup_{\theta\in\Theta}|m_{\theta,j}(x)| < \infty$ for all x.

Assumption 5: There exists a unique value for the θ, defined as $\theta_0 = \arg\max_\theta \sum_{i=1}^n \sum_{j=1}^J \mathbb{1}\{Y_{i,j} = 1\}\log[m_{\theta,j}(X_i)]$. Letting $\theta_0 = \theta$, it satisfies $\hat{\theta} - \theta = O_p(1/\sqrt{n})$.

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 too strong, because it excludes X from 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 zero 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 matrix V is invertible.

We need Assumption 4 to show the asymptotic behavior of C_n. The $\sqrt{n}$-consistency of the parametric estimation given in Assumption 5 is 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 is consistent under Assumptions 1 and 5 (see, for example, Theorem 4.1 of [41]).

4.2. Asymptotic Distribution under the Null Hypothesis

We provide a proposition about the asymptotic distribution of C_n under the null hypothesis. The proof of the proposition is provided in the Appendix.

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

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

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

Proposition 1 indicates that the asymptotic distribution of the test statistic C_n under the null hypothesis is a chi-squared distribution 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 $C_n > t_\alpha$, where t_α is the $(1 - \alpha)$ quantile of the chi-squared distribution with $J - 1$ degrees of freedom.

4.3. Asymptotic Distribution under the Alternative Hypothesis

We show that the test statistic is consistent, that is its asymptotic power is equal to one. The proof of the lemma is provided in the Appendix.

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

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

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

The proof of Lemma 1 provided in the Appendix implies that $nh^{2/q} Z_{n,j}$ diverges for some j as the sample size n increases and $\hat{V}_{j,j}$ converges to a constant that is strictly larger than zero. In addition, it is straightforward to see that the probability limits of $\hat{V}_{j,m}$ under the alternative hypothesis is

$$2K^{(2)}(0)\, \mathbb{E}[m_{\theta,j}(X_i)^2 m_{\theta,m}(X_i)^2 f(X_i)],$$

which is bounded above by Assumptions 1–3 and 5 for any $j \neq m$. Thus, the following proposition follows immediately.

Proposition 2. *Let Assumptions 1–6 hold. Then, under the alternative hypothesis, C_n diverges in probability, and thus, the asymptotic power of the test is one.*

The proof of Proposition 2 is apparent from Lemma 1 and the discussion on the probability limits of $\hat{V}_{j,m}$ under the alternative hypothesis mentioned above.

4.4. Asymptotic Distribution under the Pitman Local Alternative

We show that the test statistic C_n has nontrivial power against Pitman local alternatives approaching the null at the rate of $1/\sqrt{nh^{q/2}}$. The proof of the lemma is 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 $l_j(\cdot)$ is a known continuous function with $\mathbb{E}[l_j(\cdot)^2] < \infty$ for all j and $\delta_n \to 0$ at the rate of $1/\sqrt{nh^{q/2}}$.

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

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

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

Lemma 2 indicates that the limiting distribution of $nh^{q/2} Z_{n,j} / V_{j,j}$ is the normal distribution with mean $M_j V_{j,j}^{-1/2}$ and variance one. The following proposition shows that the test statistic can detect the local alternative with nontrivial power.

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

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

where $\tilde{\lambda} \equiv M'V^{-1}M$ is a noncentrality parameter.

The proof of Proposition 3 is straightforward from Lemma 2 and the discussion on the probability limits of $\hat{V}_{j,m}$ for $j \neq m$ in the proof of Proposition 1.

5. Bootstrap Methods

This section presents a bootstrapping method that is useful in approximating the distribution of the test statistic when the sample size is small. Specification tests for the regression function usually require the wild bootstrapping procedure to calculate the rejection region, as proven by [7]. In our case, however, the wild bootstrap does not work well. The intuitive reason is that it fails to generate a bootstrap sample for the binary response variable.

We show that the parametric bootstrap procedure works well to calculate the rejection region for the test statistic. Intuitively, this is because the binary bootstrap sample for the response variable, say Y_i^*, can be driven according to the parametrically-generated response probabilities, and there are no specific conditions that should be held by Y_i^* in multinomial choice models. This is, for example, different from

the case of the regression model in which the conditional expectation of the error term should be zero. The proof for the proposition in this section is 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 $\{\{X_{i,j}, Y_{i,j}\}_{i=1}^{n}\}_{j=1}^{J}$. For each person, we randomly choose one of J alternatives (say, alternative m_i) with individual 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 $i = 1, \cdots, n$, so that $P(Y_{i,j}^*|X_i) = m_{\hat{\theta},j}(X_i)$, 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 any other.

Assumption 5': There exists a unique value for the θ, defined as $\theta_0 = \arg\max_{\theta} \sum_{i=1}^{n} \sum_{j=1}^{J} \mathbb{1}\{Y_{i,j} = 1\} \log[m_{\theta,j}(X_i)]$. Letting $\theta_0 = \theta$, it satisfies $\hat{\theta}^* - \theta = O_p(1/\sqrt{n})$.

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 3' holds whenever Assumption 3 holds and one applies parametric models whose estimates do not exceed below zero and above one, such as the MNL or MNP model. Assumption 5' requires that $\hat{\theta}^*$ be a consistent estimator of θ. Assumption 5' is satisfied whenever Assumption 5 holds because the true value of $\hat{\theta}^*$ is $\hat{\theta}$, which converges to θ in probability.

Bootstrap Methods for C_n

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

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

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

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

6. Monte Carlo Experiments

The size and power of the test 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 test, 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 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 the power property of the test in the case of misspecification due to nonlinearity and the choice-specific coefficients. We add nonlinearity to the true function of $g_j(\cdot)$ in all true models by setting c_j and d_j at nonzero values. Choice-specific coefficients are inserted into Model 2 by setting γ_j at different values across j. 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 one, if $y_{i,j}^* > y_{i,m}^*$ for all $m \neq j$, and zero otherwise. Sample sizes are $n = 50$ and $n = 100$. The critical value is computed by $B = 100$ repetitions of the parametric bootstrap, and all results are based on $M = 1000$ simulation runs.

To calculate the test statistics, $X_{i,j}$ is 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 $h \in \{0.30, 0.35, 0.40, 0.45\}$.

Table 1 illustrates the size of the test at the 5% significance level. The first and second rows of the table show the size of the test, where the critical values are obtained by the parametric bootstrap ($t_{0.05}^*$) and asymptotic distribution of the test statistic ($t_{0.05}$), respectively. The first to fourth columns of the table illustrate the results obtained with a sample size of $n = 50$ and bandwidths h of 0.30, 0.35, 0.40 and 0.45, respectively. Similarly, the fifth to eighth columns show the result with $n = 100$. Overall, the test tends to over-reject the null hypothesis when the critical values are calculated by parametric bootstrap. The probability of rejection is close to its nominal size when $h = 0.35$ and $n = 50$. In contrast, the test tends to under-reject the null hypothesis when the critical values are the 95% quantile of the chi-squared distribution with two degrees of freedom. The probability of rejection is close to its nominal size when $h = 0.30$ and $n = 50$.

Table 1. Monte Carlo estimates of the size.

Critical Value\h	*n* = 50				*n* = 100			
	0.30	0.35	0.40	0.45	0.30	0.35	0.40	0.45
$t^*_{0.05}$	0.069	0.047	0.062	0.058	0.058	0.064	0.063	0.077
$t_{0.05}$	0.049	0.040	0.042	0.043	0.032	0.051	0.050	0.046

The significance level is 0.05.

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

Table 2 reports the proportion of rejections of the null hypothesis at the 5% significance level. The first to third rows of the table show the power of the test when the true models are Model 1, Model 2 and Model 3, respectively, where the critical values are obtained by parametric bootstrap. Similarly, the fourth to sixth rows of the table show the power, where the critical values are obtained by the 95% quantile of the chi-squared distribution with two degrees of freedom. The first to fourth columns of the table illustrate the power results obtained with a sample size of $n = 50$ and bandwidths h of 0.30, 0.35, 0.40 and 0.45, respectively. Similarly, the fifth to eighth columns show the result with $n = 100$.

