Econometric Theory, **19**, 2003, 1040–1064. Printed in the United States of America.
DOI: 10.1017/S0266466603196077

SEMIPARAMETRIC ESTIMATION OF A HETEROSKEDASTIC SAMPLE SELECTION MODEL

SONGNIAN CHEN
Hong Kong University of Science and Technology

SHAKEEB KHAN
University of Rochester

This paper considers estimation of a sample selection model subject to conditional heteroskedasticity in both the selection and outcome equations. The form of heteroskedasticity allowed for in each equation is multiplicative, and each of the two scale functions is left unspecified. A three-step estimator for the parameters of interest in the outcome equation is proposed. The first two stages involve nonparametric estimation of the "propensity score" and the conditional interquartile range of the outcome equation, respectively. The third stage reweights the data so that the conditional expectation of the reweighted dependent variable is of a partially linear form, and the parameters of interest are estimated by an approach analogous to that adopted in Ahn and Powell (1993, *Journal of Econometrics* 58, 3–29). Under standard regularity conditions the proposed estimator is shown to be $\sqrt{n}$-consistent and asymptotically normal, and the form of its limiting covariance matrix is derived.

1. INTRODUCTION AND MOTIVATION

Estimation of economic models is often confronted with the problem of sample selectivity, which is well known to lead to specification bias if not properly accounted for. Sample selectivity arises from nonrandomly drawn samples, which can be due to either self-selection by the economic agents under investigation or to the selection rules established by the econometrician. In labor economics, the most studied example of sample selectivity is the estimation of the labor supply curve, where hours worked are only observed for agents who decide to participate in the labor force. Examples include the seminal works of Gronau (1974) and Heckman (1976, 1979). It is well known that the failure to account for the presence of sample selection in the data may lead to inconsistent esti-

We are grateful to B. Honoré, R. Klein, E. Kyriazidou, L.-F. Lee, J. Powell, two anonymous referees, and the co-editor D. Andrews and also to seminar participants at Princeton, Queens, UCLA, and the University of Toronto for helpful comments. Chen's research was supported by RGC grant HKUST 6070/01H from the Research Grants Council of Hong Kong. Address correspondence to: Shakeeb Khan, Department of Economics, University of Rochester, Rochester, NY 14627, USA; e-mail: skhan@troi.cc.rochester.edu.

  0266-4666/03 $12.00

mation of the parameters aimed at capturing the behavioral relation between the variables of interest.

Econometricians typically account for the presence of sample selectivity by estimating a bivariate equation model known as the sample selection model (or, using the terminology of Amemiya, 1985, the type 2 Tobit model). The first equation, typically referred to as the "selection" equation, relates the binary selection rule to a set of regressors. The second equation, referred to as the "outcome" equation, relates a continuous dependent variable, which is only observed when the selection variable is 1, to a set of possibly different regressors.

Parametric approaches to estimating this model require the specification of the joint distribution of the bivariate disturbance term. The resulting model is then estimated by maximum likelihood or parametric "two-step" methods. This approach yields inconsistent estimators if the distribution of the disturbance vector is parametrically misspecified and/or conditional heteroskedasticity is present. This negative result has motivated estimation procedures that are robust to either distributional misspecification or the presence of conditional heteroskedasticity. Powell (1989), Choi (1990), and Ahn and Powell (1993) propose two-step estimators that impose no distributional assumptions on the disturbance vector, but none of these are robust to the presence of conditional heteroskedasticity in the outcome equation. Alternatively, Donald (1995) proposes a two-step estimator that allows for general forms of conditional heteroskedasticity but requires the disturbance vector to have a bivariate normal distribution. Chen (1999) recently relaxed the normality assumption but still requires joint symmetry of the error distribution.[1]

Given the observed characteristics of some types of microeconomic data, such as differing variability across agents with differing characteristics, and also empirical distributions exhibiting asymmetry and/or tails thicker than would be consistent with a Gaussian distribution,[2] it appears important to address the issues of both heteroskedasticity and nonnormality/asymmetry simultaneously. This paper attempts to do so by considering a model that exhibits nonparametric multiplicative heteroskedasticity in each of the two equations. This allows for conditional heteroskedasticity of general forms and does not require a parametric specification or a symmetric shape restriction for the distribution of the disturbance vector.[3]

