Econometric Theory, **22**, 2006, 587–613. Printed in the United States of America.
DOI: 10.1017/S0266466606060294

A NONPARAMETRIC BOOTSTRAP TEST OF CONDITIONAL DISTRIBUTIONS

Yanqin Fan
Vanderbilt University

Qi Li
Texas A&M University

Insik Min
Kyung Hee University

This paper proposes a bootstrap test for the correct specification of parametric conditional distributions. It extends Zheng's test (Zheng, 2000, *Econometric Theory* 16, 667–691) to allow for discrete dependent variables and for mixed discrete and continuous conditional variables. We establish the asymptotic null distribution of the test statistic with data-driven stochastic smoothing parameters. By smoothing both the discrete and continuous variables via the method of cross-validation, our test has the advantage of automatically removing irrelevant variables from the estimate of the conditional density function and, as a consequence, enjoys substantial power gains in finite samples, as confirmed by our simulation results. The simulation results also reveal that the bootstrap test successfully overcomes the size distortion problem associated with Zheng's test.

1. INTRODUCTION

Currently, there exists a substantial body of work on consistent model specification testing for regression models and for unconditional distribution (density) functions; see Bierens and Ploberger (1997), Delgado and Manteiga (2001), Fan (1994, 1997, 1998), Fan and Li (1996), Hong and White (1996), Wooldridge (1992), and the references therein. In many economic applications, however, it is the distribution of one variable conditional on some other variables that is of more direct interest. The popular parametric binary or multinomial response models are but two leading examples of conditional probability models.

We are grateful for the insightful comments from three referees and a co-editor that greatly improved the paper. Li's research is partially supported by the Private Enterprise Research Center, Texas A&M University. Fan is grateful to the National Science Foundation for research support. Address correspondence to Qi Li, Department of Economics, Texas A&M University, College Station, TX 77843-4228, USA; e-mail: qi@econmail.tamu.edu.

 0266-4666/06 $12.00

Conditional probability models also are widely deployed in risk management and insurance settings, where the dependent variable of interest may be the claim size (a continuous variable) and the explanatory variables usually contain a mixture of discrete and continuous variables such as sex, age, whether children are present, whether one smokes, and so forth. Moreover, in risk management analysis, usually one is interested in the entire (conditional) distribution, rather than only in the conditional mean itself. Hence, a conditional probability model is more useful than a regression model in risk analysis. Relatively speaking, tests for conditional probability models are scarce. Zheng (2000), using kernel density estimators, proposed a consistent test for a parametric conditional density function. He showed that the limiting distribution of his test statistic is $N(0,1)$ and that the test can detect Pitman local alternatives approaching the null distribution at the rate of $(nh^{q/2})^{-1/2}$, where n is the sample size, h is the bandwidth, and q is the dimension of the conditioning variables. To apply Zheng's test to a given data set, one needs to choose the bandwidth; no guidance is provided on how this should be accomplished. Moreover, the requirement that both the dependent variable y and conditioning variables x are continuous variables severely limits the scope of application of Zheng's test, as many economic data sets contain both continuous and discrete variables. Andrews (1997) proposed a conditional Kolmogorov (CK) test for testing a parametric conditional distribution function. His test overcomes the difficulties associated with Zheng's test; it does not involve smoothing parameters and allows for both discrete and continuous variables. The critical values of the CK test of Andrews are obtained via a parametric bootstrap procedure, and the test can detect Pitman type local alternatives that approach the null model at the rate of $O(n^{-1/2})$. Although Andrews' test can handle both continuous and discrete variables, it does not produce an estimate of the conditional density function, which is of course undesirable when the parametric distribution function is rejected. In addition, it does not distinguish between relevant and irrelevant explanatory variables.

A related literature is the work on dynamic integral probability transform models such as that outlined in Diebold, Gunther, and Tay (1998). Corradi and Swanson (2004) and Li and Tkacz (2004) have also proposed bootstrap-based tests for conditional distributions. The Corradi and Swanson (2004) procedure is a nonsmoothing test similar to that of Andrews (1997), and their test extends Andrews' test to the time series data setting. Li and Tkacz (2004) use kernel smoothing; however, like Zheng (2000), they only consider the case whereby both y and x are continuous variables. The conventional way of handling discrete variables when estimating a conditional density function involving both discrete and continuous explanatory variables is by the so-called frequency method in which the entire sample is first split into a number of distinct cells and the data in each cell are then used to estimate the conditional density that is a function of the remaining continuous variables. For economic data, however, it is typically the case that the number of discrete cells is comparable to or even larger than the sample size. This renders the nonparametric frequency approach infeasible. Moreover, one may not know which conditional variables should be included in a particular application and hence faces the danger of including potentially irrelevant variables in the estimate. This is unfortunate, particularly in nonparametric settings, as including irrelevant explanatory variables has serious consequences for the accuracy of the resulting estimate: the rate of convergence of the density estimator will deteriorate quickly with the number of irrelevant continuous variables (the "curse of dimensionality"), whereas the number of cells will increase quite quickly with the number of irrelevant discrete variables. Recently, Hall, Racine, and Li (2004) proposed estimating a conditional density by smoothing both the discrete and continuous variables and showed that the use of cross-validation can automatically remove irrelevant variables from the resulting estimate. This is because the cross-validation method selects bandwidths that converge to some optimal values for relevant variables but selects large values for irrelevant conditional variables, thereby effectively smoothing out the irrelevant variables from the resulting estimate.

In this paper, we exploit the approach of Hall et al. (2004) to establish an alternative test for a parametric conditional density function. It is constructed based on the Zheng (2000) setup; however, it improves upon Zheng's test in a number of important ways: (i) the bandwidth is automatically chosen by cross-validation, thereby avoiding potential arbitrariness in the test's outcome due to an arbitrary choice of the bandwidth; (ii) it allows for both discrete and continuous variables; and (iii) the critical values are obtained from a parametric bootstrap procedure, which corrects the size distortions present in Zheng's approach. Although (ii) and (iii) are shared by Andrews' CK test, our test automatically produces an estimate of the conditional density function when the parametric density function is rejected by the test. More importantly, by automatically smoothing both the discrete and continuous variables via the method of cross-validation, our test has the advantage of automatically removing irrelevant variables from the resulting estimate (see Hall et al., 2004) and, as a consequence, enjoys substantial power gains in finite samples, as confirmed by our simulation results. Although our proposed test can only detect Pitman local alternatives approaching the null at rates slower than $O(n^{-1/2})$, it can be shown that for high-frequency alternatives, our test can detect local alternatives that approach the null at rates $o(n^{-1/2})$ in terms of the L_1 norm of the difference between the local alternative and the null model (e.g., Fan, 1998; Fan and Li, 2000). Hence it provides a complement to Andrews' CK test.

The remainder of this paper is organized as follows. In Section 2 we review and suggest a modified version of Zheng's test statistic. We also propose a bootstrap method for approximating the null distribution of our test. Section 3 reports Monte Carlo simulation results that examine the finite-sample performance of the proposed test. Finally, Section 4 concludes. Proofs are presented in the Appendix.

2. THE NULL HYPOTHESIS AND THE TEST

2.1. Zheng's Test

We begin by briefly reviewing the test proposed by Zheng (2000). Suppose that the data consist of $\{y_i, x_i\}_{i=1}^n$, an independent and identically distributed (i.i.d.) sample drawn from the distribution of (y, x) with the joint density function $p(y, x)$. Let $p(y|x)$ denote the conditional density function of y given x. We are interested in testing whether $p(y|x)$ belongs to a particular parametric family. Let $f(y|x, \theta)$ denote a parametric conditional density function with θ being a $k \times 1$ parameter. The null hypothesis is given by

$$H_0 : \Pr[p(y_i|x_i) = f(y_i|x_i, \theta_0)] = 1 \quad \text{for some } \theta_0 \in \Theta,$$

where Θ is the parameter space that is a compact set in $\mathcal{R}^k$. The alternative hypothesis is the negation of the null:

$$H_1 : \Pr[p(y_i|x_i) = f(y_i|x_i, \theta)] < 1 \quad \text{for all } \theta \in \Theta.$$

The Kullback–Leibler information criterion (Kullback and Leibler, 1951), measuring the discrepancy of two conditional density functions, is defined as

$$I(p, f) = E\left\{\log\left[\frac{p(y_i|x_i)}{f(y_i|x_i, \theta_0)}\right]\right\}. \tag{1}$$

It is well known that $I(p, f) \geq 0$ and $I(p, f) = 0$ if and only if $p(y|x) = f(y|x, \theta_0)$ almost everywhere (a.e.). Thus, $I(p, f)$ serves as a proper measure to test H_0. For technical reasons, instead of basing his test on the information measure, Zheng (2000) considered its first-order expansion,

$$\begin{aligned} E\left\{\log\left[\frac{p(y_i|x_i)}{f(y_i|x_i, \theta_0)}\right]\right\} &\cong E\left[\frac{p(y_i|x_i)}{f(y_i|x_i, \theta_0)} - 1\right] \\ &= E\left[\frac{p(y_i|x_i) - f(y_i|x_i, \theta_0)}{f(y_i|x_i, \theta_0)}\right]. \end{aligned} \tag{2}$$

Weighting (2) by the marginal density $p_1(x)$ of the conditional variable x leads to the following measure:

$$I_1(p, f) = E\left[\frac{p(y_i, x_i) - f(y_i|x_i, \theta_0)p_1(x_i)}{f(y_i|x_i, \theta_0)}\right]. \tag{3}$$

Zheng (2000) has shown that $I_1(p, f) \geq 0$ and the equality holds if and only if H_0 is true. Therefore, $I_1(p, f)$ also serves as a proper measure to test for H_0. For continuous random variables y and x, Zheng (2000) proposed estimating $p(y_i, x_i)$ by a standard kernel density estimator and estimating $f(y_i|x_i, \theta_0)p_1(x_i)$ by a smoothed density estimator $\tilde{p}(y_i, x_i)$ given by

$$\tilde{p}(y_i, x_i) = \frac{1}{n}\sum_{j=1}^{n} \int w_{2,h_y}\left(\frac{y_i - y}{h_y}\right) W_h\left(\frac{x_i - x_j}{h}\right) f(y|x_j, \hat{\theta})\, dy, \tag{4}$$

where $w_{2,h_y}(\cdot) = h_y^{-1} w_2(\cdot)$, $w_2(\cdot)$ is a (specially defined) univariate kernel function, $W_h(\cdot)$ is a product kernel $W_h((x_i - x_j)/h) = \prod_{s=1}^{q} h_s^{-1} w((x_{is} - x_{js})/h_s)$ with $w(\cdot)$ being a standard (second-order) univariate kernel, h_y and h_s's are the smoothing parameters, and $\hat{\theta}$ is an estimator of θ_0 under the null model. The measure $I_1(p, f)$ is then estimated by

$$T_{n,h}^c = \frac{1}{n(n-1)} \times \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \left[\frac{w_{2,h_y}\left(\frac{y_i - y_j}{h_y}\right) W_h\left(\frac{x_i - x_j}{h}\right) - \int w_{2,h_y}\left(\frac{y_i - y}{h_y}\right) W_h\left(\frac{x_i - x_j}{h}\right) f(y|x_j, \hat{\theta})\, dy}{f(y_i|x_i, \hat{\theta})} \right]. \tag{5}$$

To establish the null asymptotic distribution of $T_{n,h}^c$, Zheng (2000) suggested transforming the dependent variable such that it takes values in $[0,1]$ and then choosing a special kernel function for $w_2(\cdot)$ with the property that $h_y^{-1}\int_0^1 w_2((y_i - y)/h_y)^2\, dy \to 1$ as $n \to \infty$. The use of the smoothed estimator $\tilde{p}(y, x)$ eliminates the bias of the kernel estimator of $p(y_i, x_i)$ under H_0 such that the test statistic is appropriately centered for a wide range of smoothing parameter values. Under some regularity conditions, Zheng (2000) showed that the asymptotic null distribution of $T_{n,h}^c$ is normal and provided a consistent estimator of its asymptotic variance.

2.2. Our Framework

We now extend Zheng's test to include both continuous and discrete explanatory variables (x is a mixed variable), where the dependent variable y can be discrete or continuous.

We first consider the case that y is a discrete variable. In this case, we show that the smoothed estimator $\tilde{p}(y, x)$ reduces to an average estimator. Thus, the resulting test statistic only involves summations and hence avoids the need for numerical integration.