The test does not have a decidedly nontrivial power when Model 1 is true. 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 show, the parametric model under the null hypothesis may provide a proper approximation for the response probabilities of Model 1. Therefore, the low power of the test statistic may be acceptable. In contrast, the test statistic has more nontrivial power when Model 2 or 3 is true. The greater the sample size, the better the power performance, which depends on the choice of bandwidth.
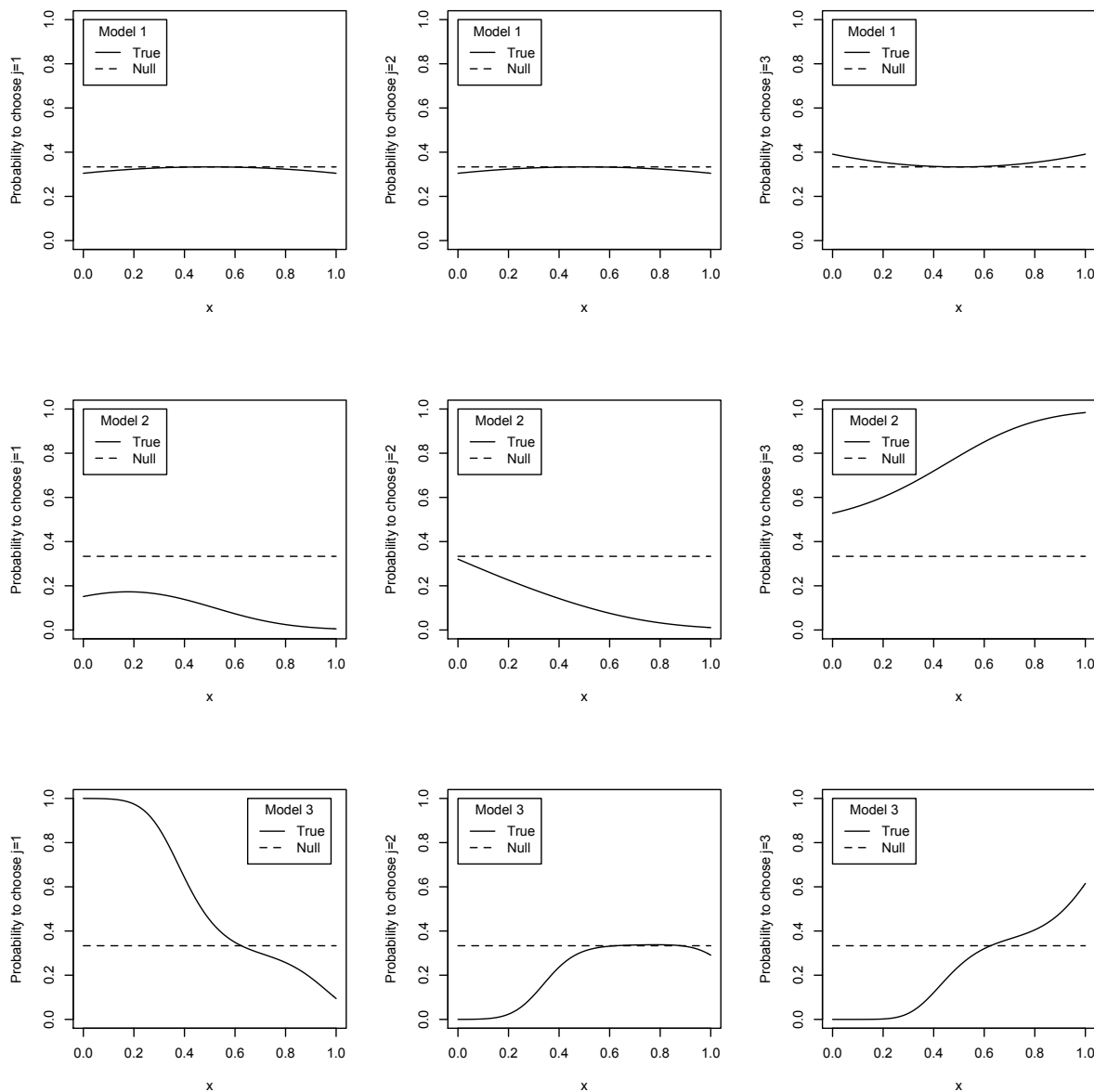
Figure 1. Discrepancy between true and estimated parametric response probabilities.

Table 2. Proportion of null hypothesis rejections based on Monte Carlo simulation.

Critical Value	Model\h	*n* = 50				*n* = 100			
		0.30	**0.35**	**0.40**	**0.45**	**0.30**	**0.35**	**0.40**	**0.45**
$t^*_{0.05}$	Model 1	0.061	0.058	0.055	0.070	0.065	0.053	0.063	0.076
	Model 2	0.810	0.888	0.935	0.960	0.998	1.000	1.000	1.000
	Model 3	0.236	0.330	0.397	0.411	0.672	0.709	0.791	0.838
$t_{0.05}$	Model 1	0.047	0.043	0.056	0.061	0.049	0.053	0.052	0.053
	Model 2	0.807	0.907	0.940	0.970	0.995	1.000	0.999	1.000
	Model 3	0.255	0.304	0.365	0.415	0.600	0.713	0.777	0.839

The significance level is 0.05.

Closer inspection of Table 2 reveals that the test performs better in terms of power when the critical values are obtained by parametric bootstrap, especially when the sample size is $n = 50$. Too see this,

we compare the results of $h = 0.35$ when critical values are $t^*_{0.05}$ with those of $h = 0.30$ when critical values are $t_{0.05}$. We compare the results with different bandwidths because the size of the test is close to its nominal size with these bandwidths (0.047 and 0.049, respectively). When the true model is Model 2, the probability of the rejection of the null hypothesis is 0.888 for $t^*_{0.05}$. The probability of the rejection is 0.807 for $t_{0.05}$. Similarly, when the true model is Model 3, the probability is 0.330 for $t^*_{0.05}$ and 0.255 for $t_{0.05}$. It is surprising that the performance of the test is not unreasonable when critical values are obtained by an asymptotic distribution. However, at least in this setting, the test shows higher power when critical values are obtained by parametric bootstrap when the sample size is small.

7. Conclusions

This study proposes a consistent specification test for unordered multinomial choice models. It tests the specifications of multiple response probabilities jointly for all choice alternatives. The test statistic is 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 of $1/\sqrt{nh^{q/2}}$. 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 test. We find that the test statistic does not have a decidedly nontrivial power when the parametric model under the null hypothesis provides a proper approximation for the response probabilities of the true model. The test statistic has more nontrivial power when the approximation of the null model is less successful. In addition, we find that the test shows higher power performance when critical values are obtained by parametric bootstrap than when they are obtained by the asymptotic distribution of the test statistic. The differences of the power performances are greater when the sample size is small. We can reduce size distortion by choosing an appropriate bandwidth, but this issue remains for future research.

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

Acknowledgments

Conflicts of Interest

The author declares no conflict of interest.

Appendix: Proofs

Proof of Proposition 1. We first prove the following:

$$nh^{q/2}\hat{Z}_{n,j} \xrightarrow{d} N(0, V_{j,j}), \tag{4}$$

$$V_{j,j} - \hat{V}_{j,j} = o_p(1), \tag{5}$$

$$V_{j,m} - \hat{V}_{j,m} = o_p(1), \tag{6}$$

where $V_{j,j}$ and $V_{j,m}$ 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:

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

Proof of (4). Under the null hypothesis, we have $m_j(\cdot) = m_{\theta,j}(\cdot)$. Thus, it follows that

$$
\begin{aligned}
&nh^{q/2}\hat{Z}_{n,j} \\
&= \frac{1}{h^{q/2}(n-1)} \sum_{i=1}^{n} \sum_{l\neq i}^{n} K\left(\frac{X_i - X_l}{h}\right) u_{\hat{\theta},i,j} u_{\hat{\theta},l,j} \\
&= \frac{1}{h^{q/2}(n-1)} \sum_{i=1}^{n} \sum_{l\neq i}^{n} K\left(\frac{X_i - X_l}{h}\right) [m_{\theta,j}(X_i) - m_{\hat{\theta},j}(X_i) + u_{i,j}][m_{\theta,j}(X_l) - m_{\hat{\theta},j}(X_l) + u_{l,j}] \\
&= \frac{1}{h^{q/2}(n-1)} \sum_{i=1}^{n} \sum_{l\neq i}^{n} K\left(\frac{X_i - X_l}{h}\right) [m_{\theta,j}(X_i) - m_{\hat{\theta},j}(X_i)][m_{\theta,j}(X_l) - m_{\hat{\theta},j}(X_l)] \\
&\quad + \frac{1}{h^{q/2}(n-1)} \sum_{i=1}^{n} \sum_{l\neq i}^{n} K\left(\frac{X_i - X_l}{h}\right) u_{i,j} u_{l,j} \\
&\quad + \frac{1}{h^{q/2}(n-1)} \sum_{i=1}^{n} \sum_{l\neq i}^{n} K\left(\frac{X_i - X_l}{h}\right) [m_{\theta,j}(X_i) - m_{\hat{\theta},j}(X_i)] u_{l,j} \\
&\quad + \frac{1}{h^{q/2}(n-1)} \sum_{i=1}^{n} \sum_{l\neq i}^{n} K\left(\frac{X_i - X_l}{h}\right) [m_{\theta,j}(X_l) - m_{\hat{\theta},j}(X_l)] u_{i,j} \\
&\equiv Z_{1,j} + Z_{2,j} + Z_{3,j} + Z_{4,j}.
\end{aligned}
$$