In this paper we show that $\sqrt{n}$-consistent estimation of the parameters in the outcome equation is still possible with the presence of nonparametric multiplicative heteroskedasticity. Our estimation approach involves three stages. The first stage concentrates on the selection equation, estimating the "propensity scores" introduced in Rosenbaum and Rubin (1983). In the second stage, nonparametric quantile regression methods are used to estimate the conditional interquartile range of the outcome equation dependent variable for the selected observations. It will be shown that the conditional interquartile range is a product of the outcome equation scale function and an unknown function of the propensity score. This fact implies that when the dependent variable and regressors are reweighted by dividing by the conditional interquartile range, the conditional expectation of the reweighted outcome equation is of the partially linear form that arises in homoskedastic sample selection models. Since the nonparametric component of the partially linear model is a function of the propensity score, the first-stage estimated values can be used in combination with the reweighted values to estimate the parameters of interest in a fashion analogous to the approaches used in Ahn and Powell (1993), Donald (1995), and Kyriazidou (1997).

The rest of the paper is organized as follows. The next section describes the model in detail and further details the estimation procedure. Sections 3 and 4 detail the regularity conditions imposed and establish the asymptotic properties of the estimator, respectively. Section 5 concludes and suggests topics for future research. The Appendix collects the proofs of the asymptotic arguments.

2. HETEROSKEDASTIC MODEL AND ESTIMATION PROCEDURE

We consider estimation of the following model:

$$d_i = I[\mu(w_i) - \sigma_1(w_i)\nu_i \geq 0], \tag{2.1}$$

$$y_i = d_i y_i^* = d_i \cdot (x_i' \beta_0 + \sigma_2(x_i)\epsilon_i), \tag{2.2}$$

where $\beta_0 \in \Re^{d+1}$ are the unknown parameters of interest, x_i, w_i are observed vectors of explanatory variables (with possibly common elements), $\mu(w_i)$, $\sigma_1(w_i), \sigma_2(x_i)$ are unknown functions of the explanatory variables, and ν_i, ϵ_i are unobserved disturbances, which are independent of the regressors but not necessarily independent of each other. The observed dependent variable d_i in the selection equation is binary, with $I[\cdot]$ denoting the usual indicator function, and the dependent variable of the outcome equation, y_i^*, is only observed when the $d_i = 1$.

Of interest is to estimate β_0 from n observations of a random sample of the quadruple $(d_i, y_i, w_i', x_i')'$, where $'$ denotes the transpose of a vector. We show in this paper that $\sqrt{n}$-consistent estimation of β_0 is possible, and we propose a three-step estimator that achieves this rate of convergence. The following three sections discuss each of the stages in detail.

2.1. First Stage: Kernel Estimation of the Propensity Score

The first stage estimates the probability of selection conditional on the selection equation regressors. Following frequently used terminology, we refer to these conditional probabilities as "propensity scores." We denote the propensity score as

$$p_i \equiv p(w_i) = E[d_i | w_i] = P(d_i = 1 | w_i) = F_\nu\left(\frac{\mu(w_i)}{\sigma_1(w_i)}\right), \tag{2.3}$$

where $F_\nu(\cdot)$ denotes the cumulative distribution function (c.d.f.) of the random variable ν_i and is assumed to be strictly monotonic. Noting that the propensity score is merely a conditional expectation, it can be estimated using nonparametric methods for estimating the conditional mean. Denoting the estimated value by $\hat{p}_i$, we consider the Nadaraya–Watson kernel estimator:

$$\hat{p}_i = \frac{\sum_{j \neq i} d_j K\left(\frac{w_j - w_i}{h_{1n}}\right)}{\sum_{j \neq i} K\left(\frac{w_j - w_i}{h_{1n}}\right)}, \tag{2.4}$$

where $K(\cdot)$ is a kernel function and h_{1n} is a "bandwidth" converging to 0 as the sample size increases to infinity. Additional conditions imposed on $K(\cdot)$ and h_{1n} are discussed in the section detailing regularity conditions.

2.2. Second Stage: Nonparametric Estimation of the Conditional Interquartile Range

In the second stage, we turn attention to the outcome equation. Specifically, we estimate two conditional quantile functions. Let z_i denote the vector of distinct elements of the vector $(w_i', x_i')'$. For a fixed number $\alpha \in (0,1)$, we denote the α conditional quantile function as

$$q_\alpha(z_i) \equiv Q_\alpha(y_i | d_i = 1, x_i), \tag{2.5}$$

where $Q_\alpha(\cdot)$ denotes the α quantile of the distribution of y_i conditional on $d_i = 1$ and x_i. This quantile function can be estimated using nonparametric methods on the subsample of observations for which $d_i = 1$. Although there exist various nonparametric estimation procedures in the literature, we adopt the local polynomial estimator introduced in Chaudhuri (1991a, 1991b), which is also used as a preliminary estimator in Chaudhuri, Doksum, and Samarov (1997), Chen and Khan (2000, 2001), Khan (2001), and Khan and Powell (2001). We first introduce some new notation that will help facilitate a description of the local polynomial procedure.