Let $x = (x^c, x^d)$, where x^c is a $q \times 1$ continuous variable and x^d is an $r \times 1$ discrete variable. We use x_{is}^c (x_{is}^d) to denote the sth component of x_i^c (x_i^d). We further assume that x_{is}^d takes the values in $\{0, 1, \ldots, c_s - 1\}$ (it takes c_s different values).

In constructing the kernel density estimate, we use different kernel functions for the discrete and continuous variables. For the discrete variable x^d, we use the Aitchison and Aitken (1976) kernel: $l(x_{is}^d, x_{js}^d, \lambda_s) = 1 - \lambda_s$ if $x_{is}^d = x_{js}^d$, and

https://doi.org/10.1017/S0266466606060294 Published online by Cambridge University Press

$l(x_{is}^d, x_{js}^d, \lambda_s) = \lambda_s/(c_s - 1)$ if $x_{is}^d \neq x_{js}^d$. Hence, the product kernel for the discrete variable is

$$L(x_i^d, x_j^d, \lambda) = \prod_{s=1}^{r} l(x_{is}^d, x_{js}^d, \lambda_s) = \prod_{s=1}^{r} \{\lambda_s/(c_s - 1)\}^{N_{is}(x)}(1 - \lambda_s)^{1-N_{is}(x)},$$

where $N_{is}(x) = I(x_{is}^d \neq x_{js}^d)$, in which $I(\cdot)$ is the usual indicator function and $\lambda_1, \ldots, \lambda_r$ are the smoothing parameters for the discrete components and are constrained by $0 \leq \lambda_s \leq (c_s - 1)/c_s$. Note that when λ_s assumes the upper extreme value, $(c_s - 1)/c_s$, $l(x_{is}^d, x_{js}^d, \lambda_s = (c_s - 1)/c_s) \equiv 1/c_s$ becomes unrelated to (x_{is}^d, x_{js}^d), i.e., the sth component of x^d is completely smoothed out when $\lambda_s = (c_s - 1)/c_s$.

For the continuous component x^c, we still use the standard (second-order) product kernel function as discussed earlier. Therefore, for the mixed type variable $x = (x^c, x^d)$, the kernel function is defined by

$$K_{\gamma,ij} = K_\gamma(x_i, x_j) \stackrel{def}{=} W_h\left(\frac{x_i^c - x_j^c}{h}\right) L(x_i^d, x_j^d, \lambda), \tag{6}$$

where $\gamma = (h, \lambda) \equiv (h_1, \ldots, h_q, \lambda_1, \ldots, \lambda_r)$.

We now discuss how to estimate $p(y_i, x_i)$ and $p_1(x_i)$. Assume that y_i is a discrete variable; then we estimate $p(y_i, x_i)$ and $p_1(x_i)$ by the following leave-one-out kernel estimators:

$$\hat{p}_{-i}(y_i, x_i) = \frac{1}{n} \sum_{j \neq i}^{n} I(y_i = y_j) K_\gamma(x_i, x_j), \tag{7}$$

$$\hat{p}_{1,-i}(x_i) = \frac{1}{n} \sum_{j \neq i}^{n} K_\gamma(x_i, x_j). \tag{8}$$

To construct the smoothed estimator of $f(y_i|x_i, \theta_0)$, we replace $W_h(\cdot)$ in (4) by $K_\gamma(x_i, x_j)$ and $\int w_{2,h_y}((y_i - y)/h_y)\, dy$ by $\sum_y I(y_i - y)$. Taking into account these modifications, we obtain

$$\begin{aligned}\tilde{p}(y_i, x_i) &= \frac{1}{n} \sum_{j=1}^{n} \sum_{y} I(y_i = y) K_\gamma(x_i, x_j) f(y|x_j, \hat{\theta}) \\ &= \frac{1}{n} \sum_{j=1}^{n} K_\gamma(x_i, x_j) f(y_i|x_j, \hat{\theta}).\end{aligned} \tag{9}$$

Using $\hat{p}_{-i}(y_i, x_i)$, $\hat{p}_{1,-i}(x_i)$, and $\tilde{p}(y_i, x_i)$ just introduced, we define our test statistic as

$$T_{n,\gamma} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \left\{ \frac{K_\gamma(x_i, x_j)}{f(y_i|x_i, \hat{\theta})} [I(y_i = y_j) - f(y_i|x_j, \hat{\theta})] \right\}. \tag{10}$$

Note that the double summation in $T_{n,\gamma}$ does not include $j = i$ terms because we have used the leave-one-out estimators for estimating $p(y_i, x_i)$ and $p_1(x_i)$. The reason for using these leave-one-out estimators is that, under H_0, the asymptotic distribution of $T_{n,\gamma}$ will be centered at zero (there is no center term).

The smoothing parameters $h_1, \ldots, h_q$ (corresponding to the continuous variable x^c) can be selected by several commonly used procedures, including the cross-validation method, the plug-in method, and some ad hoc methods. However, for $\lambda_1, \ldots, \lambda_r$, the plug-in or even an ad hoc formula is not available. Hall et al. (2004) have shown that using the cross-validation method to select $\lambda_1, \ldots, \lambda_r$ and $h_1, \ldots, h_q$ has some nice properties: when x_s^c (x_s^d) is a relevant variable, the cross-validation method will select a small $h_s(\lambda_s)$ that converges to zero at an optimal rate; when x_s^c (x_s^d) is an irrelevant variable,[1] the cross-validation method will select an extremely large value for h_s (upper bound value for λ_s) so that the irrelevant variables are (asymptotically) automatically removed (smoothed out). Indeed in the problem of nonparametric estimation of a conditional density, cross-validation comes into its own as a method with no obvious peers. Therefore, we will choose $\lambda_1, \ldots, \lambda_r$, $h_1, \ldots, h_q$ by the cross-validation method suggested in Hall et al. (2004).

Let $(h, \lambda) = (h_1, \ldots, h_q, \lambda_1, \ldots, \lambda_r)$. Hall et al. propose choosing (h, λ) by minimizing the following objective function:[2]

$$CV(h,\lambda) = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{G}_{-i}(x_i) m(x_i^c)}{\hat{p}_{1,-i}(x_i)^2} - \frac{2}{n}\sum_{i=1}^{n} \frac{\hat{p}_{-i}(x_i, y_i) m(x_i^c)}{\hat{p}_{1,-i}(x_i)}, \qquad \textbf{(11)}$$

where

$$\hat{G}_{-i}(X_i) = \frac{1}{(n-1)^2} \sum_{i_1 \neq i}^{n} \sum_{i_2 \neq i}^{n} K_\gamma(x_i, x_{i_1}) K_\gamma(x_i, x_{i_2}) I(y_{i_1} = y_{i_2}),$$

in which $\hat{p}_{1,-i}(x_i)$ and $\hat{p}_{-i}(x_i, y_i)$ are the leave-one-out kernel estimators of $p_1(x_i)$ and $p(x_i, y_i)$, respectively, and $m(x_i^c)$ is a weight function introduced to deal with the small random denominator problem; see Hall et al. (2004).

We will use $\hat{h}_1, \ldots, \hat{h}_q$ and $\hat{\lambda}_1, \ldots, \hat{\lambda}_r$ to denote the resulting smoothing parameters. Assuming that all the x variables are relevant variables, Hall et al. (2004) showed that $\hat{h}_s = a_s^0 n^{-1/(q+4)} + o_p(n^{-1/(q+4)})$ for $s = 1, \ldots, q$, and $\hat{\lambda}_s = b_s^0 n^{-2/(q+4)} + o_p(n^{-2/(q+4)})$ for $s = 1, \ldots, r$, where a_s^0 and b_s^0 are some finite constants.

THEOREM 2.1. *Under conditions (C1)–(C3) given in the Appendix, we have under* H_0

$$J_{n,\hat{\gamma}} \stackrel{def}{=} n(\hat{h}_1 \ldots \hat{h}_q)^{1/2} T_{n,\hat{\gamma}} / \sqrt{\hat{V}_{n,\hat{\gamma}}} \to N(0,1) \quad \textit{in distribution,}$$

https://doi.org/10.1017/S0266466606060294 Published online by Cambridge University Press

where $\hat{\gamma} = (\hat{h}_1, \ldots, \hat{h}_q, \hat{\lambda}_1, \ldots, \hat{\lambda}_r)$ *and* $\hat{V}_{n,\hat{\gamma}} = 2[n(n-1)]^{-1} \sum_i \sum_{j \neq i} \{K_{\hat{\gamma}}(x_i, x_j)[I(y_i = y_j) - f(y_i|x_j, \hat{\theta})]/\hat{f}(y_i|x_i, \hat{\theta})\}^2$ *is a consistent estimator of* $\sigma_0^2 = [\int W^2(v)\,dv]E[(1 - f(y_i|x_i, \theta_0))f^{-1}(y_i|x_i, \theta_0)p_1(x_i)]$, *the asymptotic variance of* $n(\hat{h}_1 \ldots \hat{h}_q)^{1/2} T_{n,\hat{\gamma}}$.

It can be shown that under H_1, $J_{n,\hat{\gamma}}$ diverges to $+\infty$. Hence, the $J_{n,\hat{\gamma}}$ test is a consistent test. Moreover, the $J_{n,\hat{\gamma}}$ test can detect local alternatives that approach the null at a rate of $O_p(n^{-1/2}(h_1 \ldots h_q)^{-1/4}) = O_p(n^{-(1/2)((8+q)/(8+2q))})$, which is slower than $O_p(n^{-1/2})$ (because $h_j = O_p(n^{-1/(4+q)})$ for all $j = 1, \ldots, q$).

We now briefly discuss the case where the dependent variable y is continuous. In this case, one can still use Zheng's test statistic given in (5) but with $w_{2,h_y}((y_i - y_j)/h_y)$ and $W_h((x_i^c - x_j^c)/h)$ being replaced by $w_{2,\hat{h}_y}((y_i - y_j)/\hat{h}_y)$ and $K_{\hat{\gamma},ij} = W_{\hat{h}}((x_i^c - x_j^c)/\hat{h})L(x_i^d, x_j^d, \hat{\lambda})$, respectively, where $(\hat{h}_y, \hat{h}, \hat{\lambda}) = (\hat{h}_y, \hat{h}_1, \ldots, \hat{h}_q, \hat{\lambda}_1, \ldots, \hat{\lambda}_r)$ denote the cross-validation selected smoothing parameters suggested by Hall et al. (2004); i.e., one chooses (h_y, h, λ) by minimizing (11), but now $G_{-i}(x_i)$ is defined as

$$G_{-i}(x_i) = \frac{1}{(n-1)^2} \sum_{i_1 \neq i}^{n} \sum_{i_2 \neq i}^{n} K_\gamma(x_i, x_{i_1}) K_\gamma(x_i, x_{i_2}) \overline{w}_{2,h_y}\left(\frac{y_{i_1} - y_{i_2}}{h_y}\right),$$

where $\overline{w}_{2,h_y}(v) = h_y^{-1}\overline{w}_2(v)$ and $\overline{w}_2(v) = \int w_2(u)w_2(v-u)\,du$ is the twofold convolution kernel derived from $w_2(\cdot)$.

With a slight abuse of notation, the resulting test statistic becomes

$$T_{n,\hat{\gamma}}^c = \frac{1}{n(n-1)} \times \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \left[\frac{w_{2,\hat{h}_y}\left(\frac{y_i - y_j}{\hat{h}_y}\right) K_{\hat{\gamma},ij} - \int w_{2,\hat{h}_y}\left(\frac{y_i - y}{\hat{h}_y}\right) K_{\hat{\gamma},ij} f(y|x_j, \hat{\theta})\,dy}{f(y_i|x_i, \hat{\theta})} \right], \tag{12}$$

where $\hat{\gamma} = (\hat{h}_y, \hat{h}_1, \ldots, \hat{h}_q, \hat{\lambda}_1, \ldots, \hat{\lambda}_r)$ contains the extra smoothing parameter $\hat{h}_y$ because y_i is continuous.

The asymptotic distribution of $T_{n,\hat{\gamma}}^c$ is given in the following theorem.