We will prove the following:

$$Z_{1,j} = o_p(1), \tag{7}$$

$$Z_{2,j} \xrightarrow{d} N(0, V_{j,j}), \tag{8}$$

$$Z_{3,j} + Z_{4,j} = o_p(1). \tag{9}$$

Proof of (7). We show that $Z_{1,j} = o_p(1)$. By Assumption 4, there is an interior point $\tilde{\theta}$ between θ and $\hat{\theta}$, such that

$$m_{\hat{\theta},j}(X_i) - m_{\theta,j}(X_i) = \frac{\partial}{\partial\theta'} m_{\tilde{\theta},j}(X_i)(\hat{\theta} - \theta). \tag{10}$$

By using this, $Z_{1,j}$ can be represented as follows:

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

$$= h^{q/2}\sqrt{n}(\hat{\theta} - \theta)' \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i}^{n} \frac{1}{h^q} K\left(\frac{X_i - X_l}{h}\right) \frac{\partial m_{\tilde{\theta},j}(X_i)}{\partial \theta} \frac{\partial m_{\tilde{\theta},j}(X_l)}{\partial \theta'} \sqrt{n}(\hat{\theta} - \theta)$$

$$\equiv h^{q/2}\sqrt{n}(\hat{\theta} - \theta)' \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i}^{n} \bar{Z}_1(X_i, X_l) \sqrt{n}(\hat{\theta} - \theta),$$

where $\bar{Z}_1(X_i, X_l)$ is a symmetric function. We apply Lemma 3.1 of [42] to the second-order U-statistic, $\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i}^{n} \bar{Z}_1(X_i, X_l)$. To do this, we need to show that $\mathbb{E}[\|\bar{Z}_1(X_i, X_l)\|^2] = o(n)$.

$$\mathbb{E}[\|\bar{Z}_1(X_i, X_l)\|^2] = \frac{1}{h^{2q}} \mathbb{E}\left[K\left(\frac{X_i - X_l}{h}\right)^2 \left\|\frac{\partial m_{\tilde{\theta},j}(X_i)}{\partial \theta}\right\|^2 \left\|\frac{\partial m_{\tilde{\theta},j}(X_l)}{\partial \theta}\right\|^2\right]$$

$$= \frac{1}{h^{2q}} \int K\left(\frac{x-y}{h}\right)^2 \left\|\frac{\partial m_{\tilde{\theta},j}(x)}{\partial \theta}\right\|^2 \left\|\frac{\partial m_{\tilde{\theta},j}(y)}{\partial \theta}\right\|^2 f(x)f(y)dxdy$$

$$= \frac{1}{h^q} \int K(u)^2 \left\|\frac{\partial m_{\tilde{\theta},j}(x)}{\partial \theta}\right\|^2 \left\|\frac{\partial m_{\tilde{\theta},j}(x-uh)}{\partial \theta}\right\|^2 f(x)f(x-uh)dxdu$$

$$= \frac{1}{h^q} K^{(2)}(0) \int \left\|\frac{\partial m_{\tilde{\theta},j}(x)}{\partial \theta}\right\|^4 f(x)^2 dx + O(h)$$

$$= O(h^{-q}) + O[n(nh^q)^{-1}] = o(n) \text{ since } nh^q \to \infty. \tag{11}$$

Applying Lemma 3.1 of [42], we obtain $\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i}^{n} \bar{Z}_1(X_i, X_l) = \mathbb{E}[\bar{Z}_1(X_i, X_l)] + o_p(1/\sqrt{n})$, where

$$\mathbb{E}[\bar{Z}_1(X_i, X_l)] = \frac{1}{h^q} \mathbb{E}\left[K\left(\frac{X_i - X_l}{h}\right) \frac{\partial m_{\tilde{\theta},j}(X_i)}{\partial \theta} \frac{\partial m_{\tilde{\theta},j}(X_l)}{\partial \theta'}\right]$$

$$= \frac{1}{h^q} \int K\left(\frac{x-y}{h}\right) \frac{\partial m_{\tilde{\theta},j}(x)}{\partial \theta} \frac{\partial m_{\tilde{\theta},j}(y)}{\partial \theta'} f(x)f(y)dxdy$$

$$= \int K(u) \frac{\partial m_{\tilde{\theta},j}(x)}{\partial \theta} \frac{\partial m_{\tilde{\theta},j}(x-uh)}{\partial \theta'} f(x)f(x-uh)dxdu$$

$$= \int \frac{\partial m_{\tilde{\theta},j}(x)}{\partial \theta} \frac{\partial m_{\tilde{\theta},j}(x)}{\partial \theta'} f(x)^2 dx$$

$$= O(1).$$

Therefore, we yield

$$Z_{1,j} = h^{q/2}\sqrt{n}(\hat{\theta} - \theta)' \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i}^{n} \bar{Z}_1(X_i, X_l) \sqrt{n}(\hat{\theta} - \theta),$$

$$= O_p(h^{q/2}) = o_p(1).$$

Proof of (8). 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 degenerate U-statistics proposed by [43],

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

if

$$\frac{\mathbb{E}[G_n^2(Z_1, Z_2)] + n^{-1}\mathbb{E}\{[u_{1,j} u_{2,j} K\left(\frac{X_1 - X_2}{h}\right)]^4\}}{\mathbb{E}\{[u_{1,j} u_{2,j} K\left(\frac{X_1 - X_2}{h}\right)]^2\}^2} \to 0 \quad \text{as } n \to \infty. \tag{12}$$

Thus, it is enough to show that (12) and

$$\frac{2}{h^q}\mathbb{E}\left\{\left[u_{1,j} u_{2,j} K\left(\frac{X_1 - X_2}{h}\right)\right]^2\right\} \to V_{j,j}, \tag{13}$$

hold.

Proof of (12). First, straightforward calculation gives

$$\mathbb{E}[G_n^2(Z_1, Z_2)] = \mathbb{E}\left[\left\{\mathbb{E}_{Z_i}\left[u_{1,j} u_{2,j} u_{i,j}^2 K\left(\frac{X_1 - X_i}{h}\right) K\left(\frac{X_2 - X_i}{h}\right)\right]\right\}^2\right]$$

$$= \mathbb{E}\left\{\sigma_j^2(X_1)\sigma_j^2(X_2)\left[\int \sigma_j^2(z) K\left(\frac{X_1 - z}{h}\right) K\left(\frac{X_2 - z}{h}\right) f(z)dz\right]^2\right\}$$

$$= h^{3q}K^{(4)}(0)\int [\sigma_j^2(x)]^4 f(x)^4 dx + O(h^{3q+1}) + o(h^{3q+1})$$

$$= O(h^{3q}). \tag{14}$$

Similarly, it can be shown that

$$\frac{1}{n}\mathbb{E}\left\{\left[u_{1,j} u_{2,j} K\left(\frac{X_1 - X_2}{h}\right)\right]^4\right\} = \frac{1}{n}\int \sigma_j^4(x)\sigma_j^4(y)\left[K\left(\frac{x - y}{h}\right)\right]^4 f(x)f(y)dxdy$$

$$= \frac{h^q}{n}\int [\sigma_j^4(x)]^2 f^2(x)dx \int [K(u)]^4 du + O\left(\frac{h^{2q}}{n}\right)$$

$$= O\left(\frac{h^q}{n}\right). \tag{15}$$

Following some calculation, we obtain

$$\mathbb{E}\left\{\left[u_{1,j} u_{2,j} K\left(\frac{X_1 - X_2}{h}\right)\right]^2\right\}^2 = \mathbb{E}\left\{\sigma^2(X_1)\sigma^2(X_2)\left[K\left(\frac{X_1 - X_2}{h}\right)\right]^2\right\}^2$$

$$= h^{2q}\left\{K^{(2)}(0)\int [\sigma^2(x)]^2 f^2(x)dx + O(h)\right\}^2$$

$$= O(h^{2q}). \tag{16}$$

Finally, (14)–(16) indicate that (12) 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 (13). From Equation (16), it is clear that

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

$$= 2K^{(2)}(0)\mathbb{E}\{[\sigma^2(x)]^2 f(x)\} + O(h)$$

$$\to V_{j,j}. \tag{17}$$

Proof of (9). We show that $Z_{3,j} + Z_{4,j} = o_p(1)$. By using (10), $Z_{3,j} + Z_{4,j}$ can be represented as follows:

$$Z_{3,j} + Z_{4,j} = \frac{1}{(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}^{n}\frac{1}{h^{q/2}}K\left(\frac{X_i-X_l}{h}\right)\{[m_{\theta,j}(X_i) - m_{\hat{\theta},j}(X_i)]u_{l,j} + [m_{\theta,j}(X_l) - m_{\hat{\theta},j}(X_l)]u_{i,j}\}$$