First, we assume that the regressor vector z_i, whose distribution function we denote by $F_Z(\cdot)$, can be partitioned as $(z_i^{(ds)}, z_i^{(c)})$, where the k_{ds}-dimensional vector $z_i^{(ds)}$ is discretely distributed and the k_c-dimensional vector $z_i^{(c)}$ is continuously distributed.

We let $C_n(z_i)$ denote the cell of observation z_i and let h_{2n} denote the sequence of bandwidths that govern the size of the cell. For some observation z_j, $j \neq i$, we let $z_j \in C_n(z_i)$ denote that $z_j^{(ds)} = z_i^{(ds)}$ and $z_j^{(c)}$ lies in the k_c-dimensional cube centered at $z_i^{(c)}$ with side length h_{2n}.[4]

Next, we let ℓ denote the assumed order of differentiability of the quantile functions with respect to $z_i^{(c)}$, and we let A denote the set of all k_c-dimensional vectors of nonnegative integers b_l, where the sum of the components of each b_l, which we denote by $[b_l]$, is less than or equal to ℓ. We order the elements of the set A such that its first element corresponds to $[b_l] = 0$, and we let $s(A)$ denote the number of elements in A.

For any $s(A)$-dimensional vector Ξ, we let $\Xi_{(l)}$ denote its lth component, and for any two $s(A)$-dimensional vectors a, b, we let a^b denote the product of each component of a raised to the corresponding component of b. Finally, we let $I[\cdot]$ be an indicator function, taking the value 1 if its argument is true and 0 otherwise. The local polynomial estimator of the conditional αth quantile function at an observation z_i involves α-quantile regression (see Koenker and Bassett, 1978) on observations that lie in the defined cells of z_i. Specifically, let the vector

$$(\hat{\Xi}_{(1)}, \hat{\Xi}_{(2)}, \ldots \hat{\Xi}_{(s(A))})$$

minimize the objective function

$$\sum_{j=1}^{n} d_j I[z_j \in C_n(z_i)] \rho_\alpha \left(y_j - \sum_{l=1}^{s(A)} \Xi_{(l)} (z_j^{(c)} - z_i^{(c)})^{b_l} \right), \tag{2.6}$$

where we recall that $\rho_\alpha(\cdot) \equiv \alpha|\cdot| + (2\alpha - 1)(\cdot) I[\cdot < 0]$. The conditional quantile estimator that will be used in the first stage will be the value $\hat{\Xi}_{(1)}$.

A computational advantage of using this estimator is that its evaluation can be carried out by linear programming techniques, whereby a solution can be reached in a finite number of simplex iterations. Since the objective function is globally convex, the solution found is guaranteed to be a global minimizer.

The second stage of the estimation procedure involves this local polynomial estimation procedure of the conditional quantile function, at two particular quantiles, $0 < \alpha_1 < \alpha_2 < 1$.[5] From these two estimators, we can estimate the conditional interquartile range, defined as

$$\Delta q(z) \equiv q_{\alpha_2}(z) - q_{\alpha_1}(z). \tag{2.7}$$

We let $\Delta\hat{q}(z) \equiv \hat{q}_{\alpha_2}(z) - \hat{q}_{\alpha_1}(z)$ denote the second-stage estimator.

2.3. Third Stage: Reweighting and Estimation of a Partially Linear Model

The third-stage estimator is based on the relationship between the propensity score estimated in the first stage and the conditional interquartile range estimated in the second stage. This relationship arises from the result that the conditional distribution function of the outcome equation, given the selection variable is 1 and the regressors, can be expressed as a function of the propensity score.[6] That is,

$$P(\epsilon_i \leq c | d_i = 1, z_i = z) = P(\epsilon_i \leq c | \nu_i \leq \mu(w_i)/\sigma_1(w_i), z_i = z) \tag{2.8}$$

$$= \mathcal{G}(c, \mu(w)/\sigma_1(w)) \tag{2.9}$$

$$= \mathcal{H}(c, p), \tag{2.10}$$

where $\mathcal{G}$ is the unknown conditional distribution function of ϵ_i given the event involving ν_i, $p = P(\nu_i \leq \mu(w)/\sigma_1(w)) \equiv F_{\nu_i}(\mu(w)/\sigma_1(w))$, and the last equality is based on $\mu(w)/\sigma_1(w) = F_{\nu_i}^{-1}(p)$.[7] It immediately follows that any functional of this conditional distribution function can be expressed as a function of the propensity score.