THEOREM 2.2. *Under conditions (C1)–(C3) given in the Appendix, we have under* H_0,

$$J_{n,\hat{\gamma}}^c \stackrel{def}{=} n(\hat{h}_y \hat{h}_1 \ldots \hat{h}_q)^{1/2} T_{n,\hat{\gamma}}^c \Big/ \sqrt{\hat{V}_{n,\hat{\gamma}}^c} \to N(0,1) \quad \textit{in distribution},$$

where $\hat{V}_{n,\hat{\gamma}}^c = (2\hat{h}_1 \ldots \hat{h}_q)/n(n-1) \sum_i \sum_{j \neq i} K_{\hat{\gamma},ij}^2$ *is a consistent estimator of* $\sigma_{0,c}^2 = 2[\int W^2(v)\,dv]E[p_1(x_i)]$, *the asymptotic variance of* $n(\hat{h}_y \hat{h}_1 \ldots \hat{h}_q)^{1/2} T_{n,\hat{\gamma}}^c$.

The proof of Theorem 2.2 is similar to that of Theorem 1 in Zheng (2000) and is omitted here.

2.3. A Parametric Bootstrap Test

Theorems 2.1 and 2.2 provide, respectively, the asymptotic null distribution of $J_{n,\hat{\gamma}}$ and $J_{n,\hat{\gamma}}^c$. Consequently, one can perform tests for H_0 by comparing the value of $J_{n,\hat{\gamma}}$ (or $J_{n,\hat{\gamma}}^c$) with its asymptotic critical value. However, it is well known that consistent nonparametric tests often suffer from substantial finite-sample size distortions. Our simulations reveal that the $J_{n,\hat{\gamma}}$ ($J_{n,\hat{\gamma}}^c$) shares this drawback. To overcome this problem, we propose a bootstrap procedure to more accurately approximate the finite-sample null distribution of $J_{n,\hat{\gamma}}$ ($J_{n,\hat{\gamma}}^c$). It involves the following steps.

Step (i). Generate the ith bootstrap value of the dependent variable y from the parametric conditional distribution $f(\cdot|x_i,\hat{\theta})$. Denote this value by y_i^* ($i = 1,\ldots,n$). We have the complete bootstrap sample $\{x_i, y_i^*\}_{i=1}^n$.

Step (ii). Based on the parametric null model, estimate θ using the bootstrap sample. Let $\hat{\theta}^*$ denote the resulting estimator. Compute the bootstrap statistic $J_{n,\hat{\gamma}}^*$ ($J_{n,\hat{\gamma}}^{c*}$) in the same way as $J_{n,\hat{\gamma}}$ ($J_{n,\hat{\gamma}}^c$) except that $\{y_i\}_{i=1}^n$ and $\hat{\theta}$ are replaced by $\{y_i^*\}_{i=1}^n$ and $\hat{\theta}^*$, respectively. Note that we use the same cross-validation selected smoothing parameter $\hat{\gamma}$ in computing the bootstrap statistics. There is no re-cross-validation in computing $T_{n,\hat{\gamma}}^*$ ($T_{n,\hat{\gamma}}^{c*}$).

Step (iii). Repeat steps (i) and (ii) a large number of times, say, B times, and use the empirical distribution of the B bootstrap statistics $\{J_{n,\hat{\gamma}}^*\}_{j=1}^B$ ($\{J_{n,\hat{\gamma}}^{c*}\}_{j=1}^B$) to approximate the null distribution of $J_{n,\hat{\gamma}}$ ($J_{n,\hat{\gamma}}^c$).

Step (iv). The bootstrap test rejects H_0 at significance level α if $J_{n,\hat{\gamma}}$ ($J_{n,\hat{\gamma}}^c$) exceeds the empirical αth percentile of $\{J_{n,\hat{\gamma}}^*\}_{j=1}^B$ ($\{J_{n,\hat{\gamma}}^{c*}\}_{j=1}^B$).

The following theorem justifies the asymptotic validity of the bootstrap test.

THEOREM 2.3. *Assume the same conditions as in Theorem 2.1 (Theorem 2.2) except the null hypothesis. We have*

$$\sup_{z\in\mathcal{R}}|P(J_{n,\hat{\gamma}}^* \le z|\{x_i,y_i\}_{i=1}^n) - \Phi(z)| = o_p(1), \quad \textbf{(13)}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable.

The proof of Theorem 2.3 is given in the Appendix.

In words, Theorem 2.3 states that $J_{n,\hat{\gamma}}^*$ converges to $N(0,1)$ in distribution in probability. Other authors show that some bootstrap method works using the concept of convergence with probability one, where one states that the left-hand side of (13) is $o(1)$ with probability one (i.e., convergence in distribution with probability one). Here we choose to use the concept of convergence in distribution in probability because our test statistic involves nonparametric estimation and it is easier to work with "convergence in probability" than "convergence with probability one."

Note that Theorem 2.3 holds true regardless of whether the null hypothesis is true or not. Therefore, (i) when the null hypothesis is true, the bootstrap procedure will lead to (asymptotically) correct size of the test, because $J_{n,\hat{\gamma}}$ converges in distribution to the same $N(0,1)$ limiting distribution under H_0; (ii) when the null hypothesis is false, because the test statistic $T_{n,\hat{\gamma}}$ will converge to $+\infty$ in probability, whereas asymptotically the bootstrap critical value is still finite (say, the 95th quantile from the $N(0,1)$ distribution), the bootstrap procedure leads to a consistent test.

3. MONTE CARLO SIMULATION RESULTS

In this section, we present Monte Carlo simulation results to examine the finite-sample performance of our $J_{n,\hat{\gamma}}$ ($J^c_{n,\hat{\gamma}}$) test.

3.1. Discrete Dependent Variable

In this simulation experiment, the dependent variable y is a $\{0,1\}$ binary variable. We use a slightly different notation in this section; x denotes x^c and z denotes x^d. The data generating process (DGP) for the null model is given by

$$DGP_0^a: \quad y_i = 1 \quad \text{if } \beta_0 + \beta_1 x_i + \beta_2 z_i + u_i > 0,$$
$$y_i = 0 \quad \text{otherwise,}$$

where $\{x_i\}_{i=1}^n$ is a random sample from $N(0,1)$, z_i takes binary values $\{0,1\}$ with case (i) $\Pr[z_i = 1] = \frac{1}{2}$ and $\Pr[z_i = 0] = \frac{1}{2}$ and case (ii) $\Pr[z_i = 1] = 0.8$ and $\Pr[z_i = 0] = 0.2$, and the error term $\{u_i\}$ is i.i.d. $N(0,1)$. Moreover, x_i, z_i, and u_i are all independent of each other. The true parameters are $\{\beta_0, \beta_1\} = \{1,1\}$ and $\beta_2 = \{1, 0.3, 0\}$; $\beta_2 = 0$ corresponds to the case that z_i is in fact an irrelevant variable. This leads to the following null hypothesis:

$$H_0 : p(y|x,z,\theta) = y\Phi(\beta_0 + \beta_1 x + \beta_2 z) + (1-y)[1 - \Phi(\beta_0 + \beta_1 x + \beta_2 z)],$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. The parametric conditional density of the null model is estimated by the maximum likelihood (ML) method.

The following two alternative DGPs are constructed to examine the power of the $J_{n,\hat{\gamma}}$ test; one has a nonlinear term in the index, and the other has a conditional heteroskedastic error:

$$DGP_1^a: \quad y_i = 1 \quad \text{if } \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i^2 + u_i > 0,$$
$$y_i = 0 \quad \text{otherwise;}$$
$$DGP_2^a: \quad y_i = 1 \quad \text{if } \beta_0 + \beta_1 x_i + \beta_2 z_i + x_i u_i > 0,$$
$$y_i = 0 \quad \text{otherwise,}$$

where x_i, z_i, and u_i are all generated in the same way as before. Also, $\beta_0, \beta_1, \beta_2$ take the same values as previously, whereas $\beta_3 = 1$. We use the parametric bootstrap described earlier to approximate the null distribution of the test statistic $J_{n,\hat{\gamma}}$.

Our test will be compared with the CK test of Andrews (1997) with test statistic (CK_n) defined as

$$CK_n = \max_{1 \le j \le n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [I(y_i \le y_j) - F(y_j | x_i, z_i, \hat{\theta})] I(x_i \le x_j) I(z_i \le z_j) \right|, \quad \textbf{(14)}$$

where $F(\cdot|\cdot,\cdot,\theta)$ is the parametric conditional distribution function and $\hat{\theta}$ is the ML estimator of θ_0.

The sample sizes considered are $n = 100$ and 200, the numbers of simulations are 5,000 for size estimation and 2,000 for power estimation, and the number of bootstraps is $B = 1{,}000$ for all cases. The simulation results for discrete y_i with relevant covariates only are reported in Table 1.

From Table 1 we observe that for different values of β_2 (with $\beta_2 = 1, 0.3$) and different values of $\Pr(z_i = 1)$ (0.5, 0.8), the performances of the $J_{n,\hat{\gamma}}$ and Andrews' tests are qualitatively the same. Overall the estimated sizes are quite close to their nominal sizes for both tests. The power performances are mixed for the two alternative models. For the alternative DGP_1^a with an extra quadratic term, our test $J_{n,\hat{\gamma}}$ shows higher power than Andrews' test for the sample sizes considered. However, for some cases of DGP_2^a with a heteroskedastic error term, Andrews' test is slightly more powerful than ours. The simulation results show that our $J_{n,\hat{\gamma}}$ test complements Andrews' test.

Next we consider the case with an irrelevant covariate. We use the same DGP as before except that now we choose $\beta_2 = 0$ so that the binary discrete variable z becomes an irrelevant covariate. Because this information is unknown a priori, we still compute the conditional probability of y conditional on both x and z. In this case we expect that the cross-validation method tends to select the upper bound value of $\lambda = \frac{1}{2}$ so that the irrelevant covariate z is smoothed out automatically, resulting in a finite-sample power gain for the $J_{n,\hat{\gamma}}$ test.

From Table 2 we observe that the power of the $J_{n,\hat{\gamma}}$ test improves substantially compared with those reported in Table 1. It is interesting to observe that for DGP_2^a, the power performance of the $J_{n,\hat{\gamma}}$ test is quite comparable to that of Andrews' test. Thus, the simulation results confirm that our cross-validation-based test indeed has the ability to remove irrelevant covariates and enjoys superior finite-sample power performance.

3.2. Continuous Dependent Variable

In this section we consider the case where both y and x are continuous variables, and we compare the finite-sample performance of Zheng's original test with our $J_{n,\hat{\gamma}}$ test. The first DGP we use is the same as that in Zheng. The null model is a linear regression model with normal homoskedastic errors:

TABLE 1. DGP^a: The case of discrete y_i with relevant covariates

	$J_{n,\hat{\gamma}}$			Andrews (1997)		
	1%	5%	10%	1%	5%	10%
a. z_i is relevant ($\beta_2 = 1$) with $\Pr[z_i = 1] = 0.5$						
DGP_0 (size)						
$N = 100$	0.9	4.3	9.2	0.7	4.1	10.1
$N = 200$	1.1	5.5	11.2	1.1	4.7	9.5
DGP_1^a (power)						
$N = 100$	9.2	31.2	45.2	4.5	18.4	29.4
$N = 200$	31.2	57.2	70.8	12.2	34.8	53.2
DGP_2^a (power)						
$N = 100$	28.2	48.7	60.4	23.2	51.2	62.4
$N = 200$	56.8	77.2	84.4	50.2	78.3	85.7
b. z_i is relevant ($\beta_2 = 1$) with $\Pr[z_i = 1] = 0.8$						
DGP_0 (size)						
$N = 100$	0.8	5.2	9.5	0.8	4.8	9.4
$N = 200$	1.2	5.6	10.0	1.3	5.8	11.0
DGP_1^a (power)						
$N = 100$	23.0	46.3	60.3	7.5	27.7	44.2
$N = 200$	57.7	80.2	89.8	27.3	66.0	80.5
DGP_2^a (power)						
$N = 100$	31.4	55.8	70.9	31.3	55.0	70.0
$N = 200$	63.7	82.5	91.7	62.2	83.7	91.5
c. z_i is relevant ($\beta_2 = 0.3$) with $\Pr[z_i = 1] = 0.8$						
DGP_0 (size)						
$N = 100$	0.8	5.5	10.7	0.8	5.3	9.8
$N = 200$	1.4	5.3	10.0	1.5	6.1	11.2
DGP_1^a (power)						
$N = 100$	23.2	51.7	63.2	6.1	20.3	34.2
$N = 200$	53.7	75.2	87.5	17.5	46.7	68.8
DGP_2^a (power)						
$N = 100$	31.6	59.1	67.5	30.1	54.2	60.8
$N = 200$	60.8	78.5	89.1	56.0	76.2	83.7

$$DGP_0^b : y_i = \beta_0 + \beta_1 x_i + u_i,$$

where $\{x_i\}_{i=1}^n$ is a random sample from $N(0,1)$ and the error term $\{u_i\}$ is i.i.d. $N(0,\sigma^2)$. Moreover, x_i and u_i are independent of each other. The true parameters are $\{\beta_0, \beta_1, \sigma\} = \{1,1,1\}$. This leads to the following null hypothesis:

$$H_0 : p(y|x,\theta) = \phi[(y - \beta_0 - \beta_1 x)/\sigma]/\sigma,$$

https://doi.org/10.1017/S0266466606060294 Published online by Cambridge University Press