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

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

$$\equiv \frac{h^{q/2}}{(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}^{n}\bar{Z}_3(X_i,X_l)(\hat{\theta}-\theta),$$

where $\bar{Z}_3(X_i, X_l)$ is a symmetric function. We apply Lemma 3.1 of [42] to the second-order U-statistic, $\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}^{n}\bar{Z}_3(X_i, X_l)$. To do this, we need to show that: $\mathbb{E}[\|\bar{Z}_3(X_i, X_l)\|^2] = o(n)$.

$$\mathbb{E}[|\bar{Z}_3(X_i,X_l)|^2] \leq \frac{2}{h^{2q}}\mathbb{E}\left\{K\left(\frac{X_i-X_l}{h}\right)^2\left\|\frac{\partial m_{\tilde{\theta},j}(X_i)}{\partial\theta'}\right\|^2 u_{l,j}^2\right\}$$

$$+ \frac{2}{h^{2q}}\mathbb{E}\left\{K\left(\frac{X_i-X_l}{h}\right)^2\left\|\frac{\partial m_{\tilde{\theta},j}(X_l)}{\partial\theta'}\right\|^2 u_{i,j}^2\right\}$$

$$= \frac{2}{h^{2q}}\int K\left(\frac{x-y}{h}\right)^2\left\|\frac{\partial m_{\tilde{\theta},j}(x)}{\partial\theta'}\right\|^2\sigma_j^2(y)f(x)f(y)dxdy$$

$$+ \frac{2}{h^{2q}}\int K\left(\frac{x-y}{h}\right)^2\left\|\frac{\partial m_{\tilde{\theta},j}(y)}{\partial\theta'}\right\|^2\sigma_j^2(x)f(x)f(y)dxdy$$

$$= \frac{2}{h^q}\int K(u)^2\left\|\frac{\partial m_{\tilde{\theta},j}(x)}{\partial\theta'}\right\|^2\sigma_j^2(x-uh)f(x)f(x-uh)dxdu$$

$$+ \frac{2}{h^q}\int K(v)^2\left\|\frac{\partial m_{\tilde{\theta},j}(y)}{\partial\theta'}\right\|^2\sigma_j^2(y+vh)f(y+vh)f(y)dvdy$$

$$= \frac{2}{h^q}K^{(2)}(0)\int\left\|\frac{\partial m_{\tilde{\theta},j}(x)}{\partial\theta'}\right\|^2\sigma_j^2(x)f(x)^2 dx$$

$$+ \frac{2}{h^q}K^{(2)}(0)\int\left\|\frac{\partial m_{\tilde{\theta},j}(y)}{\partial\theta'}\right\|^2\sigma_j^2(y)f(y)^2 dy + O(h)$$

$$= O(h^{-q}) + O[n(nh^q)^{-1}] = o(n) \text{ since } nh^q \to \infty. \tag{18}$$

Applying Lemma 3.1 of [42], we obtain $\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}^{n}\bar{Z}_3(X_i, X_l) = \mathbb{E}[\bar{Z}_3(X_i, X_l)] + o_p(1/\sqrt{n})$, where $\mathbb{E}[\bar{Z}_3(X_i, X_l)] = 0$. Therefore,

$$Z_{3,j} + Z_{4,j} = \frac{nh^{q/2}}{n(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}^{n}\bar{Z}_3(X_i, X_l)(\hat{\theta} - \theta)$$

$$= nh^{q/2}o_p(1/\sqrt{n})O_p(1/\sqrt{n})$$

$$= o_p(h^{q/2}) = o_p(1).$$

Proof of (5) and (6). 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 $V_{j,m}$. From the results of (7)–(9), it is clear that $\mathbb{E}(Z_{2,j}Z_{2,m}) \to V_{j,m}$ 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

$$\mathbb{E}(Z_{2,j}Z_{2,m})$$

$$= \frac{1}{(n-1)^2 h^q}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{l\neq 1}^{n}K\left(\frac{X_i - X_l}{h}\right)u_{i,j}u_{l,j}\sum_{s=1}^{n}\sum_{t\neq s}^{n}K\left(\frac{X_s - X_t}{h}\right)u_{s,m}u_{t,m}\right]$$

$$= \frac{2}{(n-1)^2 h^q}\mathbb{E}\left\{\sum_{i=1}^{n}\sum_{l\neq 1}^{n}u_{i,j}u_{l,j}u_{i,m}u_{l,m}\left[K\left(\frac{X_i - X_l}{h}\right)\right]^2\right\}$$

$$= \frac{2n}{(n-1)h^q}\int\sigma_{j,m}(x)\sigma_{j,m}(y)\left[K\left(\frac{x - y}{h}\right)\right]^2 f(x)f(y)dxdy$$

$$= 2K^{(2)}(0)\int[\sigma_{j,m}(x)]^2 f^2(x)dx + O(h)$$

$$\to V_{j,m}. \tag{19}$$

Thus, the proofs of (5) and (6) are straightforward from (17) and (19).

Let $Z_2 = (Z_{2,1}, Z_{2,2}, \ldots, Z_{2,J-1})'$. Similarly to the proof of (8), it can be straightforwardly shown that $t'Z_2 \overset{d}{\to} N(0, t'Vt)$ for any $(J-1) \times 1$ vector t, where V is a $(J-1) \times (J-1)$ variance-covariance matrix whose (j, m) elements are $V_{j,m}$. Then, by the Cramér–Wold device, Z_2 converges to a multivariate normal distribution with $(J-1) \times 1$ mean vector consists of zero and variance-covariance matrix V. Therefore, C_n, which is the quadratic form of $nh^{q/2}\hat{Z}_{n,j}$, converges to a chi-squared distribution with $J-1$ degrees of freedom. $\square$

Proof of Lemma 1. Under the alternative hypothesis, $nh^{q/2}Z_{n,j}$ can be represented as follows:

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\equiv A_{1,j} + A_{2,j} + A_{3,j} + A_{4,j} + A_{5,j} + A_{6,j} + A_{7,j} + A_{8,j} + A_{9,j},$$

where $A_{5,j} = Z_{1,j} = o_p(1)$ and $A_{6,j} + A_{8,j} = Z_{3,j} + Z_{4,j} = o_p(1)$. We show that

$$\frac{1}{nh^{2/q}}(A_{2,j} + A_{3,j} + A_{4,j} + A_{7,j} + A_{9,j}) = o_p(1). \tag{20}$$

Then, $\hat{Z}_{n,j} = \frac{1}{nh^{2/q}} A_{1,j} + o_p(1)$. Thus, it is enough to show that (20), and the following holds:

$$\frac{1}{nh^{2/q}} A_{1,j} = \mathbb{E}\{[m_{\theta,j}(X_i) - m_j(X_i)]^2 f(X_i)\} + o_p(1), \tag{21}$$

$$\hat{V}_{j,j}^{Zh} = 2K^{(2)}(0)\,\mathbb{E}\{m_{\theta,j}(X_i)^2[1 - m_{\theta,j}(X_i)]^2 f(X_i)\} + o_p(1). \tag{22}$$

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 a probability under the alternative hypothesis, the proofs of (22) is straightforward.

Proof of (20). First, we show $\frac{1}{nh^{2/q}}(A_{2,j} + A_{4,j}) = o_p(1)$. $\frac{1}{nh^{2/q}}(A_{2,j} + A_{4,j})$ can be represented as a second-order U-statistic as follows:

$$\frac{1}{nh^{2/q}}(A_{2,j} + A_{4,j})$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i}^{n} \frac{1}{h^q} K\left(\frac{X_i - X_l}{h}\right)$$

$$\left\{ [m_j(X_i) - m_{\theta,j}(X_i)][m_{\theta,j}(X_l) - m_{\hat{\theta},j}(X_l)] + [m_{\theta,j}(X_i) - m_{\hat{\theta},j}(X_i)][m_j(X_l) - m_{\theta,j}(X_l)] \right\}$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i}^{n} \frac{1}{h^q} K\left(\frac{X_i - X_l}{h}\right)$$

$$\left\{ [m_j(X_i) - m_{\theta,j}(X_i)] \frac{\partial}{\partial \theta'} m_{\tilde{\theta},j}(X_l) + [m_j(X_l) - m_{\theta,j}(X_l)] \frac{\partial}{\partial \theta'} m_{\tilde{\theta},j}(X_i) \right\} (\hat{\theta} - \theta)$$

$$\equiv \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i}^{n} \bar{A}_2(X_i, X_l)(\hat{\theta} - \theta),$$

where $\bar{A}_2(X_i, X_l)$ is a symmetric function. We show that $\mathbb{E}[\|\bar{A}_2(X_i, X_l)\|^2] = o(n)$.