The third-stage estimator is based on three such functionals. First, we note that

$$q_{\alpha_2}(z_i) = x_i'\beta_0 + \sigma_2(x_i)Q_{\alpha_2}(\epsilon_i | d_i = 1, z_i) \tag{2.11}$$

$$= x_i'\beta_0 + \sigma_2(x_i)\lambda_{\alpha_2}(p_i), \tag{2.12}$$

where the second equality results from the established result that the conditional quantile of ϵ_i depends on w_i only through the propensity score p_i, with $\lambda_{\alpha_2}(\cdot)$ denoting the unknown selection correction function. Similarly we have

$$q_{\alpha_1}(z_i) = x_i'\beta_0 + \sigma_2(x_i)\lambda_{\alpha_1}(p_i). \tag{2.13}$$

Letting $\Delta\lambda(p_i) = \lambda_{\alpha_2}(p_i) - \lambda_{\alpha_1}(p_i)$, we now have

$$\Delta q(z_i) = \sigma_2(x_i)\Delta\lambda(p_i). \tag{2.14}$$

This result will prove very useful when combined with the more conventional selection correction equation

$$E[y_i | d_i = 1, z_i] = x_i'\beta_0 + \sigma_2(x_i)\lambda(p_i), \tag{2.15}$$

where $\lambda(p_i) = E[\epsilon_i | d_i = 1, z_i]$. We notice that (2.15) has a similar form to the conditional expectation in Donald (1995). Following a similar approach, we now combine (2.14) and (2.15). Define the transformed variables

$$\tilde{y}_i = \frac{y_i}{\Delta q(z_i)}, \tag{2.16}$$

$$\tilde{x}_i = \frac{x_i}{\Delta q(z_i)} \tag{2.17}$$

and define

$$\tilde{\lambda}(p_i) = \frac{\lambda(p_i)}{\Delta\lambda(p_i)}. \tag{2.18}$$

Then we have the following relationship:

$$\bar{\bar{y}}_i \equiv E[\tilde{y}_i | d_i = 1, z_i] = \tilde{x}_i' \beta_0 + \tilde{\lambda}(p_i). \tag{2.19}$$

This now looks like the partially linear form of the conditional expectation function in the homoskedastic sample selection model. Assuming the values $\tilde{y}_i, \tilde{x}_i$ were observed for the selected observations, we could estimate β_0 by the same procedure used in Ahn and Powell (1993).[8] The idea behind their procedure is that for a pair of observations, indexed by i, j, if $p_i = p_j$, then by differencing the conditional expectations, we have

$$\Delta \bar{\bar{y}}_{ij} = \Delta \tilde{x}_{ij}' \beta_0, \tag{2.20}$$

where $\Delta \tilde{x}_{ij} \equiv \tilde{x}_j - \tilde{x}_i$ and $\Delta \bar{\bar{y}}_{ij} \equiv \bar{\bar{y}}_j - \bar{\bar{y}}_i$. As pointed out in Powell (1989) and Ahn and Powell (1993), if p_i is continuously distributed, such ties where $p_i = p_j$ will occur with probability 0, making it impossible to translate the condition in (2.20) into a feasible estimator of β_0. Nonetheless, if the function $\tilde{\lambda}(\cdot)$ is smooth, then first differencing can be applied to pairs where p_i is close to p_j, resulting in (2.20) holding approximately.

This suggests an estimator based on weighting across pairs, assigning relatively large weight to pairs of observations for which the propensity scores are close and relatively low weight to pairs of observations for which the propensity scores are far apart. If $\tilde{y}_i$ and $\tilde{x}_i$ were observed, one could follow Ahn and Powell (1993) and assign kernel weights to pairs of observations based on the distance between their estimated propensity scores. This would lead to an infeasible weighted least squares estimator of the form

$$\hat{\beta}_{IF} = \left(\sum_{i \neq j} d_i d_j k\left(\frac{\hat{p}_i - \hat{p}_j}{h_{3n}} \right) \Delta \tilde{x}_{ij} \Delta \tilde{x}_{ij}' \right)^{-1} \sum_{i \neq j} d_i d_j k\left(\frac{\hat{p}_i - \hat{p}_j}{h_{3n}} \right) \Delta \tilde{x}_{ij} \Delta \tilde{y}_{ij}, \tag{2.21}$$

where $k(\cdot)$ is a kernel function assigning weights to pairs and h_{3n} is a bandwidth converging to 0 as the sample size increases, ensuring that in the limit, only pairs with identical propensity score values are assigned positive weight.