TABLE 2. The case of discrete y_i with irrelevant covariates

	$J_{n,\hat{\gamma}}$			Andrews (1997)		
	1%	5%	10%	1%	5%	10%
a. z_i is irrelevant ($\beta_2 = 0$) with $\Pr[z_i = 1] = 0.5$						
DGP_0 (size)						
$N = 100$	0.8	5.8	11.5	1.2	5.7	11.2
$N = 200$	1.3	5.6	10.9	1.3	5.6	11.0
DGP_1^a (power)						
$N = 100$	30.0	53.2	66.4	14.5	28.8	43.6
$N = 200$	71.6	89.6	94.4	36.4	70.0	83.2
DGP_2^a (power)						
$N = 100$	47.6	71.2	79.2	42.4	60.8	70.2
$N = 200$	78.9	89.0	94.7	76.8	88.4	92.4
b. z_i is irrelevant ($\beta_2 = 0$) with $\Pr[z_i = 1] = 0.8$						
DGP_0 (size)						
$N = 100$	0.8	5.9	11.2	1.2	5.8	11.5
$N = 200$	1.2	5.7	11.0	1.5	5.9	10.8
DGP_1^a (power)						
$N = 100$	35.3	61.8	72.3	12.5	27.8	44.0
$N = 200$	69.5	90.0	92.0	34.8	68.7	81.5
DGP_2^a (power)						
$N = 100$	45.8	70.8	78.8	42.4	60.8	70.2
$N = 200$	81.5	93.0	97.0	67.8	85.8	89.8

where $\phi(\cdot)$ is the standard normal density function. The parameter θ is estimated by the ML estimation method.

Two alternative models are considered: one is designed to test misspecification in the regression (DGP_1^b), and the second is to test homoskedasticity of the error term (DGP_2^b):

$$DGP_1^b : y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i,$$

$$DGP_2^b : y_i = \beta_0 + \beta_1 x_i + x_i u_i,$$

where β_2 is set to be 1 in the experiment. We also report Andrews' test for comparison purposes. The simulation results are reported in Table 3a.

We observe from Table 3a that the parametric bootstrap method successfully overcomes the size distortion of Zheng's test. The estimated sizes of the bootstrap test are all close to their nominal values, whereas Zheng's test based on the asymptotic normal approximation is significantly undersized. For the alternatives DGP_1^b and DGP_2^b, we observe that the bootstrap test $J_{n,\hat{\gamma}}^c$ is much more powerful than Zheng's test. There are two reasons for this: the first is that the

TABLE 3. DGP^b: The case of continuous y_i

	$J^c_{n,\hat{\gamma}}$			Zheng (2000)			Andrews (1997)		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
a. Continuous variable case without irrelevant covariates ($\beta_2 = 1$)									
DGP^b_0 (size)									
$N = 50$	1.4	5.2	10.5	1.3	1.9	2.7	0.8	4.5	9.0
$N = 100$	1.2	5.5	10.9	1.5	2.7	3.9	0.9	4.2	10.6
$N = 200$	0.9	4.5	8.9	1.7	2.5	4.0	1.2	4.0	9.8
DGP^b_1 (power)									
$N = 50$	84.8	96.0	98.4	56.3	75.6	84.0	72.7	92.2	96.7
$N = 100$	99.6	99.8	100.0	93.3	96.9	97.5	96.5	99.0	100
$N = 200$	100	100	100.0	100	100	100	99.7	100	100
DGP^b_2 (power)									
$N = 50$	48.0	73.2	82.4	28.4	44.3	56.8	19.8	29.0	35.8
$N = 100$	96.4	100	100	81.0	88.8	91.2	25.0	41.8	58.0
$N = 200$	100	100	100	96.3	96.6	96.7	32.0	51.3	69.6
b. Continuous variable case with an irrelevant covariate ($\beta_2 = 0$)									
DGP^c_0 (size)									
$N = 100$	0.07	4.8	10.0	1.6	4.2	5.8	0.05	3.0	6.5
$N = 200$	0.08	4.9	9.6	2.4	3.8	6.4	0.06	3.1	6.9
DGP^c_1 (power)									
$N = 50$	36.4	71.0	83.0	25.6	36.4	43.4	20.8	38.2	57.6
$N = 100$	75.2	93.6	97.8	50.4	65.4	71.8	41.2	65.0	78.4
DGP^c_2 (power)									
$N = 50$	73.4	93.6	97.8	50.4	65.2	71.8	24.8	39.0	48.7
$N = 100$	97.6	99.6	100	87.2	93.2	95.8	35.0	46.8	68.0

bootstrap test corrects the undersize problem of Zheng's test and hence improves the finite-sample power performance; the second reason is that we use the data-driven cross-validation method to select the smoothing parameters that lead to optimal smoothing in estimating the unknown conditional density functions, whereas Zheng suggested using some ad hoc method to select the smoothing parameters. It turns out that the use of optimal smoothing also enhances the finite-sample power of the test. For DGP^b_1, Andrews' test has similar power as the $J_{n,\hat{\gamma}}$ test, whereas for DGP^b_2, Andrews' test is less powerful than the $J_{n,\hat{\gamma}}$ test.

Finally we consider a case that there exists an irrelevant continuous variable. We will use basically the same setup as in DGP^b except that we set $\beta_2 = 0$ now. Therefore, x_{2i} becomes an irrelevant variable. However, this information is not used in the estimation. That is, all estimation methods still use the full data set $\{y_i, x_{1i}, x_{2i}\}_{i=1}^n$. Because our cross-validation-based $J_{n,\hat{\gamma}}$ has the advantage of (asymptotically) removing the irrelevant variable x_2, we expect that the $J_{n,\hat{\gamma}}$ test will enjoy further power gains. The simulation results are reported in Table 3b.

From Table 3b we observe that the $J_{n,\hat{\gamma}}$ test has good estimated sizes. Zheng's test still underestimates the sizes at the 5% and 10% levels. Andrews' test is also somewhat undersized when an irrelevant variable exists. From the estimated power results, we see substantial power gain of the $J_{n,\hat{\gamma}}$ test over Zheng's test. Essentially, Zheng's test is based on a two-dimensional nonparametric conditional density estimate because the smoothing parameters in Zheng's test are selected by some ad hoc rules that cannot detect the irrelevant variable x_2, whereas our $J_{n,\hat{\gamma}}$ test estimates, asymptotically, a one-dimensional conditional density function because x_{2i} will be smoothed out asymptotically. The $J_{n,\hat{\gamma}}$ test is also more powerful than Andrews' test for this DGP (when there is an irrelevant continuous variable). Of course here we only report a limited simulation result, from the local power analysis; we expect that there exist data generating processes for which Andrews' test will be more powerful than the $J_{n,\hat{\gamma}}$ test. Our simulation results show that the $J_{n,\hat{\gamma}}$ test can serve as a useful complement to Andrews' test when one is interested in testing a parametric conditional distribution.

4. CONCLUSIONS

This paper proposes a kernel-based bootstrap test for parametric conditional distribution functions. We separately consider the case where y is a discrete variable and where y is a continuous variable. In either case, the conditional variables can contain both discrete and continuous variables. By automatically smoothing both the discrete and continuous variables via the method of cross-validation, our test has the advantage of automatically removing irrelevant variables from the estimate of the conditional density function and, as a consequence, enjoys substantial power gains in finite-sample applications, as confirmed by our simulation results. The test is potentially applicable in a wide variety of applications and should prove useful to applied researchers.

NOTES

1. We say that x_s is an irrelevant variable if $p(y|x)$ is independent of x_s.

2. Hall et al. (2004) show that, up to an additive constant term that does not depend on (h,λ), $CV(h,\lambda)$ is a consistent estimator of the weighted integrated squared error: $\int\{\hat{p}(y|x) - p(y|x)\}^2 p_1(x) w(x^c)\,dxdy$, where $\int dxdy = \sum_{x^d}\int dx^c\,dy$ if y is a continuous variable and $\int dxdy = \sum_{x^d}\sum_y \int dx^c$ if y is a discrete variable.

REFERENCES

Aitchison, J. & C.G.G. Aitken (1976) Multivariate binary discrimination by the kernel method. *Biometrika* 63, 413–420.

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

Bhattacharya, R.N. & R.R. Rao (1986) *Normal Approximations and Asymptotic Expansions*. Krieger.
Bierens, H.J. & W. Ploberger (1997) Asymptotic theory of integrated conditional moment tests. *Econometrica* 65, 1129–1151.
Billingsley, P. (1968) *Probability and Measure*. Wiley.
Corradi, V. & N. Swanson (2004) Bootstrap conditional distribution tests under dynamic misspecification. *Journal of Econometrics*, forthcoming.
de Jong, R.M. (1996) The Bierens test under data dependence. *Journal of Econometrics* 72, 1–32.
Delgado, M.A. & W.G. Manteiga (2001) Significance testing in nonparametric regression based on the bootstrap. *Annals of Statistics* 29, 1469–1507.
Diebold, F.X., T. Gunther, & A.S. Tay (1998) Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39, 863–883.
Fan, Y. (1994) Testing the goodness-of-fit of a parametric density function by kernel method. *Econometric Theory* 10, 316–356.
Fan, Y. (1997) Goodness-of-fit tests for a multivariate distribution by the empirical characteristic function. *Journal of Multivariate Analysis* 62, 36–63.
Fan, Y. (1998) Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters. *Econometric Theory* 14, 604–621.
Fan, Y. & Q. Li (1996) Consistent model specification tests: Omitted variables and semi-parametric functional forms. *Econometrica* 64, 865–890.
Fan, Y. & Q. Li (2000) Consistent model specification tests: Nonparametric versus Bierens' ICM tests. *Econometric Theory* 16, 1016–1041.
Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* 14, 1–16.
Hall, P., J. Racine, & Q. Li (2004) Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association* 99, 1015–1026.
Hong, Y. & H. White (1996) Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133–1159.
Ichimura, H. (2000) Asymptotic Distribution of Non-parametric and Semiparametric Estimators with Data Dependent Smoothing Parameters. Manuscript, University College, London.
Kullback, S. & R.A. Leibler (1951) On information and sufficiency. *Annals of Mathematical Statistics* 22, 79–86.
Li, F. & G. Tkacz (2004) A consistent bootstrap test for conditional density functions with time-dependent data. *Journal of Econometrics*, forthcoming.
Mammen, E. (1992) *When does bootstrap work? Asymptotic results and simulations.* Springer-Verlag.
Ossiander, M. (1987) A central limit theorem under metric entropy with L_2 bracketing. *Annals of Probability* 15, 897–919.
Powell, J.L., J.H. Stock, & T.M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57, 1403–1430.
Wooldridge, J. (1992) A test for functional form against nonparametric alternatives. *Econometric Theory* 8, 452–475.
Zheng, J.X. (2000) A consistent test of conditional parametric distributions. *Econometric Theory* 16, 667–691.

APPENDIX

We first state conditions that are used to prove Theorem 2.1.

(C1) $\{y_i, x_i\}_{i=1}^n$ are i.i.d. data with a joint density $p(y,x)$. The first-order derivatives of $p(.,.)$ with respect to its continuous arguments are uniformly bounded. The marginal density $p_1(x)$ of x_i and its first-order derivatives with respect to its continuous arguments are uniformly bounded.

https://doi.org/10.1017/S0266466606060294 Published online by Cambridge University Press

(C2) (i) The parameter space Θ is a compact and convex subset of R^k. Let $\|\cdot\|$ denote the euclidean norm of $\cdot$; then $f(y|x,\theta_0)^{-1}$, $\|(\partial f(y|x,\theta))/\partial\theta\|$, $\|(\partial^2 \log f(y|x,\theta))/\partial\theta\partial\theta'\|$, and $\|(\partial \log f(y|x,\theta))/\partial\theta \times (\partial \log f(y|x,\theta))/\partial\theta'\|$ are all bounded by a nonnegative function $b(x,y)$ with $\int b(x,y)^s < \infty$ $(s = 1,2)$, where $\int$ denotes integration for the continuous variable and summation for the discrete variable. (ii) $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$ under H_0.