$$\mathbb{E}[\|\bar{A}_2(X_i, X_l)\|^2] \leq \frac{2}{h^{2q}} \mathbb{E}\left[K\left(\frac{X_i - X_l}{h}\right)^2 |m_j(X_i) - m_{\theta,j}(X_i)|^2 \left\|\frac{\partial}{\partial \theta'} m_{\tilde{\theta},j}(X_l)\right\|^2 \right]$$

$$+ \frac{2}{h^{2q}} \mathbb{E}\left[K\left(\frac{X_i - X_l}{h}\right)^2 |m_j(X_l) - m_{\theta,j}(X_l)|^2 \left\|\frac{\partial}{\partial \theta'} m_{\tilde{\theta},j}(X_i)\right\|^2 \right]$$

$$= \frac{2}{h^{2q}} \int K\left(\frac{x - y}{h}\right)^2 |m_j(x) - m_{\theta,j}(x)|^2 \left\|\frac{\partial}{\partial \theta'} m_{\tilde{\theta},j}(y)\right\|^2 f(x)f(y)dxdy$$

$$+ \frac{2}{h^{2q}} \int K\left(\frac{x - y}{h}\right)^2 |m_j(y) - m_{\theta,j}(y)|^2 \left\|\frac{\partial}{\partial \theta'} m_{\tilde{\theta},j}(x)\right\|^2 f(x)f(y)dxdy$$

$$= \frac{2}{h^q} \int K(u)^2 |m_j(x) - m_{\theta,j}(x)|^2 \left\|\frac{\partial}{\partial \theta'} m_{\tilde{\theta},j}(x - uh)\right\|^2 f(x)f(x - uh)dxdu$$

$$+ \frac{2}{h^q} \int K(v)^2 |m_j(y) - m_{\theta,j}(y)|^2 \left\|\frac{\partial}{\partial \theta'} m_{\tilde{\theta},j}(y + vh)\right\|^2 f(y + vh)f(y)dvdy$$

$$= \frac{2}{h^q} K^{(2)}(0) \int |m_j(x) - m_{\theta,j}(x)|^2 \left\|\frac{\partial}{\partial \theta'} m_{\tilde{\theta},j}(x)\right\|^2 f(x)^2 dx$$

$$+ \frac{2}{h^q} K^{(2)}(0) \int |m_j(y) - m_{\theta,j}(y)|^2 \left\|\frac{\partial}{\partial \theta'} m_{\tilde{\theta},j}(y)\right\|^2 f(y)^2 dy + O(h)$$

$$= O(h^{-q}) + O[n(nh^q)^{-1}] = o(n) \text{ since } nh^q \to \infty.$$

Applying Lemma 3.1 of [42], we obtain $\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i}^{n} \bar{A}_2(X_i, X_l) = \mathbb{E}[\bar{A}_2(X_i, X_l)] + o_p(1)$, where

$$\mathbb{E}[\bar{A}_2(X_i, X_l)] = \frac{1}{h^q} \int K\left(\frac{x - y}{h}\right) [m_j(x) - m_{\theta,j}(x)] \frac{\partial m_{\tilde{\theta},j}(y)}{\partial \theta'} f(x)f(y)dxdy$$

$$+ \frac{1}{h^q} \int K\left(\frac{x - y}{h}\right) [m_j(y) - m_{\theta,j}(y)] \frac{\partial m_{\tilde{\theta},j}(x)}{\partial \theta'} f(x)f(y)dxdy$$

$$= \int K(u)[m_j(x) - m_{\theta,j}(x)] \frac{\partial m_{\tilde{\theta},j}(x - uh)}{\partial \theta'} f(x)f(x - uh)dxdu$$

$$+ \int K(v)[m_j(y) - m_{\theta,j}(y)] \frac{\partial m_{\tilde{\theta},j}(y + vh)}{\partial \theta'} f(y + vh)f(y)dvdy$$

$$= \int [m_j(x) - m_{\theta,j}(x)] \frac{\partial m_{\tilde{\theta},j}(x)}{\partial \theta'} f(x)^2 dx$$

$$+ \int [m_j(y) - m_{\theta,j}(y)] \frac{\partial m_{\tilde{\theta},j}(y)}{\partial \theta'} f(y)^2 dy + O(h)$$

$$= O(1).$$

Therefore, we yield

$$\frac{1}{nh^{2/q}}(A_{2,j} + A_{4,j}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l\neq i}^{n} \bar{A}_2(X_i, X_l)(\hat{\theta} - \theta) = O_p(1/\sqrt{n}) = o_p(1).$$

Next, we show $\frac{1}{nh^{2/q}}(A_{3,j} + A_{7,j}) = o_p(1)$. $\frac{1}{nh^{2/q}}(A_{3,j} + A_{7,j})$ can be represented as a second-order U-statistic as follows:

$$\frac{A_{3,j} + A_{7,j}}{nh^{2/q}} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l\neq i}^{n} \frac{1}{h^q} K\left(\frac{X_i - X_l}{h}\right) \{[m_j(X_i) - m_{\theta,j}(X_i)]u_{l,j} + [m_j(X_l) - m_{\theta,j}(X_l)]u_{i,j}\}$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l\neq i}^{n} \bar{A}_3(X_i, X_j),$$

where $\bar{A}_3(X_i, X_j)$ is a symmetric function. We show that $\mathbb{E}[\|\bar{A}_3(X_i, X_l)\|^2] = o(n)$.

$$\mathbb{E}[\|\bar{A}_3(X_i, X_l)\|^2] \leq \frac{2}{h^{2q}} \mathbb{E}\left\{ K\left(\frac{X_i - X_l}{h}\right)^2 |m_j(X_i) - m_{\theta,j}(X_i)|^2 u_{l,j}^2 \right\}$$

$$+ \frac{2}{h^{2q}} \mathbb{E}\left\{ K\left(\frac{X_i - X_l}{h}\right)^2 |m_j(X_l) - m_{\theta,j}(X_l)|^2 u_{i,j}^2 \right\}$$

$$= \frac{2}{h^{2q}} \int K\left(\frac{x-y}{h}\right)^2 |m_j(x) - m_{\theta,j}(x)|^2 \sigma_j^2(y) f(x) f(y) dx dy$$

$$+ \frac{2}{h^{2q}} \int K\left(\frac{x-y}{h}\right)^2 |m_j(y) - m_{\theta,j}(y)|^2 \sigma_j^2(x) f(x) f(y) dx dy$$

$$= \frac{2}{h^q} \int K(u)^2 |m_j(x) - m_{\theta,j}(x)|^2 \sigma_j^2(x - uh) f(x) f(x - uh) dx du$$

$$+ \frac{2}{h^q} \int K(v)^2 |m_j(y) - m_{\theta,j}(y)|^2 \sigma_j^2(y + vh) f(y + vh) f(y) dv dy$$

$$= \frac{2}{h^q} K^{(2)}(0) \int |m_j(x) - m_{\theta,j}(x)|^2 \sigma_j^2(x) f(x)^2 dx$$

$$+ \frac{2}{h^q} K^{(2)}(0) \int |m_j(y) - m_{\theta,j}(y)|^2 \sigma_j^2(y) f(y)^2 dy + O(h)$$

$$= O(h^{-q}) + O[n(nh^q)^{-1}] = o(n) \text{ since } nh^q \to \infty.$$

Applying Lemma 3.1 of [42], we obtain $\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l\neq i}^{n} \bar{A}_3(X_i, X_l) = \mathbb{E}[\bar{A}_3(X_i, X_l)] + o_p(1)$, where $\mathbb{E}[\bar{A}_3(X_i, X_l)] = 0$. Therefore, we yield

$$\frac{A_{3,j} + A_{7,j}}{nh^{2/q}} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l\neq i}^{n} \bar{A}_3(X_i, X_j) = o_p(1).$$

Finally, we show that $n^{-1} h^{-2/q} A_{9,j} = o_p(1)$. It is clear that $\frac{1}{nh^{2/q}} A_{9,j}$ is a second-order U-statistic. This satisfies the condition for Lemma 3.1 of [42] as follows:

$$\mathbb{E}\left[\left\|\frac{1}{h^q} K\left(\frac{X_i - X_l}{h}\right) u_{i,j} u_{l,j}\right\|^2\right] = \frac{1}{h^{2q}} \int K\left(\frac{x-y}{h}\right)^2 \sigma_j^2(x) \sigma_j^2(y) f(x) f(y) dx dy$$

$$= \frac{1}{h^q} \int K(u)^2 [\sigma_j^2(x)]^2 f(x)^2 dx du + O(h)$$

$$= O(h^{-q}) = O(n(nh^q)^{-1}) = o(n) \text{ since } nh^q \to \infty.$$

Applying Lemma 3.1 of [42], we obtain $\frac{1}{nh^{q/2}} A_{9,j} = \mathbb{E}\left[h^{-q} u_{i,j} u_{l,j} K\left(\frac{X_i - X_l}{h}\right)\right] + o_p(1)$, where $\mathbb{E}\left[h^{-q} u_{i,j} u_{l,j} K\left(\frac{X_i - X_l}{h}\right)\right] = 0$.