This estimator is infeasible because the values $\tilde{y}_i, \tilde{x}_i$ are not observed. However, a feasible estimator is immediately suggested, since the second stage of the estimation procedure can be used to construct estimators of $\tilde{y}_i, \tilde{x}_i$ as follows:

$$\hat{\tilde{y}}_i = \frac{y_i}{\Delta \hat{q}(z_i)},$$

$$\hat{\tilde{x}}_i = \frac{x_i}{\Delta \hat{q}(z_i)}.$$

So we can define the third-stage estimator as

$$\hat{\beta} = \left(\sum_{i \neq j} \tau_i \tau_j d_i d_j k\left(\frac{\hat{p}_i - \hat{p}_j}{h_{3n}} \right) \Delta \hat{\tilde{x}}_{ij} \Delta \hat{\tilde{x}}_{ij}' \right)^{-1} \sum_{i \neq j} \tau_i \tau_j d_i d_j k\left(\frac{\hat{p}_i - \hat{p}_j}{h_{3n}} \right) \Delta \hat{\tilde{x}}_{ij} \Delta \hat{\tilde{y}}_{ij}, \tag{2.22}$$

where the trimming functions $\tau_i \equiv \tau(z_i)$ are incorporated in the estimator. These functions trim away observations for which z_i lies outside $\mathcal{Z}$, a compact subset of $\Re^k$, to ensure that only precise estimators of the conditional interquartile range are used.

We conclude this section by commenting on some important issues regarding implementing our procedure in practice.

Remark 2.1. Our procedure requires the choice of multiple smoothing parameters by the econometrician. Although rate conditions on these parameters are given in the following section, they give no indication on how to select values for a given sample. However, the semiparametric literature provides many guidelines for smoothing parameter selection when estimating nonparametric components of the model. For example, h_{2n} may be selected by adopting data driven approaches that accommodate prespecified rate conditions, such as proposed in Chen and Khan (2000) and Khan (2001). Also, the parameters h_{1n} and h_{3n} may be selected by the cross validation procedures mentioned in Ahn and Powell (1993) and Chen and Khan (2003).

Other parameters that are at the discretion of the econometrician are the quantile pair α_1, α_2 and the trimming function. For the quantile pair, it was mentioned that one natural choice would be $\alpha_1 = 0.25$, $\alpha_2 = 0.75$. However, as detailed in the next section, the asymptotic variance of our estimator depends on the choice of quantile pair. The expression for this asymptotic variance can provide some insight on pair selection for maximizing asymptotic efficiency, which would involve a trade-off between having the quantiles as spread out as possible and having them away from the extremes, where they are estimated imprecisely. Expanding on this point, efficiency can be further improved by combining various estimators of β_0 that are based on various quantile pairs, perhaps by extending the generalized method of moments (GMM) framework suggested in Buchinsky (1998). Finally, we feel the choice of trimming function is not as important as the other smoothing parameters. Trimming functions are included primarily for technical reasons, to help alleviate the common problem of imprecise nonparametric estimation in low density areas. In practice they can simply be indicator functions that throw away observations where the regressors take extreme values.

Remark 2.2. Focusing on equation (2.20), we consider the situation where β_0 includes an intercept term, in which case the first component of x_i, denoted by $x_i^{(0)}$, equals 1. Therefore, we have

$$\Delta\tilde{x}_{ij}^{(0)} = \frac{1}{\sigma_2(x_i)\Delta\lambda(p_i)} - \frac{1}{\sigma_2(x_j)\Delta\lambda(p_j)}.$$

We note that under homoskedasticity the preceding difference is 0 whenever $p_i = p_j$, implying that our kernel-weighted regressor matrix will be singular in the limit. However, homoskedasticity is a testable restriction, and if it cannot be rejected Ahn and Powell (1993) can be used to estimate the slope coefficients. In this situation, the intercept term, which is of interest in the treatment effects literature, can be estimated separately by exploiting its "identification at infinity" (Heckman, 1990) as is done in Andrews and Schafgans (1998). On the other hand, from standard results involving the estimability of a subset of regression parameters (see, e.g., Amemiya, 1985), our estimator of the slope coefficients will still be consistent under homoskedasticity, and if heteroskedasticity is indeed present, our procedure will simultaneously estimate the intercept term and slope coefficients at the parametric (root-n) rate.