(C3) $w(\cdot)$ is a nonnegative, bounded, symmetric function with $\int w(v)\,dv = 1$ and $\int w(v)v^2\,dv = c(< \infty)$.

The preceding conditions are basically the same as those used in Zheng (2000).

We give the central limit theorem (CLT) of Hall (1984, Thm. 3.1) for degenerate U-statistics as a lemma here.

LEMMA A.1. *Let*

$$U_n = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} H_n(z_i, z_j)$$

be a second-order U-statistic, where $\{z_i\}_{i=1}^n$ *is i.i.d. Suppose* $E[H_n(z_i,z_j)|z_i] = 0$ *(for* $i \neq j$*,* U_n *is a degenerate U-statistic) and define* $G_n(z_1,z_2) = E[H_n(z_3,z_1) H_n(z_3,z_2)|z_1,z_2]$*. If*

$$\frac{E[G_n^2(z_i,z_j)] + n^{-1}E[H_n^4(z_i,z_j)]}{\{E[H_n^2(z_i,z_j)]\}^2} \to 0 \quad \text{as } n \to \infty, \tag{A.1}$$

then

$$U_n \Big/ \left\{ \frac{2E[H_n^2(z_i,z_j)]}{n^2} \right\} \to N(0,1) \quad \text{in distribution.}$$

In the proof presented subsequently, we will replace $\hat{h}_1,\ldots,\hat{h}_q$, $\hat{\lambda}_1,\ldots,\hat{\lambda}_r$ by their nonstochastic leading terms: $(h_1,\ldots,h_q) = (a_1^0 n^{-1/(q+4)},\ldots,a_q^0 n^{-1/(q+4)})$ and $(\lambda_1,\ldots,\lambda_r) = (b_1^0 n^{-2/(q+4)},\ldots,b_r^0 n^{-2/(q+4)})$. This will greatly simplify the arguments in the proof. By the stochastic equicontinuity result of Ichimura (2000) (see Lemma A.4, which follows), we know that the conclusion holds provided $\hat{h}_s - h_s = o_p(h_s)$ $(s = 1,\ldots,q)$ and $\hat{\lambda}_s - \lambda_s = o_p(\lambda_s)$ $(s = 1,\ldots,r)$, which are true by Theorem 3.1 of Hall et al. (2004).

Using the shorthand notations $I_{ij} = I(y_i = y_j)$, $\hat{f}_i = f(y_i|x_i,\hat{\theta})$, $f_i = f(y_i|x_i,\theta_0)$, $\hat{f}_{ij} = f(y_i|x_j,\hat{\theta})$, $f_{ij} = f(y_i|x_j,\theta_0)$, and the identity

$$\frac{1}{\hat{f}_i} = \frac{1}{f_i} + \frac{f_i - \hat{f}_i}{f_i^2} + \frac{(f_i - \hat{f}_i)^2}{f_i^2 \hat{f}_i}, \tag{A.2}$$

we can write $T_{n,\gamma} = T_{n1} + T_{n2} + T_{3n}$, where

$$T_{n1} = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} \frac{K_{\gamma,ij}}{f_i} [I_{ij} - \hat{f}_{ij}],$$

$$T_{n2} = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} \frac{K_{\gamma,ij}}{f_i^2} [I_{ij} - \hat{f}_{ij}](f_i - \hat{f}_i),$$

$$T_{n3} = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} \frac{K_{\gamma,ij}}{f_i^2 \hat{f}_i} [I_{ij} - \hat{f}_{ij}](f_i - \hat{f}_i)^2.$$

Let $f_{ij}^{(1)} = [(\partial/\partial\theta) f(y_i|x_j,\theta)]|_{\theta=\theta_0}$ and $\tilde{f}_{ij}^{(2)} = [(\partial^2/\partial\theta\partial\theta') f(y_i|x_j,\theta)]|_{\theta=\tilde{\theta}}$, where $\tilde{\theta}$ is between the line segment of $\hat{\theta}$ and θ_0. By Taylor expansion, we have

$$
\begin{aligned}
T_{n1} &= \frac{1}{n(n-1)} \sum_i \sum_{j\neq i} \frac{K_{\gamma,ij}}{f_i} [I_{ij} - f_{ij}] - \frac{1}{n(n-1)} \sum_i \sum_{j\neq i} \frac{K_{\gamma,ij}}{f_i} f_{ij}^{(1)} (\hat{\theta} - \theta_0) \\
&\quad + (\hat{\theta} - \theta_0)' \frac{1}{2n(n-1)} \sum_i \sum_{j\neq i} \frac{K_{\gamma,ij}}{f_i} \tilde{f}_{ij}^{(2)} (\hat{\theta} - \theta_0) \\
&\equiv T_{n1,1} + T_{n1,2}(\hat{\theta} - \theta_0) + (\hat{\theta} - \theta_0) T_{n1,3}(\hat{\theta} - \theta_0),
\end{aligned}
$$

where the definitions of $T_{n1,j}$ $(j = 1,2,3)$ should be apparent.

The term $T_{n1,1}$ can be written as a second-order U-statistic $(z_i = (x_i, y_i))$:

$$
T_{n1,1} = \frac{2}{n(n-1)} \sum_i \sum_{j>i} H_n(z_i, z_j),
$$

where

$$
\begin{aligned}
H_n(z_i, z_j) &= \left(\frac{1}{2}\right) \left\{ \frac{K_{\gamma,ij}}{f_i} [I_{ij} - f_{ij}] + \frac{K_{\gamma,ij}}{f_j} [I_{ij} - f_{ji}] \right\} \\
&\equiv \left(\frac{1}{2}\right) \{ J_n(z_i, z_j) + J_n(z_j, z_i) \}.
\end{aligned}
$$

It is easy to check that

$$
\begin{aligned}
E[J_n(z_i,z_j)|z_i] &= E\{K_{\gamma,ij}[I_{ij} - f(y_i|x_j)] f_i^{-1} | z_i\} \\
&= f_i^{-1} \{ E[K_{\gamma,ij} I_{ij} | z_i] - E[K_{\gamma,ij} f(y_i|x_j) | z_i] \} \\
&= f_i^{-1} \left\{ \sum_{y_j} \int p(y_j, x_j) K_{\gamma,ij} I_{ij}\, dx_j - \int K_{\gamma,ij} p_1(x_j) f(y_i|x_j)\, dx_j \right\} \\
&= f_i^{-1} \left\{ \int K_{\gamma,ij} [p(y_i, x_j) - p_1(x_j) f(y_i|x_j)]\, dx_j \right\} = 0 \\
&\quad \text{because } p(y_i, x_j) - p_1(x_j) f(y_i|x_j) = 0.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
E[J_n(z_j,z_i)|z_i] &= E\{K_{\gamma,ij}[I_{ij} - f_{ji}] f_j^{-1} | z_i\} = E[K_{\gamma,ij} I_{ij} f_j^{-1} | z_i] - E[K_{\gamma,ij} f_{ji} f_j^{-1} | z_i] \\
&= \sum_{y_j} \int p(y_j, x_j) K_{\gamma,ij} I_{ij} f(y_j|x_j)^{-1}\, dx_j - \sum_{y_j} \int K_{\gamma,ij} f_j^{-1} f_{ji}\, dx_j \\
&= \int p(y_i, x_j) K_{\gamma,ij} f(y_i|x_j)^{-1}\, dx_j - \int K_{\gamma,ij} p_1(x_j)\, dx_j \sum_{y_j} f(y_j|x_i) \\
&\quad (\text{because } f_{ij}^{-1} p(y_i, x_j) = p_1(x_j)) \\
&= \int K_{\gamma,ij} [p(y_i, x_j) f(y_i|x_j)^{-1} - p_1(x_j)]\, dx_j = 0 \\
&\quad \left(\text{because } \sum_{y_j} f(y_j|x_i) = 1 \text{ and } p(y_i, x_j) f(y_i|x_j) = p_1(x_j) \right).
\end{aligned}
$$

https://doi.org/10.1017/S0266466606060294 Published online by Cambridge University Press

Thus, $E[H_n(z_i, z_j)|z_i] = 0$ and $T_{n1,1}$ is a degenerate U-statistic.

$$
\begin{aligned}
E[H_n(z_i, z_j)^2] &= E[J_n(z_i, z_j)^2] = E[K^2_{\gamma,ij}(I_{ij} - f_{ij})^2/f_i^2] \\
&= E\{E[K^2_{\gamma,ij}(I_{ij} + f_{ij}^2 - 2I_{ij}f_{ij})/f_i^2|z_i, x_j]\} \\
&= E[K^2_{\gamma,ij}(f_{ij} + f_{ij}^2 - 2f_{ij}^2)/f_i^2] = E[K^2_{\gamma,ij}f_{ij}(1 - f_{ij})/f_i^2] \\
&= \sum_{y_i}\sum_{y_j}\sum_{x_i^d}\sum_{x_j^d}\iint K^2_{n,ij}f_{ij}(1-f_{ij})f_i^{-2}p(y_i,x_i)p(y_j,x_j)\,dx_i^c\,dx_j^c \\
&= \sum_{y_i}\sum_{y_j}\sum_{x_i^d}\iint K^2_{n,ij}f_{ij}(1-f_{ij})f_i^{-2}p(y_i,x_i)p(y_j,x_j)\,dx_i^c\,dx_j^c \\
&\quad + \sum_{y_i}\sum_{y_j}\sum_{x_i^d}\sum_{x_j^d\neq x_i^d}\iint K^2_{n,ij}f_{ij}(1-f_{ij})f_i^{-2}p(y_i,x_i)p(y_j,x_j)\,dx_i^c\,dx_j^c \\
&= (h_1\dots h_q)^{-1}\sum_{y_i}\sum_{y_j}\sum_{x_i^d}\iint W^2(v)(1-f_i)f_i^{-1}p(y_i,x_i)p(y_j,x_i)\,dx_i^c\,dv \\
&\quad + O\left((h_1\dots h_q)^{-1}\left(\sum_{j=1}^q h_j^2 + \sum_{j=1}^r \lambda_j\right)\right) \\
&= (h_1\dots h_q)^{-1}\left\{\left[\int W^2(v)\,dv\right]E[(1-f_i)f_i^{-1}p_1(x_i)] + O_p(\eta_n)\right\} \\
&= (h_1\dots h_q)^{-1}\left[\left(\frac{1}{2}\right)\sigma_0^2 + O_p(\eta_n)\right],
\end{aligned}
$$

where $\eta_n = \sum_{j=1}^q h_j^2 + \sum_{j=1}^r \lambda_j$, we have used $\sum_{y_j} p(y_j, x_i) = p_1(x_i)$, $K_{\gamma,ij} = W_{h,ij}L_{\lambda,ij}$, and $L_{\lambda,ij} = O(\sum_{s=1}^r \lambda_s)$ if $x_i^d \neq x_j^d$.