Proof of (21). $\frac{1}{nh^{2/q}} A_{1,j}$ can be represented as a second-order U-statistic as follows:

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

$$\equiv \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i}^{n} \bar{A}_1(X_i, X_l),$$

where $\bar{A}_1(X_i, X_l)$ is a symmetric function. Similar to (11), it can be straightforwardly shown that $\mathbb{E}[\|\bar{A}_1(X_i, X_l)\|^2] = o(n)$. The only difference from (11) is we have $\|m_j(x) - m_{\theta,j}(x)\|^2$, which is uniformly bounded, as a part of the integrand instead of $\|\partial m_{\tilde{\theta},j}(x)/\partial\theta\|^2$. By applying Lemma 3.1 of [42], we yield $\frac{1}{nh^{2/q}} A_{1,j} = \mathbb{E}[\bar{A}_1(X_i, X_l)] + o_p(1)$, where

$$\mathbb{E}[\bar{A}_1(X_i, X_l)] = \frac{1}{h^q} \int K\left(\frac{x-y}{h}\right) [m_j(x) - m_{\theta,j}(x)][m_j(y) - m_{\theta,j}(y)] f(x) f(y) dx dy$$

$$= \int K(u)[m_j(x) - m_{\theta,j}(x)][m_j(x - uh) - m_{\theta,j}(x - uh)] f(x) f(x - uh) dx du$$

$$= \int [m_j(x) - m_{\theta,j}(x)]^2 f(x)^2 dx + O(h)$$

$$= \mathbb{E}\{[m_{\theta,j}(X_i) - m_j(X_i)]^2 f(X_i)\} + O(h).$$

$\square$

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\equiv B_{1,j} + B_{2,j} + B_{3,j} + B_{4,j} + B_{5,j} + B_{6,j} + B_{7,j} + B_{8,j} + B_{9,j},$$

where $B_{1,j} = Z_{1,j} = o_p(1)$, $B_{4,j} + B_{5,j} = Z_{3,j} + Z_{4,j} = o_p(1)$ and $B_{9,j} = Z_{2,j} \xrightarrow{d} N(0, V_{j,j})$. It suffices to show the following:

$$B_{2,j} + B_{3,j} = o_p(1), \tag{23}$$

$$B_{6,j} \xrightarrow{p} \mathbb{E}[l_j(x)^2 f(x)], \tag{24}$$

$$B_{7,j} + B_{8,j} = o_p(1). \tag{25}$$

Proof of (23). We show that $B_{2,j} + B_{3,j} = o_p(1)$. $B_{2,j} + B_{3,j}$ can be represented as follows:

$$B_{2,j} + B_{3,j}$$

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

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

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

$$\equiv \frac{nh^{q/2}}{n(n-1)} \sum_{i=1}^{n} \sum_{l\neq 1}^{n} \bar{B}_2(X_i, X_l)(\hat{\theta} - \theta)\delta_n,$$

where $\bar{B}_2(X_i, X_l)$ is a symmetric function. Similar to (18), it can be straightforwardly shown that $\mathbb{E}[|\bar{B}_2(X_i, X_l)|^2] = o(n)$. The only difference from (18) is that we have $l_j(\cdot)^2$ instead of $\sigma_j^2(\cdot)$ as a part of the integrand, where $\mathbb{E}[l_j(\cdot)^2]$ is assumed to be bounded. By applying Lemma 3.1 of [42], we yield $\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l\neq 1}^{n} \bar{B}_2(X_i, X_l) = \mathbb{E}[\bar{B}_2(X_i, X_l)] + o_p(1)$, where

$$\mathbb{E}[\bar{B}_2(X_i, X_l)] = \frac{1}{h^q} \int K\left(\frac{x-y}{h}\right) \left[\frac{\partial m_{\tilde{\theta},j}(x)}{\partial\theta'}l_j(y) + \frac{\partial m_{\tilde{\theta},j}(y)}{\partial\theta'}l_j(x)\right] f(x)f(y)dxdy$$

$$= \int K(u) \left[\frac{\partial m_{\tilde{\theta},j}(x)}{\partial\theta'}l_j(x-uh) + \frac{\partial m_{\tilde{\theta},j}(x-uh)}{\partial\theta'}l_j(x)\right] f(x)f(x-uh)dxdu$$

$$= 2 \int \frac{\partial m_{\tilde{\theta},j}(x)}{\partial \theta'} l_j(x) f(x)^2 dx + O(h)$$

$$= O(1).$$

Therefore,

$$B_{2,j} + B_{3,j} = \frac{nh^{q/2}}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq 1}^{n} \bar{B}_2(X_i, X_l)(\hat{\theta} - \theta)\delta_n$$

$$= nh^{q/2} O(1) O_p(1/\sqrt{n}) O(1/\sqrt{nh^{q/2}})$$

$$= O_p(\sqrt{h^{q/2}}) = o_p(1).$$

Proof of (24). We show that $B_{6,j}$ converges to $\mathbb{E}[l_j(x)^2 f(x)]$ as $n \to \infty$. $B_{6,j}$ can be represented as follows:

$$B_{6,j} = \frac{nh^{q/2}}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq 1}^{n} \frac{1}{h^q} K\left(\frac{X_i - X_l}{h}\right) l_j(X_i) l_j(X_l) \delta_n^2$$

$$\equiv \frac{nh^{q/2}}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq 1}^{n} \bar{B}_6(X_i, X_l) \delta_n^2,$$

where $\bar{B}_6(X_i, X_l)$ is a symmetric function. Similar to (11), it can be straightforwardly shown that $\mathbb{E}[|\bar{B}_6(X_i, X_l)|^2] = o(n)$. The only difference from (11) is that we have $l_j(\cdot)^2$ instead of $\left\|\partial m_{\tilde{\theta},j}(X_i)/\partial \theta\right\|^2$ as a part of the integrand, where $\mathbb{E}[l_j(\cdot)^2]$ is assumed to be bounded. By applying Lemma 3.1 of [42], we yield $\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq 1}^{n} \bar{B}_6(X_i, X_l) = \mathbb{E}[\bar{B}_6(X_i, X_l)] + o_p(1)$, where

$$\mathbb{E}[\bar{B}_6(X_i, X_l)] = \frac{1}{h^q} \int K\left(\frac{x-y}{h}\right) l_j(y) l_j(x) f(x) f(y) dx dy$$

$$= \int K(u) l_j(x - uh) l_j(x) f(x) f(x - uh) dx du$$

$$= \int l_j^2(x) f(x)^2 dx + O(h).$$

Therefore,

$$B_{6,j} = \frac{nh^{q/2}}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq 1}^{n} \bar{B}_6(X_i, X_l) \delta_n^2$$

$$= nh^{q/2} \{\mathbb{E}[l_j^2(x) f(x)] + o_p(1)\} \delta_n^2 \xrightarrow{p} \mathbb{E}[l_j^2(x) f(x)].$$

Proof of (25). We show that $B_{7,j} + B_{8,j} = o_p(1)$. $B_{7,j} + B_{8,j}$ can be represented as follows:

$$B_{7,j} + B_{8,j} = \frac{nh^{q/2}}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq 1}^{n} \frac{1}{h^q} K\left(\frac{X_i - X_l}{h}\right) [l_j(X_i) u_{l,j} + l_j(X_l) u_{i,j}] \delta_n$$

$$\equiv \frac{nh^{q/2}}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq 1}^{n} \bar{B}_7(X_i, X_l) \delta_n,$$

where $\bar{B}_7(X_i, X_l)$ is a symmetric function. Similar to (18), it can be straightforwardly shown that $\mathbb{E}[|\bar{B}_7(X_i, X_l)|^2] = o(n)$. The only difference from (18) is that we have $l_j(\cdot)^2$ instead of

$\left\| \partial m_{\tilde{\theta},j}(X_i)/\partial\theta \right\|^2$ as a part of the integrand, where $\mathbb{E}[l_j(\cdot)^2]$ is assumed to be bounded. By applying Lemma 3.1 of [42], we yield $\frac{1}{n(n-1)}\sum_{i=1}^n \sum_{l\neq 1}^n \bar{B}_7(X_i, X_l) = \mathbb{E}[\bar{B}_7(X_i, X_l)] + o_p(1/\sqrt{n})$, where $\mathbb{E}[\bar{B}_7(X_i, X_l)] = 0$. Therefore,

$$B_{7,j} + B_{8,j} = \frac{nh^{q/2}}{n(n-1)}\sum_{i=1}^n \sum_{l\neq 1}^n \bar{B}_7(X_i, X_l)\delta_n$$
$$= nh^{q/2}o_p(1/\sqrt{n})\delta_n = o_p(1).$$

□

Proofs of Proposition 4. Proposition 4 can be proven along the same lines as Proposition 1. Let $u_{i,j}^* = Y_{i,j}^* - m_j^*(X_i)$, where $m_j^*(X_i) \equiv \mathbb{E}(Y_{i,j}^*|X_i) = m_{\hat{\theta},j}(X_i)$, and therefore, $\mathbb{E}(u_{i,j}^*|X_i) = 0$. Then, the boundedness of $\sigma_j^{*4}(x) \equiv \mathbb{E}[u_{i,j}^{*4}|X_i = x]$ corresponding to (15) 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.