Remark 2.3. It is worth noting the similarities and differences our approach has with the estimation procedure introduced in Carroll (1982) and Robinson (1987) for the heteroskedastic linear model and in Andrews (1994) for the heteroskedastic partially linear model.[9] In their approach, both the dependent and independent variables were weighted by a nonparametric estimator of a conditional variance function that was based on residuals of the ordinary least squares (OLS) estimator. Unfortunately, this generalized least squares (GLS) type procedure cannot be applied here; this is because one cannot use existing estimators such as those of Ahn and Powell (1993) in a preliminary stage to get "residuals," because these estimators are generally inconsistent as a result of the heteroskedasticity in the outcome equation.

The following sections discuss the asymptotic properties of this three-stage procedure.

3. REGULARITY CONDITIONS

The conditions we need for developing the limiting distribution of the estimator are similar to but more detailed than those required in Ahn and Powell (1993).[10] Because of the extra step of nonparametrically estimating the conditional interquartile range, additional assumptions on both the bandwidth sequence used in the local polynomial estimator and the conditional distribution of the residual associated with the two conditional quantile functions are required. We first state the identification condition on which our estimation procedure is based.

Assumption I (Identification). The distribution of the propensity score p_i has a density with respect to Lebesgue measure, with density function denoted by $f_p(\cdot)$. Defining the following functions of p_i:

$$
\begin{aligned}
f_i &= f_p(p_i),\\
\mu_{\tau i} &= E[\tau_i | p_i],\\
\mu_{\tau x i} &= E[\tau_i \tilde{x}_i | p_i],\\
\mu_{\tau x x i} &= E[\tau_i \tilde{x}_i \tilde{x}_i' | p_i],
\end{aligned}
$$

we require that the matrix

$$\Sigma_{xx} = 2E[p_i^2 f_i(\mu_{\tau i}\mu_{\tau xxi} - \mu_{\tau xi}\mu'_{\tau xi})]$$

have full rank.

We next impose the following conditions on the third-stage kernel function used to match propensity score values and on its bandwidth sequence.

Assumption K3 (Third-Stage Kernel Function). The kernel function $k(\cdot)$ used in the third stage is assumed to have the following properties.

K3.1. $k(\cdot)$ is twice continuously differentiable with bounded second derivative and has compact support.
K3.2. $k(\cdot)$ is symmetric about 0.
K3.3. $k(\cdot)$ is a fourth-order kernel:

$$\int u^l k(u)\,du = 0 \quad \text{for } l = 1,2,3,$$

$$\int u^4 k(u)\,du \neq 0.$$

Assumption H3 (Third-Stage Bandwidth Sequence). The bandwidth sequence h_{3n} used in the third stage is of the form

$$h_{3n} = c_3 n^{-\delta},$$

where c_3 is some constant and $\delta \in (\frac{1}{8}, \frac{1}{6})$.

The following assumption characterizes the order of smoothness of density and conditional expectation functions.

Assumption S1 (Order of Smoothness of Functions of the Propensity Score).

S1.1. The functions $f_p(\cdot)$, $E[\tau_i | p_i = \cdot]$, and $E[\tau_i \tilde{x}_i | p_i = \cdot]$ have order of differentiability of 4, with fourth-order derivatives that are bounded. The function $\tilde{\lambda}(\cdot)$ is fifth-order differentiable, with bounded fifth derivative.

We next impose conditions associated with estimation of the interquartile range. This involves smoothness assumptions on the conditional quantile functions and on the distributions of z_i and the residuals associated with the quantile functions.

Assumption RD1 (Regressor Distribution). The regressor vector can be decomposed as $z_i = (z_i^{(c)'}, z_i^{(ds)'})'$ where the k_c-dimensional vector $z_i^{(c)}$ is continuously distributed and the k_{ds}-dimensional vector $z_i^{(ds)}$ is discretely distributed. Letting $f_{Z^{(c)}|Z^{(ds)}}(\cdot | z^{(ds)})$ denote the conditional density function of $z_i^{(c)}$ given $z_i^{(ds)} = z^{(ds)}$, we assume it is bounded, bounded away from 0, and Lipschitz continuous on $\mathcal{Z}$.

Letting $f_{Z^{(ds)}}(\cdot)$ denote the mass function of $z_i^{(ds)}$, we assume a finite number of mass points on $\mathcal{Z}$. Finally, we let $f_Z(\cdot)$ denote $f_{Z^{(c)}|Z^{(ds)}}(\cdot|\cdot) f_{Z(ds)}(\cdot)$.

Assumption S2 (Order of Smoothness of Conditional Quantile Functions).

S2.1. For all values of $z_i^{(ds)}$, the functions $q_{\alpha_2}(\cdot), q_{\alpha_1}(\cdot)$ are bounded and M_2 times continuously differentiable with bounded M_2^{th} derivatives, with respect to $z_i^{(c)}$ on $\mathcal{Z}$.