Therefore, we have

$$
V_{n,\gamma} \stackrel{def}{=} E\{[n(n-1)(h_1\dots h_q)]^{1/2}T_{n1}\}^2 = \frac{2(h_1\dots h_q)}{n(n-1)}\sum_i\sum_{j\neq i}E[H_n^2(z_i,z_j)] = \sigma_0^2 + o(1). \tag{A.3}
$$

Equation (A.3) implies that $\{E[H_n^2(z_i, z_j)]\}^{-1} = O(h_1 \dots h_q)$. Similarly, one can show that $E[H_n^4(z_i, z_j)] = O((h_1 \dots h_q)^{-3})$. Define $G_n(z_1, z_2) = E[H_n(z_3, z_2) H_n(z_3, z_1)|z_1, z_2]$. One can show that $E[G_n^2(z_i, z_j)] = O((h_1 \dots h_q)^{-1})$. Thus, equation (A.1) becomes

$$
\begin{aligned}
&O((h_1\dots h_q)^2)\{O((h_1\dots h_q)^{-1}) + n^{-1}O((h_1\dots h_q)^{-3})\} \\
&\quad = O((h_1\dots h_q) + n^{-1}(h_1\dots h_q)^{-1}) = o(1).
\end{aligned}
$$

Thus by Lemma A.1 we know that

$$
n(h_1\dots h_q)^{1/2}T_{n1,1}\big/\sqrt{V_{n,\gamma}} \to N(0,1) \quad \text{in distribution.} \tag{A.4}
$$

https://doi.org/10.1017/S0266466606060294 Published online by Cambridge University Press

Define

$$\hat{V}_{n,\gamma} = \frac{2(h_1 \dots h_q)}{n(n-1)} \sum_i \sum_{j \neq i} \hat{H}_n(z_i, z_j)^2, \tag{A.5}$$

where $\hat{H}_n(z_i, z_j)$ is defined in the same way as $H_n(z_i, z_j)$ except that θ_0 is replaced by $\hat{\theta}$. Applying Lemma 3.1 of Powell, Stock, and Stoker (1989) or Lemma 1 of Zheng (2000), it is straightforward to show that $\hat{V}_{n,\hat{\gamma}} - V_{n,\gamma} = o_p(1)$. Thus, we have

$$n(h_1 \dots h_q)^{1/2} T_{n1,1} \Big/ \sqrt{\hat{V}_{n,\gamma}} \to N(0,1) \quad \text{in distribution.} \tag{A.6}$$

Applying Taylor expansion to T_{n2}, i.e., using $\hat{f}_{ij} = f_{ij} + \tilde{f}_{ij}(\hat{\theta} - \theta_0)$ and $\hat{f}_i - f_i = f_i^{(1)}(\hat{\theta} - \theta_0) + (\frac{1}{2})(\hat{\theta} - \theta_0)' \tilde{f}_i^{(2)}(\hat{\theta} - \theta_0)$, we obtain

$$\begin{aligned} T_{n2} &= -\frac{1}{n(n-1)} \sum_i \sum_{j \neq i} \frac{K_{\gamma,ij}}{f_i^2} [I_{ij} - f_{ij}] f_i^{(1)}(\hat{\theta} - \theta_0) \\ &\quad + (\hat{\theta} - \theta_0)' \left\{ \frac{1}{2n(n-1)} \sum_i \sum_{j \neq i} \frac{K_{\gamma,ij}}{f_i^2} [(I_{ij} - f_{ij}) \tilde{f}_i^{(2)} + (f_i^{(1)})' \tilde{f}_{ij}^{(1)}] \right\} (\hat{\theta} - \theta_0) \\ &\equiv -T_{n2,1}(\hat{\theta} - \theta_0) + (\hat{\theta} - \theta_0)' T_{n2,2}(\hat{\theta} - \theta_0), \end{aligned} \tag{A.7}$$

where $T_{n2,1} = 1/(n(n-1)) \sum_i \sum_{j \neq i} K_{\gamma,ij}(I_{ij} - f_{ij}) f_i^{(1)}/f_i^2$ and $T_{n2,2} = 1/(2n(n-1)) \sum_i \sum_{j \neq i} (K_{\gamma,ij}/f_i^2)[(I_{ij} - f_{ij}) \tilde{f}_i^{(2)} + (f_i^{(1)})' \tilde{f}_{ij}^{(1)}]$.

Lemma A.2, which follows, shows that $T_{n1,2} = O_p(n^{-1/2})$ and $T_{n2,1} = O_p(n^{-1/2})$, and Lemma A.3 shows that $T_{n1,3} = O_p(1)$, $T_{n2,2} = O_p(1)$, and $T_{n3} = O_p(n^{-1})$. These results together with $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$ lead to

$$T_{n,\gamma} = T_{n1,1} + O_p(n^{-1}). \tag{A.8}$$

Expressions (A.6) and (A.8) together complete the proof of Theorem 2.1. ■

LEMMA A.2.

(i) $T_{n1,2} = O_p(n^{-1/2})$.
(ii) $T_{n2,1} = O_p(n^{-1/2})$.

Proof of (i).

$$T_{n1,2} = -\frac{1}{n(n-1)} \sum_i \sum_{j \neq i} K_{\gamma,ij} f_{ij}^{(1)}/f_i.$$

First note that $E[T_{n1,2}] = 0$ because

$$\begin{aligned} E[K_{\gamma,ij} f_{ij}^{(1)}/f_i] &= E\{K_{\gamma,ij} E(f_{ij}^{(1)}/f_i | x_i, x_j)\} \\ &= E\left\{K_{\gamma,ij} \frac{\partial}{\partial \theta} \sum_{y_j} f(y_i | x_j, \theta)\right\} = E\left\{K_{\gamma,ij} \frac{\partial}{\partial \theta}[1]\right\} = 0. \end{aligned}$$

Hence,

$$E\{[T_{n1,2}]^2\} = \frac{1}{n^2(n-1)^2} \sum_i \sum_{j \neq i} \sum_{i'} \sum_{j' \neq i'} E[K_{\gamma,ij} f_{ij}^{(1)} f_i^{-1} K_{\gamma,i'j'} f_{i'j'}^{(1)} f_{i'}^{-1}].$$

The preceding expression is zero if i, j, i', j' all take different values (because $E[K_{\gamma,ij} f_{ij}^{(1)}/f_i] = 0$). Therefore, for $E\{[T_{n1,2}]^2\}$ to be nonzero, we must have either (i) i, j, i', j' take three different values or (ii) i, j, i', j' take two different values. For these two cases it is easy to show that

$$E\{[T_{n1,2,(i)}]^2\} = \frac{1}{n^2(n-1)^2} O(n^3) = O(n^{-1});$$

$$E\{[T_{n1,2,(ii)}]^2\} = \frac{1}{n^2(n-1)^2} O(n^2(h_1 \dots h_q)^{-1}) = o(n^{-1}).$$

Hence, $E\{[T_{n1,2}]^2\} = O(n^{-1})$, and consequently, $T_{n1,2} = O_p(n^{-1/2})$. ■

Proof of (ii).

$$T_{n2,1} = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} K_{\gamma,ij}(I_{ij} - f_{ij}) f_{ij}^{(1)}/f_i = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} A_{1n}(z_i, z_j),$$

where

$$A_{1n}(z_i, z_j) = K_{\gamma,ij} f_{ij}^{(1)}(I_{ij} - f_{ij})/f_i \equiv A_{1n,1}(z_i, z_j) - A_{1n,2}(z_i, z_j).$$

$$E[A_{1n,1}(z_i, z_j)|z_i, x_j] = E[K_{\gamma,ij} f_{ij}^{(1)} I_{ij}/f_i | z_i, z_j] = \sum_{y_i} K_{\gamma,ij} f_{ij}^{(1)} I(y_i = y_j) f(y_j|x_j)/f_i$$

$$= K_{\gamma,ij} f_{ij}^{(1)} f(y_i|x_j)/f_i \equiv A_{1n,2}(z_i, z_j).$$

Hence, $E[A_{1n}(z_i, z_j)] = E[A_{1n,1}(z_i, z_j)] - E[A_{1n,2}(z_i, z_j)] = 0$.

One can write $T_{n2,1} = [2/n(n-1)] \sum_i \sum_{j>i} V_{1n}(z_i, z_j)$ as a second-order U-statistic, where $V_{1n}(z_i, z_j) = (\frac{1}{2})[A_{1n}(z_i, z_j) + A_{1n}(z_j, z_i)]$.

$$\begin{aligned}
E[V_{1n}(z_i, z_j)|z_i] &= \left(\frac{1}{2}\right)\{E[A_{1n}(z_i, z_j)|z_i] + E[A_{1n}(z_j, z_i)|z_i]\} \\
&= 0 + \left(\frac{1}{2}\right) E[K_{\gamma,ij} f_{ji}^{(1)}(I_{ij} - f_{ji})/f_j | z_i] \\
&= \left(\frac{1}{2}\right) \sum_{y_j} \int K_{\gamma,ij} f_{ji}^{(1)}(I_{ij} - f_{ji}) p(y_j, x_j) f_j^{-1} \, dx_j \\
&= \left(\frac{1}{2}\right) \int K_{\gamma,ij} f_i^{(1)} p_1(x_j) \, dx_j - \left(\frac{1}{2}\right) \sum_{y_j} \int K_{\gamma,ij} f_{ji}^{(1)} f_{ji} p_1(x_j) \, dx_j \\
&= \left(\frac{1}{2}\right) f_i^{(1)} p_1(x_i) - \left(\frac{1}{2}\right) \sum_{y_j} f_{ji}^{(1)} f_{ji} p_1(x_i) + (s.o.) \\
&= \left(\frac{1}{2}\right) p_1(x_i)[f_i^{(1)} - E(f_{ji}^{(1)}|x_i)] + (s.o.) \\
&= \left(\frac{1}{2}\right) p_1(x_i)[f_i^{(1)} - E(f_i^{(1)}|x_i)] + (s.o.) \equiv v_{1i} + (s.o.),
\end{aligned}$$

where in the preceding expression, $A_i = B_i + (s.o.)$ means that $\sum_i A_i = \sum_i B_i + o_p(\sum_i B_i)$, i.e., $\sum_i B_i$ is the leading term of $\sum_i A_i$. Here $v_{1i} = (\frac{1}{2})p_1(x_i)[f_i^{(1)} - E(f_i^{(1)}|x_i)]$, and we have used $E[f_{ji}^{(1)}|x_i] = E[f_i^{(1)}|x_i] = \sum_y f(y|x_i)^2$.

Using the H-decomposition, we have

$$T_{n2,1} = 0 + \frac{2}{n}\sum_i E[H_{1n}(z_i, z_j)|z_i] + (s.o.)$$

$$= \frac{2}{n}\sum_i v_{1i} + (s.o.) = O_p(n^{-1/2}) \quad \text{because } E(v_{1i}) = 0. \qquad \blacksquare$$

LEMMA A.3.

(i) $T_{n1,3} = O_p(1)$.
(ii) $T_{n2,2} = O_p(1)$.
(iii) $T_{n3} = O_p(n^{-1})$.

Proof of (i). Here $T_{n1,3} = [1/2n(n-1)]\sum_i \sum_{j\neq i} K_{\gamma,ij}\tilde{f}_{ij}^{(2)}/f_i$. By assumption (C2) ($b(.,.)$ is the bound function for $f^{(2)}(\cdot)$):

$$E[\|T_{n1,3}\|] \leq CE[K_{\gamma,ij}b(x_i, y_j)]$$

$$= C\sum_{y_j}\iint K_{\gamma,ij}b(y_i, x_j)p(y_i, x_i)p_1(x_j)\,dx_i\,dx_j$$

$$= C\sum_{x_i^d}\sum_{x_j^d}\sum_{y_i}\iint W(v)L(x_i^d, x_j^d, \lambda)b(y_i, x_i^c + hv, x_j^d)$$

$$\times p(y_i, x_i)p_1(x_i^c + hv, x_j^d)\,dx_i^c\,dv$$

$$= C\sum_{x_i^d}\sum_{y_i}\int b(y_i, x_i)p(y_i, x_i)p_1(x_i)\,dx_i^c + o_p(1)$$

$$= CE[p_1(x_i)b(y_i, x_i)] + o(1) = O(1),$$

which implies that $T_{n1,3} = O_p(1)$. $\blacksquare$

Proof of (ii). It is similar to the proof of (i) and is thus omitted here.