We first prove the following:

$$nh^{q/2}\hat{Z}_{n,j}^* \xrightarrow{d} N(0, V_{j,j}^*), \tag{26}$$
$$V_{j,j}^* - \hat{V}_{j,j}^* = o_p(1), \tag{27}$$
$$V_{j,m}^* - \hat{V}_{j,m}^* = o_p(1), \tag{28}$$

where $V_{j,j}^*$ and $V_{j,m}^*$ are the asymptotic variance of $nh^{q/2}\hat{Z}_{n,j}^*$ and 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:

$$V_{j,j}^* \equiv 2K^{(2)}(0)\mathbb{E}\{[\sigma_j^{*2}(x)]^2 f(x)dx\},$$
$$V_{j,m}^* \equiv 2K^{(2)}(0)\mathbb{E}\{[\sigma_{j,m}^*(x)]^2 f(x)dx\},$$

where $\sigma_j^{*2}(x)$ is the conditional variance of $u_{i,j}^*$ and $\sigma_{j,m}^*(x)$ is the covariance between $u_{i,j}^*$ and $u_{i,m}^*$.

Proof of (26). Let $u_{\hat{\theta},i,j}^* = Y_{i,j}^* - m_{\hat{\theta}^*,j}(X_i)$ and $Y_{i,j}^* = m_{\hat{\theta},j}(X_i) + u_{i,j}^*$, where $E(u_{i,j}^*|X_i) = 0$ by definition. Then,

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

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

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

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

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

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

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

We will prove that

$$Z_{1,j}^* = o_p(1), \tag{29}$$

$$Z_{2,j}^* \xrightarrow{d} N(0, V_{j,j}^*), \tag{30}$$

$$Z_{3,j}^* + Z_{4,j}^* = o_p(1). \tag{31}$$

Proof of (29). $Z_{1,j}^*$ can be represented as follows:

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

$$[m_{\hat{\theta},j}(X_l) - m_{\theta,j}(X_l) + m_{\theta,j}(X_l) - m_{\hat{\theta}^*,j}(X_l)]$$

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

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

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

$$\equiv Z_{1,j}^{*\prime} + Z_{1,j}^{*\prime\prime} + Z_{1,j}^{*\prime\prime\prime},$$

where $Z_{1,j}^{*\prime} = Z_{1,j} = o_p(1)$. By Assumption 4, there is an interior point $\tilde{\theta}^*$ between θ and $\hat{\theta}^*$, such that

$$m_{\hat{\theta}^*,j}(X_i) - m_{\theta,j}(X_i) = \frac{\partial}{\partial \theta'} m_{\tilde{\theta}^*,j}(X_i)(\hat{\theta}^* - \theta). \tag{32}$$

Therefore, $Z_{1,j}^{*\prime\prime} = o_p(1)$ and $Z_{1,j}^{*\prime\prime\prime} = o_p(1)$ can be also shown similar to the proof for $Z_{1,j} = o_p(1)$ by using the above mean value theorem instead of (10), because $\theta - \hat{\theta}^* = O_p(1/\sqrt{n})$ for all j under appropriate parametric models and Assumption 5.

Proof of (30). It is clear that $Z_{2,j}^*$ can be treated as second order degenerate U-statistic as follows:

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

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 limited theorem for degenerated U-statistics proposed by [43],

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

if

$$\frac{\mathbb{E}[G_n^{*2}(Z_1^*, Z_2^*)] + n^{-1}\mathbb{E}\{[u_{1,j} u_{2,j} K((X_1 - X_2)/h)]^4\}}{\mathbb{E}\{[u_{1,j}^* u_{2,j}^* K((X_1 - X_2)/h)]^2\}^2} \to 0 \quad \text{as } n \to \infty. \tag{33}$$

Thus, it suffices to show that (33) and

$$2h^{-q}\mathbb{E}\{[u_{1,j}^* u_{2,j}^* K((X_1 - X_2)/h)]^2\} \to V_{j,j}^*, \tag{34}$$

hold.

Proof of (33). First, note that

$$\mathbb{E}[G_n^{*2}(Z_1, Z_2)] = \mathbb{E}\left[\left\{\mathbb{E}_{Z_i}\left[u_{1,j}^* u_{2,j}^* u_{i,j}^{*2} K\left(\frac{X_1 - X_i}{h}\right) K\left(\frac{X_2 - X_i}{h}\right)\right]\right\}^2\right]$$

$$= \mathbb{E}\left\{\sigma_j^{*2}(X_1)\sigma_j^{*2}(X_2)\left[\int \sigma_j^{*2}(z) K\left(\frac{X_1 - z}{h}\right) K\left(\frac{X_2 - z}{h}\right) f(z)dz\right]^2\right\}$$

$$= h^{3q} K^{(4)}(0) \int [\sigma_j^{*2}(x)]^4 f(x)^4 dx + O(h^{3q+1}) + o(h^{3q+1})$$

$$= O(h^{3q}). \tag{35}$$

In the same way as above, we obtain

$$n^{-1}\mathbb{E}\left\{\left[u_{1,j}^* u_{2,j}^* K\left(\frac{X_1 - X_2}{h}\right)\right]^4\right\} = n^{-1}\int \sigma_j^{*4}(x)\sigma_j^{*4}(y)\left[K\left(\frac{x-y}{h}\right)\right]^4 f(x)f(y)dxdy$$

$$= n^{-1}h^q \int [\sigma_j^{*4}(x)]^2 f^2(x)dx \int [K(u)]^4 du + O\left(\frac{h^{2q}}{n}\right)$$

$$= O\left(\frac{h^q}{n}\right). \tag{36}$$

Following some calculation, we can obtain

$$\mathbb{E}\left\{\left[u_{1,j}^* u_{2,j}^* K\left(\frac{X_1 - X_2}{h}\right)\right]^2\right\}^2 = \mathbb{E}\left\{\sigma^{*2}(X_1)\sigma^{*2}(X_2)\left[K\left(\frac{X_1 - X_2}{h}\right)\right]^2\right\}^2$$

$$= h^{2q}\left\{K^{(2)}(0)\int [\sigma^{*2}(x)]^2 f^2(x)dx + O(h)\right\}^2$$

$$= O(h^{2q}). \tag{37}$$

Thus, (33) holds by (35)–(37) because $\frac{O(h^{3q}) + O\left(\frac{h^q}{n}\right)}{O(h^{2q})} = \frac{O(h^q) + O\left(\frac{1}{nh^q}\right)}{O(1)} \to 0$, as $h \to 0$ and $nh^q \to \infty$.

Proof of (34). From Equation (37), it is clear that

$$\frac{2}{h^q}\mathbb{E}\left\{\left[u_{1,j}^* u_{2,j}^* K\left(\frac{X_1 - X_2}{h}\right)\right]^2\right\} = 2K^{(2)}(0)\int [\sigma^{*2}(x)]^2 f^2(x)dx + O(h)$$

$$= 2K^{(2)}(0)\mathbb{E}\{[\sigma^{*2}(x)]^2 f(x)\} + O(h)$$

$$\to V_{j,j}^*. \tag{38}$$

Thus, we have $Z_{2,j}^* \xrightarrow{d} N(0, V_{j,j}^*)$.

Proof of (31). $Z_{3,j}^* + Z_{4,j}^*$ can be represented as follows:

$$Z_{3,j}^* + Z_{4,j}^* = \frac{1}{(n-1)h^{q/2}}\sum_{i=1}^{n}\sum_{l\neq i}^{n} K\left(\frac{X_i - X_l}{h}\right)[m_{\hat{\theta},j}(X_i) - m_{\hat{\theta}^*,j}(X_i)]u_{l,j}^*$$

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

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

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

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

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

$$\equiv Z_{3,j}^{*'} + Z_{3,j}^{*''} + Z_{4,j}^{*'} + Z_{4,j}^{*''}.$$

The only difference between $Z_{3,j}^{*'} + Z_{4,j}^{*'}$ and $Z_{3,j} + Z_{4,j}$ in (9) is that the former contains $u_{l,j}^*$ and $u_{i,j}^*$ instead of $u_{l,j}$ and $u_{i,j}$, respectively. However, since $\mathbb{E}(u_{i,j}^*|X_i) = 0$ and $\mathbb{E}(u_{l,j}^*|X_l) = 0$ from the definition, $Z_{3,j}^{*'} + Z_{4,j}^{*'} = o_p(1)$ can be proven as with the proof of (9). Moreover, $Z_{3,j}^{*''} + Z_{4,j}^{*''} = o_p(1)$ can also be proven as with the proof of (9) by using (32) instead of (10).