S2.2. The polynomial used for the second-stage quantile function estimators is of order M_2.

Assumption CED (Outcome Equation Conditional Error Distribution). The homoskedastic component of the outcome equation error term, conditional on selection and z_i, denoted by $\epsilon_i | d_i = 1, z_i$, has a continuous distribution with density function that is bounded, positive, and continuous, for all $z_i \in \mathcal{Z}$.

Assumption H2 (Second-Stage Bandwidth Sequence for Interquartile Range Estimation). The bandwidth sequence used to estimate the conditional interquartile range is of the form

$$h_{2n} = c_2 n^{-\gamma_2},$$

where c_2 is a constant and $\gamma_2 \in \left((\delta + \frac{1}{2})/M_2, (1 - 4\delta)/3k_c\right)$.

The final set of assumptions involve restrictions for the first-stage kernel estimator of the propensity score. This involves smoothness conditions on both the propensity scores p_i and the distribution of the regressors in the selection equation and the rate at which the first-stage bandwidth sequence decreases to 0.

Assumption RD2 (Distribution of Selection Equation Regressors). The regressor vector in the selection equation can be decomposed as $w_i = (w_i^{(c)'}, w_i^{(ds)'})'$ where the w_c-dimensional vector $w_i^{(c)}$ is continuously distributed and the w_d-dimensional vector $w_i^{(ds)}$ is discretely distributed.

Letting $f_{W^{(c)}|W^{(d)}}(\cdot|w^{(ds)})$ denote the conditional density function of $w_i^{(c)}$ given $w_i^{(ds)} = w^{(ds)}$, we assume it is bounded away from 0 and ∞ on $\mathcal{W}$, a predetermined compact subset of the support of w_i. Following Ahn and Powell (1993), we redefine d_i to take the value 0 if $w_i \notin \mathcal{W}$.

Letting $f_{W^{(ds)}}(\cdot)$ denote the mass function of $w_i^{(ds)}$, we assume a finite number of mass points on $\mathcal{W}$. Finally, we let $f_W(\cdot)$ denote $f_{W^{(c)}|W^{(ds)}}(\cdot|\cdot) f_{W^{(ds)}}(\cdot)$.

Assumption PS (Order of Smoothness of Functions Involving the Selection Regressors). For each $w^{(d)}$ in the support of $w_i^{(d)}$, p_i and $f_W(\cdot)$ are bounded and continuously differentiable of order M_1 functions of $w_i^{(c)}$, where M_1 is an even integer satisfying $M_1 > w_c/(\frac{1}{3} - 2\delta)$.

Assumption K1 (First-Stage Kernel Function Condition). The kernel function $K(\cdot)$ used in the first stage is of bounded variation, has bounded support, integrates to 1, and is of order M_1, where the order of a kernel function was defined in Assumption K3.

Assumption H1 (Rate Condition on First-Stage Bandwidth Sequence). The bandwidth sequence h_{1n} of the first-stage kernel estimator of the propensity score is of the form

$$h_{1n} = c_1 n^{-\gamma},$$

where c_1 is some constant and γ satisfies

$$\gamma \in \left(\frac{1}{2M_1}, \frac{\frac{1}{6} - \delta}{w_c} \right),$$

where δ is regulated by Assumption H3.

Remark 3.1. These regularity conditions are quite standard when compared to other estimators in the semiparametric literature. They are very similar to those required in Ahn and Powell (1993), and we state the important similarities and differences here.

(i) The identification in Assumption I is analogous to Assumption 3.4 in Ahn and Powell (1993). As discussed in that paper, with the assumption of a linear index in the selection equation, an exclusion restriction is required for identification of β_0. Specifically, it is required that a component included in w_i be excluded from x_i. We impose this exclusion restriction for the results in this paper as the special case where $\sigma_2(\cdot) \equiv 1$ corresponds to the homoskedastic model. Although it may be possible to identify the parameter β_0 without an exclusion restriction by the "nonlinearity" induced by the presence of the scale functions, we are not comfortable with requiring heteroskedasticity for identification without exclusion.

(ii) Assumptions RD1, S2, and CED, which impose conditions on the regressors and error term in the outcome equation, are generally not required for the homoskedastic model considered in Ahn and Powell (1993). They are imposed here to ensure uniform rates of convergence for the conditional quantile estimator used in the second stage of our procedure.