Proof of (iii). Using $\hat{f}_i - f_i = f^{(1)}(y_i|x_i, \tilde{\theta})(\hat{\theta} - \theta_0)$, where $\tilde{\theta}$ is between the line segment of $\hat{\theta}$ and θ_0, we have $T_{n3} = (\hat{\theta} - \theta_0)'T_{n3,1}(\hat{\theta} - \theta_0)$, where

$$T_{n3,1} = \frac{1}{n(n-1)}\sum_i\sum_{j\neq i} K_{\gamma,ij}I_{ij}[I_{ij} - \hat{f}_{ij}]\tilde{f}_i^{(1)}(\tilde{f}_i^{(1)})'/(f_i^2\hat{f}_i)$$

$$= \frac{1}{n(n-1)}\sum_i\sum_{j\neq i} K_{\gamma,ij}I_{ij}[I_{ij} - f_{ij}]f_i^{(1)}(f_i^{(1)})'/(f_i^3) + (s.o.)$$

$$\equiv T_{n3,1,0} + (s.o.).$$

https://doi.org/10.1017/S0266466606060294 Published online by Cambridge University Press

It is easy to show that $E[\|T_{3n,1,0}\|] = O(1)$. Hence, $T_{3n,1,0} = O_p(1)$, which implies $T_{3n,1} = O_p(1)$ and $T_{3n} = O_p(n^{-1})$ because $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$. ■

LEMMA A.4. *(Ichimura).*

$$J_{n,\hat{\gamma}} - J_{n,\gamma} = o_p(1),$$

where $\gamma = (h_1,\dots,h_q,\lambda_1,\dots,\lambda_r)$ *with* $h_s = a_s^0 n^{-1/(q+4)}$ $(s = 1,\dots,q)$, $\lambda_s = b_s^0 n^{-2/(q+4)}$ $(s = 1,\dots,r)$, *and* $a_s^0 > 0$ *and* $b_s^0 \geq 0$ *are uniquely defined constants as given in Hall et al. (2004).*

Ichimura (2000) has proved a general result that includes Lemma A.4 as a special case. Here, we provide an alternative proof for Lemma A.4 using a simple tightness argument (e.g., Mammen, 1992). Our proof consists of two parts: (i) $(n\hat{h}_1,\dots,\hat{h}_q)^{1/2} T_{n,\hat{\gamma}} - (nh_1,\dots,h_q)^{1/2}T_{n,\gamma} = o_p(1)$ under H_0; (ii) $\hat{V}_{n,\hat{\gamma}} - \hat{V}_{n,\gamma} = o_p(1)$. Because the proofs are similar, we only provide the proof for (i).

Proof of (i). Writing $\hat{h}_s = \hat{a}_s n^{-1/(q+4)}$ and $\hat{\lambda}_s = \hat{b}_s n^{-2/(q+4)}$, by Theorem 3.1 of Hall et al. (2004), we know that $\hat{h}_s/h_s^0 - 1 \to 0$ and $\hat{\lambda}_s/\lambda_s^0 - 1 \to 0$ (in probability). This implies that $\hat{a}_s \to a_s^0$ and $\hat{b}_s \to b_s^0$ in probability. Let $\mathcal{C} = \prod_{s=1}^{q}[a_{1s},a_{2s}] \times \prod_{t=1}^{r}[b_{1t},b_{2t}]$, where a_{js} and b_{jt} $(j = 1,2)$ are some positive constants with $a_{1s} < a_s^0 < a_{2s}$ $(s = 1,\dots,q)$ and $b_{1t} < b_t^0 < b_{2t}$ $(t = 1,\dots,r)$. Denote $c = (a_1,\dots,a_q,b_1,\dots,b_r)$, $c_0 = (a_1^0,\dots,a_q^0,b_1^0,\dots,b_r^0)$, $\hat{c} = (\hat{a}_1,\dots,\hat{a}_q,\hat{b}_1,\dots,\hat{b}_r)$. Then Lemma A.5, which follows, shows that $A_n(c) \equiv n(h_1 \dots h_q)^{1/2}T_{n,\gamma}$ (with $h_s = a_s n^{-1/(q+4)}$ and $\lambda_s = b_s n^{-2/(q+4)}$) is tight in $c \in \mathcal{C}$.

Define $B_n(c) = A_n(c) - A_n(c_0)$. Then (i) becomes $B_n(\hat{c}) = o_p(1)$; i.e., we want to show that, for all $\epsilon > 0$

$$\lim_{n\to\infty} \Pr[|B_n(\hat{c})| < \epsilon] = 1. \tag{A.9}$$

For any $\delta > 0$, denote the δ-ball centered at c_0 by $C_\delta = \{c : \|c - c_0\| \leq \delta\}$, where $\|\cdot\|$ denotes the euclidean norm of a vector. By Lemma A.5 we know that $A_n(\cdot)$ is tight. By the Arzela–Ascoli theorem (see Billingsley, 1968, Thm. 8.2, p. 55) we know that tightness implies the following stochastic equicontinuous condition: for all $\epsilon > 0$, $\eta_1 > 0$, there exist a δ $(0 < \delta < 1)$ and an N_1 such that

$$\Pr\left[\sup_{\|c'-c\|<\delta} |A_n(c') - A_n(c)| > \epsilon\right] < \eta_1 \tag{A.10}$$

for all $n \geq N_1$.

Expression (A.10) implies that

$$\Pr[|B_n(\hat{c})| > \epsilon, \hat{c} \in C_\delta] \leq \Pr\left[\sup_{c\in C_\delta} |B_n(c)| > \epsilon\right] < \eta_1 \tag{A.11}$$

for all $n \geq N_1$.

Also, from $\hat{c} \to c_0$ in probability we know that for all $\eta_2 > 0$, and for the δ given previously, there exists an N_2 such that

$$\Pr[\hat{c} \notin C_\delta] \equiv \Pr[\|\hat{c} - c_0\| > \delta] < \eta_2 \tag{A.12}$$

for all $n \geq N_2$.

Therefore,

$$\Pr[|B_n(\hat{c})| > \epsilon] = \Pr[|B_n(\hat{c})| > \epsilon, \hat{c} \in C_\delta] + \Pr[|B_n(\hat{c})| > \epsilon, \hat{c} \notin C_\delta] < \eta_1 + \eta_2 \quad \textbf{(A.13)}$$

for all $n \geq \max\{N_1, N_2\}$ by (A.11) and (A.12), where we have also used the fact that $\{|B_n(\hat{c})| > \epsilon, \hat{c} \notin C_\delta\}$ is a subset of $\{\hat{c} \notin C_\delta\}$.

Equation (A.13) is equivalent to (A.9). This completes the proof of (i). ■

LEMMA A.5. *Let $A_n(c) = n(h_1 \dots h_q)^{1/2} T_{n,\gamma}$, where $\gamma = (h, \lambda)$, $h_s = a_s n^{-1/(q+4)}$, $\lambda_s = b_s n^{-2/(q+4)}$, $c = (a_1, \dots, a_q, b_1, \dots, b_r)$, $c_s \in [C_{1s}, C_{2s}]$ with $0 < C_{1s} < C_{2s} < \infty$ $(s = 1, \dots, q + r)$.*

Then the stochastic process $A_n(c)$ indexed by c is tight under the sup-norm.

Proof. Write $K_{\gamma,ij}$ as $(h_1 \dots h_q)^{-1} K_{c,ij}$ with $h_s = a_s n^{1/(q+4)}$ and $\lambda_s = b_s n^{-2/(q+4)}$, where $K_{c,ij} = W((X_j - X_i)/h) L(X_j^d, X_i^d, \lambda)$. Also, denote by $\delta = q/(4+q)$, $C_1 = (a_1, \dots, a_q)'$ and $C_2 = (b_1, \dots, b_r)'$, $\bar{C}_1 = \prod_{s=1}^q a_s$ and $\bar{C}_2 = \prod_{s=1}^r b_s$. Then we have $(h_1 \dots h_q)^{-1} K_{c,ij} = \bar{C}_1 n^\delta W_{C_1,ij} L_{C_2,ij}$. Also note that $|L_{C_2',ij} - L_{C_2,ij}| \leq \sum_{s=1}^r |b_s - b_s'| \leq r\|C_2 - C_2'\|$; we have

$$\begin{aligned}
&|(h_1' \dots h_q')^{1/2} K_{\gamma',ij} - (h_1 \dots h_q)^{1/2} K_{\gamma,ij}| \\
&\quad = |(h_1' \dots h_q')^{-1/2} K_{C',ij} - (h_1 \dots h_q)^{-1/2} K_{C,ij}| \\
&\quad = |n^\delta \{(\bar{C}_1')^{-1} W_{C_1',ij} L_{C_2',ij} - \bar{C}_1^{-1} W_{C_1,ij} L_{C_2,ij}\}| \\
&\quad = |n^\delta \{(\bar{C}_1')^{-1} W_{C_1',ij} [L_{C_2',ij} - L_{C_2,ij}] + [(\bar{C}_1')^{-1} W_{C_1',ij} - \bar{C}_1^{-1} W_{C_1,ij}] L_{C_2,ij}\}| \\
&\quad \leq D_1 \left\{ (h_1' \dots h_q')^{-1} W_{C_1',ij} \|C_2' - C_2\| + (h_1 \dots h_q)^{-1} G\left(\frac{x_j - x_i}{h}\right) \|C_1' - C_1\| \right\}, \quad \textbf{(A.14)}
\end{aligned}$$

where $D_1 > 0$ is a finite constant. In the last equality we used $|L_{C_2,ij}| \leq 1$ and assumption (C3). Also, we replaced one of the $(\bar{C}_1')^{-1/2}$ by $\bar{C}_1^{-1/2}$ because $a_s \in [C_{1s}, C_{2s}]$ are all bounded from above and below. The difference can be absorbed into D_1.

By noting that $A_n(c') - A_n(c)$ is a degenerate U-statistic, and using (A.14), we have

$$\begin{aligned}
&E\{[A_n(c') - A_n(c)]^2\} \\
&\quad = E\{(I_{ij} - f_{ij})^2 f_i^{-2} [(h_1' \dots h_q')^{-1/2} K_{c',ij} - (h_1 \dots h_q)^{-1/2} K_{c,ij}]^2\} \\
&\quad \sim E\{[(h_1' \dots h_q')^{-1/2} K_{c',ij} - (h_1 \dots h_q)^{-1/2} K_{c,ij}]^2\} \\
&\quad \leq 4E\left\{ \left[(h_1' \dots h_q')^{-1} W^2\left(\frac{x_j - x_i}{h'}\right) \|C_2' - C_2\|^2 \right.\right. \\
&\qquad \left.\left. + (h_1 \dots h_q)^{-1} G\left(\frac{x_j - x_i}{h}\right) \|C_1' - C_1\|^2 \right] \right\} \\
&\quad \leq 4D_1 \left\{ \left[\iint f(x_i) f(x_i + hv) W^2(v)\, dx_i\, dv \right] \|C_2' - C_2\|^2 \right. \\
&\qquad \left. + \left[\iint f(x_i) f(x_i + w) G(w)^2\, dx_i\, dw \right] \|C_1' - C_1\|^2 \right\} \\
&\quad \leq 4D_1 \sup_x f(x) \left\{ \left[\int W^2(v)\, dv \right] \|C_2' - C_2\|^2 + \left[\int G(w)^2\, dw \right] \|C_1' - C_1\|^2 \right\} \\
&\quad \leq D\|C' - C\|^2, \quad \textbf{(A.15)}
\end{aligned}$$

where in the preceding expression $A \sim B$ means A and B having the same order of magnitude and D is a finite positive constant. Therefore, $A_n(\cdot)$ (hence, $B_n(\cdot)$) is tight by Theorem 3.1 of Ossiander (1987). ■

Proof of Theorem 2.3. We will provide a proof for the discrete dependent variable case. The continuous case is similar. To prove (13), similar to the decomposition of $T_{n,\gamma}$, we decompose $T^*_{n,\hat{\gamma}}$ as $T^*_{n,\hat{\gamma}} = T^*_{n1} + T^*_{n2} + T^*_{n3}$, where the definitions of T^*_{nj} are similar to those of T_{nj} with the proper changes; i.e., $y_i, \hat{\theta}, \gamma$ need to be changed to $y_i^*, \hat{\theta}^*, \hat{\gamma}$. We further decompose T^*_{n1} to $T^*_{n1} = T^*_{n1,1} + T^*_{n1,2}(\hat{\theta}^* - \hat{\theta}) + (\hat{\theta}^* - \hat{\theta})' T^*_{n1,3}(\hat{\theta}^* - \hat{\theta})$, where the definitions of $T^*_{n1,j}$ are similar to $T_{n1,j}$ with the proper changes ($j = 1,2,3$).

The term $T^*_{n1,1}$ can be written as a second-order U-statistic ($z_i^* = (x_i^*, y_i^*) = (x_i, y_i^*)$):

$$T^*_{n1,1} = \frac{2}{n(n-1)} \sum_i \sum_{j>i} H_n^*(z_i^*, z_j^*),$$

where

$$H_n^*(z_i^*, z_j^*) = \left(\frac{1}{2}\right)\left\{ \frac{K_{\hat{\gamma},ij}}{f_i^*}[I_{ij}^* - f_{ij}^*] + \frac{K_{\hat{\gamma},ij}}{f_j^*}[I_{ij}^* - f_{ji}^*] \right\}$$

$$\equiv \left(\frac{1}{2}\right)\{J_n^*(z_i^*, z_j^*) + J_n^*(z_j^*, z_i^*)\}$$

with $f_i^* = f(y_i^*|x_i, \hat{\theta})$, $f_{ij}^* = f(y_i^*|x_j, \hat{\theta})$, and $I_{ij}^* = I(y_i^* = y_j^*)$.