Proof of (27) and (28). 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 $V_{j,m}^*$. From the results of (29)–(31), it is clear that $\mathbb{E}(Z_{2,j}^* Z_{2,m}^*) \to V_{j,m}^*$ 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

$$\mathbb{E}(Z_{2,j}^* Z_{2,m}^*) = \frac{1}{(n-1)^2 h^q} \mathbb{E}\left[\sum_{i=1}^{n} \sum_{l \neq 1}^{n} K\left(\frac{X_i - X_l}{h}\right) u_{i,j}^* u_{l,j}^* \sum_{s=1}^{n} \sum_{t \neq s}^{n} K\left(\frac{X_s - X_t}{h}\right) u_{s,m}^* u_{t,m}^* \right]$$

$$= \frac{2}{(n-1)^2 h^q} \mathbb{E}\left\{ \sum_{i=1}^{n} \sum_{l \neq 1}^{n} u_{i,j}^* u_{l,j}^* u_{i,m}^* u_{l,m}^* \left[K\left(\frac{X_i - X_l}{h}\right) \right]^2 \right\}$$

$$= \frac{2n}{(n-1)h^q} \int \sigma_{j,m}^*(x) \sigma_{j,m}^*(y) \left[K\left(\frac{x-y}{h}\right) \right]^2 f(x) f(y) dx dy$$

$$= 2K^{(2)}(0) \int [\sigma_{j,m}^*(x)]^2 f^2(x) dx + O(h)$$

$$\to V_{j,m}^*. \tag{39}$$

Thus, proofs of (27) and (28) are straightforward from (38) and (39).

By the Cramér–Wold device and a similar calculation to the proof of (30), it can be straightforwardly shown that Z_2 converges to a multivariate normal distribution with the $(J-1) \times 1$ mean vector consisting of zero and variance-covariance matrix V^*, where V^* is a $(J-1) \times (J-1)$ variance-covariance matrix whose (j,m) elements are $V_{j,m}^*$. Therefore, C_n^*, which is the quadratic form of $nh^{q/2}\hat{Z}_{n,j}^*$, converges to a chi-squared distribution with $J-1$ degrees of freedom. $\square$

References

1. McFadden, D. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*; Zarembka, P., Ed.; Academic Press: New York, NY, USA, 1974; pp. 105–142.

2. Hausman, J.A.; Wise, D.A. A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences. *Econometrica* **1978**, *46*, 403–426.

3. Berry, S.; Levinsohn, J.; Pakes, A. Automobile Prices in Market Equilibrium. *Econometrica* **1995**, *63*, 841–890.

4. Goldberg, P.K. Product Differentiation and Oligopoly in International Markets: The Case of the US Automobile Industry. *Econometrica* **1995**, *63*, 891–951.

5. Heckman, J.J. The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Ann. Econ. Soc. Meas.* **1976**, *5*, 475–492.

6. Dubin, J.A.; McFadden, D.L. An Econometric Analysis of Residential Electric Appliance Holdings and Consumption. *Econometrica* **1984**, *52*, 345–362.

7. Härdle, W.; Mammen, E. Comparing Nonparametric versus Parametric Regression Fits. *Ann. Stat.* **1993**, *21*, 1926–1947.

8. Bierens, H.J. Consistent Model Specification Tests. *J. Econ.* **1982**, *20*, 105–134.

9. Bierens, H.J. Model Specification Testing of Time Series Regressions. *J. Econom.* **1984**, *26*, 323–353.

10. Bierens, H.J. A Consistent Conditional Moment Test of Functional Form. *Econometrica* **1990**, *58*, 1443–1458.

11. Delgado, M.A. Testing the Equality of Nonparametric Regression Curves. *Stat. Probab. Lett.* **1993**, *17*, 199–204.

12. De Jong, R.M. The Bierens Test under Data Dependence. *J. Econom.* **1996**, *72*, 1–32.

13. Andrews, D.W. A Conditional Kolmogorov Test. *Econometrica* **1997**, *65*, 1097–1128.

14. Bierens, H.J.; Ploberger, W. Asymptotic Theory of Integrated Conditional Moment Tests. *Econometrica* **1997**, *65*, 1129–1151.

15. Stute, W. Nonparametric Model Checks for Regression. *Ann. Stat.* **1997**, *25*, 613–641.

16. Stinchcombe, M.B.; White, H. Consistent Specification Testing with Nuisance Parameters Present Only under the Alternative. *Econom. Theory* **1998**, *14*, 295–325.

17. Chen, X.; Fan, Y. Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series. *J. Econom.* **1999**, *91*, 373–401.

18. Whang, Y.J. Consistent Bootstrap Tests of Parametric Regression Functions. *J. Econom.* **2000**, *98*, 27–46.

19. Eubank, R.L.; Spiegelman, C.H. Testing the Goodness of fit of a Linear Model via Nonparametric Regression Techniques. *J. Am. Stat. Assoc.* **1990**, *85*, 387–392.

20. Le Cessie, S.; van Houwelingen, J.C. A Goodness-of-fit Test for Binary Regression Models, Based on Smoothing Methods. *Biometrics* **1991**, *47*, 1267–1282.

21. Wooldridge, J.M. A Test for Functional Form Against Nonparametric Alternatives. *Econom. Theory* **1992**, *8*, 452–475.

22. Yatchew, A.J. Nonparametric Regression Tests Based on an Infinite Dimensional Least Squares Procedure. *Econom. Theory* **1992**, *8*, 435–451.

23. Gozalo, P.L. A Consistent Model Specification Test for Nonparametric Estimation of Regression Function Models. *Econom. Theory* **1993**, *9*, 451–477.

24. Aït-Sahalia, Y.; Bickel, P.J.; Stoker, T.M. Goodness-of-fit Tests for Kernel Regression with an Application to Option Implied Volatilities. *J. Econom.* **2001**, *105*, 363–412.

25. Delgado, M.A.; Stengos, T. Semiparametric Specification Testing of non-Nested Econometric Models. *Rev. Econ. Stud.* **1994**, *61*, 291–303.

26. Horowitz, J.L.; Härdle, W. Testing a Parametric Model Against a Semiparametric Alternative. *Econom. Theory* **1994**, *10*, 821–848.

27. Hong, Y.; White, H. Consistent Specification Testing via Nonparametric Series Regression. *Econometrica* **1995**, *63*, 1133–1159.

28. Fan, Y.; Li, Q. Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms. *Econometrica* **1996**, *64*, 865–890.

29. Lavergne, P.; Vuong, Q.H. Nonparametric Selection of Regressors: The Nonnested Case. *Econometrica* **1996**, *64*, 207–219.

30. Zheng, J.X. A Consistent Test of Functional Form via Nonparametric Estimation Techniques. *J. Econom.* **1996**, *75*, 263–289.

31. Li, Q.; Wang, S. A Simple Consistent Bootstrap Test for a Parametric Regression Function. *J. Econom.* **1998**, *87*, 145–165.

32. Lavergne, P.; Vuong, Q. Nonparametric Significance Testing. *Econom. Theory* **2000**, *16*, 576–601.

33. Fan, Y.; Li, Q. Consistent Model Specification Tests. *Econom. Theory* **2000**, *16*, 1016–1041.

34. Mora, J.; Moro-Egido, A.I. On Specification Testing of Ordered Discrete Choice Models. *J. Econom.* **2008**, *143*, 191–205.

35. Fan, J.; Huang, L.S. Goodness-of-fit Tests for Parametric Regression Models. *J. Am. Stat. Assoc.* **2001**, *96*, 640–652.

36. Horowitz, J.L.; Spokoiny, V.G. An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative. *Econometrica* **2001**, *69*, 599–631.

37. Spokoiny, V. Data-Driven Testing the fit of Linear Models. *Math. Methods Stat.* **2001**, *10*, 465–497.

38. Baraud, Y.; Huet, S.; Laurent, B. Adaptive Tests of Linear Hypotheses by Model Selection. *Ann. Stat.* **2003**, *31*, 225–251.

39. Zhang, C.M. Adaptive Tests of Regression Functions via Multiscale Generalized Likelihood Ratios. *Can. J. Stat.* **2003**, *31*, 151–171.

40. Guerre, E.; Lavergne, P. Data-Driven Rate-Optimal Specification Testing in Regression Models. *Ann. Stat.* **2005**, *33*, 840–870.

41. Härdle, W.; Müller, M.; Sperlich, S.; Werwatz, A. *Nonparametric and Semiparametric Models*; Springer: Berlin Heidelberg, Germany, 2004; p. 92.

42. Powell, J.L.; Stock, J.H.; Stoker, T.M. Semiparametric Estimation of Index Coefficients. *Econometrica* **1989**, *57*, 1403–1430.

43. Hall, P. Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators. *J. Multivar. Anal.* **1984**, *14*, 1–16.