It is easy to check that

$$\begin{aligned} E^*[J_n^*(z_i^*, z_j^*)|z_i^*] &= K_{\hat{\gamma},ij} f_i^{*-1} E^*\{[I_{ij}^* - f(y_i^*|x_j)]|z_i^*\} \\ &= f_i^{*-1} K_{\hat{\gamma},ij} \left\{ \sum_{y_j^*} f(y_j^*|x_j, \hat{\theta}) I(y_i^* = y_j^*) - f(y_i^*|x_j, \hat{\theta}) \right\} \\ &= f_i^{*-1} K_{\hat{\gamma},ij} \{ f(y_i^*|x_j, \hat{\theta}) - f(y_i^*|x_j, \hat{\theta})\} = 0. \end{aligned}$$

Similarly,

$$\begin{aligned} E^*[J_n^*(z_j^*, z_i^*)|z_i^*] &= K_{\hat{\gamma},ij} E^*\{[I_{ij}^* - f_{ji}^*] f_j^{*-1}|z_i^*\} \\ &= K_{\hat{\gamma},ij}\{E^*[I_{ij}^* f_j^{*-1}|z_i^*] - E^*[f_{ji}^* f_j^{*-1}|z_i^*]\} \\ &= K_{\hat{\gamma},ij}\left\{ \sum_{y_j^*} I(y_i^* = y_j^*) - \sum_{y_j^*} f^*(y_j^*|x_i, \hat{\theta}) \right\} = K_{\hat{\gamma},ij}\{1 - 1\} = 0. \end{aligned}$$

Hence, $E^*[H_n^*(z_i^*, z_j^*)|z_i^*] = 0$. Thus, conditional on the random sample $\{x_i, y_i\}_{i=1}^n$, $T^*_{n1,1}$ is a degenerate U-statistic.

Denote $U^*_{n,ij} = [2/(n(n-1))] H_n^*(z_i^*, z_j^*)$ and define $U_n^* = [2/(n(n-1))] \sum_i \sum_{j>i} H_n^*(z_i^*, z_j^*) \equiv T^*_{n1,1}$. We apply the CLT of de Jong (1996) for generalized quadratic forms to derive the asymptotic distribution of $U_n^*|\{x_i, y_i\}_{i=1}^n$. The reason for using de Jong's central CLT instead of the one in Hall (1984) is that in the bootstrap world, the function $H_n^*(z_i^*, z_j^*)$ depends on i and j, because $z_i^* = (x_i, y_i^*)$. By de Jong (1996, Prop. 3.2) we know that $U_n^*/S_n^* \to N(0,1)$ in distribution in probability if G_I^*, G_{II}^*, and G_{IV}^* are all $o_p(S_n^{*4})$, where $S_n^{*2} = E^*[U_n^{*2}]$, $G_I^* = \sum_i \sum_{j>i} E^*[U^{*4}_{n,ij}]$,

https://doi.org/10.1017/S0266466606060294 Published online by Cambridge University Press

$G_{II}^* = \sum_i \sum_{j>i} \sum_{l>j>i} [E^*(U_{n,ij}^{*2} U_{n,il}^{*2}) + E^*(U_{n,ji}^{*2} U_{n,jl}^{*2}) + E^*(U_{n,li}^{*2} U_{n,lj}^{*2})]$, and $G_{IV}^* = (\frac{1}{2}) \sum_i \sum_{j>i} \sum_s \sum_{t>s} E^*(U_{n,is}^{*2} U_{n,sj}^{*2} U_{n,ti}^{*2} U_{n,js}^{*2})$.

Now,

$$E^*[H_n^{*2}(z_i^*, z_j^*)] = E^*[J_n^{*2}(z_i^*, z_j^*)] = K_{\hat{\gamma},ij}^2 E^*[(I_{ij}^* - f_{ij}^*)^2 / f_i^{*2}]$$

$$= K_{\hat{\gamma},ij}^2 E^*\{f_i^{*-2} E^*[(I_{ij}^* + f_{ij}^{*2} - 2I_{ij}^* f_{ij}^*) | y_i^*]\}$$

$$= K_{\hat{\gamma},ij}^2 E^*\{f_i^{*-2}(f_{ij}^* + f_{ij}^{*2} - 2f_{ij}^{*2})\}$$

$$= K_{\hat{\gamma},ij}^2 E^*\{[f_{ij}^*(1 - f_{ij}^*)]/f_i^{*2}\} = K_{\hat{\gamma},ij}^2 \sum_{y_i^*} \{[f_{ij}^*(1 - f_{ij}^*)]/f_i^{*2}\}.$$

Hence, $S_n^{*2} = [4/n^2(n-1)^2] \sum_i \sum_{j>i} E^*[H_n(z_i^*, z_j^*)^2] = [4/n^2(n-1)^2] \sum_i \sum_{j>i} K_{\hat{\gamma},ij}^2 \sum_{y_i^*} \{[f_{ij}^*(1 - f_{ij}^*)]/f_i^{*2}\}$. By using a proof similar to the proof of Lemma A.4, one can show that S_n^{*2} has the same order as $\bar{S}_n^{*2}$ where $\bar{S}_n^{*2}$ is the same as S_n^{*2} except that $\hat{\gamma}$ is replaced by γ. Hence we only need to establish the order of $\bar{S}_n^{*2}$. Because discrete regressors do not affect its order, for clarity, we will establish the order of $\bar{S}_n^{*2}$ for the case with continuous regressors only. We have

$$E|\bar{S}_n^{*2}| = n^{-2}(\hat{h}_1 \ldots \hat{h}_q)^{-2} \Bigg[\sum_{y_i^*} \iint W^2\left(\frac{x_i - x_j}{\hat{h}}\right) \frac{f(y_i^*|x_j, \theta_0)[1 - f(y_i^*|x_j, \theta_0)]}{f^2(y_i^*|x_i, \theta_0)}$$

$$\times p_1(x_i) p_1(x_j)\, dx_i\, dx_j + o(1) \Bigg]$$

$$= n^{-2}(\hat{h}_1 \ldots \hat{h}_q)^{-1} \left[\int W^2(u)\, du \sum_{y_i^*} \int \frac{[1 - f(y_i^*|x_i, \theta_0)] p_1^2(x_i)}{f(y_i^*|x_i, \theta_0)}\, dx_i + o(1) \right]$$

$$= n^{-2}(\hat{h}_1 \ldots \hat{h}_q)^{-1} [C + o(1)],$$

where $C > 0$ is a constant, which implies that $(\bar{S}_n^{*2})^{-1} = O_p(n^2(\hat{h}_1 \ldots \hat{h}_q))$. Hence, $1/S_n^{*2} = O_p(n^2(\hat{h}_1 \ldots \hat{h}_q))$ and $1/S_n^{*4} = O_p(n^4(\hat{h}_1 \ldots \hat{h}_q)^2)$.

Next, $E^*[H_n^{*4}(z_i^*, z_j^*)] = E[J_n^{*4}(z_i^*, z_j^*)] = K_{\hat{\gamma},ij}^4 E^*[(I_{ij}^* - f_{ij}^*)^4 / f_i^{*4}]$. Similar to S_n^{*2}, one can show that

$$G_I^* = \frac{16}{n^4(n-1)^4} \sum_i \sum_{j>i} E^*[U_{n,ij}^{*4}] = \frac{16}{n^4(n-1)^4} \sum_i \sum_{j>i} K_{\hat{\gamma},ij}^4 E^*[(I_{ij}^* - f_{ij}^*)^4 / f_i^{*4}]$$

$$= O_p(n^{-6}(\hat{h}_1 \ldots \hat{h}_q)^{-3}),$$

given that $n^{-8} \sum_i \sum_{j>i} K_{\hat{\gamma},ij}^4 = O_p(n^{-6}(\hat{h}_1 \ldots \hat{h}_q)^{-3})$.

From the preceding calculation it should be apparent that the probability orders of G_I^*, G_{II}^*, and G_{IV}^* are solely determined by the factor of n's and $(\hat{h}_1 \ldots \hat{h}_p)$'s through $K_{ij,\hat{\gamma}}$. Therefore, tedious but straightforward calculations show that

https://doi.org/10.1017/S0266466606060294 Published online by Cambridge University Press

$$G_{II}^* \sim n^{-8} \sum_i \sum_{j>i} \sum_{s>j>i} [K_{ij,\hat{\gamma}}^2 K_{is,\hat{\gamma}}^2 + K_{js,\hat{\gamma}}^2 K_{ji,\hat{\gamma}}^2 + K_{si,\hat{\gamma}}^2 K_{sj,\hat{\gamma}}^2] = O_p(n^{-5}(n\hat{h}_1 \dots \hat{h}_q)^{-2}),$$

$$G_{IV}^* \sim n^{-8} \sum_i \sum_{j>i} \sum_s \sum_{t>s} [K_{si,\hat{\gamma}} K_{sj,\hat{\gamma}} K_{ti,\hat{\gamma}} K_{tj,\hat{\gamma}}] = O_p(n^{-4}(\hat{h}_1 \dots \hat{h}_q)^{-1}).$$

Therefore, $G_k^*/S_n^{*4} = o_p(1)$ for all $k = I, II, IV$, and we know that

$$U_n^*/S_n^* \to N(0,1) \quad \text{in distribution in probability.} \tag{A.16}$$

Next, define

$$V_{n,\hat{\gamma}}^* \stackrel{def}{=} E^*\{[n(n-1)(\hat{h}_1 \dots \hat{h}_q)]^{1/2} T_{n1,1}^*\}^2 = \frac{2(\hat{h}_1 \dots \hat{h}_q)}{n(n-1)} \sum_i \sum_{j \neq i} E^*[H_n^{*2}(z_i^*, z_j^*)]$$

and

$$\hat{V}_{n,\hat{\gamma}}^* \stackrel{def}{=} \frac{2(\hat{h}_1 \dots \hat{h}_q)}{n(n-1)} \sum_i \sum_{j \neq i} \hat{H}_n^{*2}(z_i^*, z_j^*),$$

where $\hat{H}_n^*(z_i^*, z_j^*)$ is defined in the same way as $H_n^*(z_i^*, z_j^*)$ except that $\hat{\theta}$ is replaced by $\hat{\theta}^*$. Similar to the analysis of S_n^{*2}, one can show that $\hat{V}_{n,\hat{\gamma}}^* - V_{n,\hat{\gamma}}^* = o_p(1)$ and that $V_{n,\hat{\gamma}}^* - (n^2\hat{h}_1 \dots \hat{h}_q) S_n^{*2} = o_p(1)$. These results together with (A.16) lead to

$$n(\hat{h}_1 \dots \hat{h}_q)^{1/2} T_{n1,1}^* / \sqrt{\hat{V}_{n,\hat{\gamma}}^*} \to N(0,1) \quad \text{in distribution in probability.}$$

The analysis of $T_{n1,2}^*$, $T_{n1,3}^*$, T_{n2}^*, and T_{n3}^* is similar to that of their counterparts in the proof of Theorem 2.1. One can show that $T_{n1,1}^*$ is the leading term of T_n^*. For example, in Lemma A.2(i) we have shown that $T_{n1,2} = O_p(n^{-1/2})$ by proving that $E[T_{n1,2}^2] = O(n^{-1})$. By similar arguments one can show that $E^*[T_{n1,2}^{*2}] = O_p(n^{-1})$. The details are omitted here to save space. Therefore, we conclude that $n(\hat{h}_1 \dots \hat{h}_q)^{1/2} T_n^* / \sqrt{\hat{V}_{n,\hat{\gamma}}^*}$ has the same asymptotic distribution as that of $n(\hat{h}_1 \dots \hat{h}_q)^{1/2} T_{n1,1}^* / \sqrt{\hat{V}_{n,\hat{\gamma}}^*}$. Hence,

$$n(\hat{h}_1 \dots \hat{h}_q)^{1/2} T_n^* / \sqrt{\hat{V}_{n,\hat{\gamma}}^*} \to N(0,1) \quad \text{in distribution in probability.}$$

Because $N(0,1)$ is a continuous distribution, by Polyā's theorem (Bhattacharya and Rao, 1986), we obtain Theorem 2.3. ■