4. ASYMPTOTIC PROPERTIES

Here we briefly discuss the asymptotic properties of the proposed three-stage estimation procedure. The first result, illustrated in the following lemma, establishes the asymptotic difference between the proposed estimator and the infeasible estimator that assumes $\tilde{y}_i$ and $\tilde{x}_i$ are observed for observations with $d_i = 1$.[11] The details of the proof are left to the Appendix.

LEMMA 4.1. *Let $\hat{\beta}_{IF}$ be defined as in equation (2.21), with the trimming functions τ_i, τ_j included. Then under the regularity conditions detailed in the previous section,*

$$\hat{\beta} = \hat{\beta}_{IF} + \Sigma_{xx}^{-1} \frac{1}{n} \sum_{i=1}^{n} \psi_{2i} + o_p(n^{-1/2}) \tag{4.1}$$

with

$$\begin{aligned} \psi_{2i} = -2\tau_i d_i (& f_{U_2|Z}(0|z_i)^{-1}(I[y_i \le q_{2i}] - \alpha_2) \\ & - (f_{U_1|Z}(0|z_i)^{-1}(I[y_i \le q_{1i}] - \alpha_1)) \\ & \times \Delta q_i^{-1} f_i \tilde{\lambda}(p_i) p_i^2 (\tilde{x}_i \mu_{\tau i} - \mu_{\tau x i}), \end{aligned} \tag{4.2}$$

where q_{1i}, q_{2i} denote $q_{\alpha_1}(z_i)$ and $q_{\alpha_2}(z_i)$, respectively, and $f_{U_2|Z}, f_{U_1|Z}$ denote the conditional density functions of the residuals associated with the conditional quantile functions.

An immediate implication of this lemma is that the arguments developed in Powell (1989) and Ahn and Powell (1993) can be used to derive the limiting distribution of the feasible estimator $\hat{\beta}$. The limiting distribution is characterized in the following theorem.

THEOREM 4.1. *The feasible estimator is $\sqrt{n}$-consistent and asymptotically normal; specifically, we have*

$$\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, \Sigma_{xx}^{-1} \Omega_{xx} \Sigma_{xx}^{-1}), \tag{4.3}$$

where Ω_{xx} is the covariance matrix of the k-dimensional vector $\psi_i \equiv \psi_{1i} + \psi_{2i}$ with

$$\psi_{1i} = 2\tau_i p_i f_i (\tilde{x}_i \mu_{\tau i} - \mu_{\tau x i}) \cdot (d_i \tilde{u}_i + p_i \tilde{\lambda}'(p_i)(d_i - p_i)) \tag{4.4}$$

with $\tilde{\lambda}'(\cdot)$ denoting the derivative of $\tilde{\lambda}(\cdot)$ and $\tilde{u}_i$ is the "residual"

$$\tilde{u}_i = \tilde{y}_i - \tilde{x}_i' \beta_0 - \tilde{\lambda}(p_i). \tag{4.5}$$

For the purpose of conducting inference we propose a consistent estimator of the asymptotic variance matrix. To estimate the component Σ_{xx} we propose a standard "plug-in" estimator that replaces unknown values and expectations with estimated values and sample averages, respectively:

$$\hat{\Sigma}_{xx} = \frac{1}{n(n-1)h_{3n}} \left(\sum_{i \ne j} \tau_i \tau_j d_i d_j k\left(\frac{\hat{p}_i - \hat{p}_j}{h_{3n}} \right) \Delta \hat{\tilde{x}}_{ij} \Delta \hat{\tilde{x}}_{ij}' \right). \tag{4.6}$$

Estimation of Ω_{xx} is more involved, because it involves the unknown selection correction function and its derivative. We propose the following estimator, which is analogous to those found in Powell (1989) and Ahn and Powell (1993):

$$\hat{\Omega}_{xx} = \frac{1}{n} \sum_{i=1}^{n} (\hat{\psi}_{1i} + \hat{\psi}_{2i})(\hat{\psi}_{1i} + \hat{\psi}_{2i})', \tag{4.7}$$

where $\hat{\psi}_{1i} \equiv \hat{\psi}_{11i} + \hat{\psi}_{12i}$ with

$$\hat{\psi}_{11i} = 2\tau_i d_i \left(\frac{1}{nh_{3n}} \sum_{j=1}^{n} \tau_j d_j k\left(\frac{\hat{p}_i - \hat{p}_j}{h_{3n}} \right) \Delta\hat{\hat{v}}_{ij} \Delta\hat{\hat{x}}_{ij} \right) \tag{4.8}$$

with $\Delta\hat{\hat{v}}_{ij} = \Delta\hat{\hat{y}}_{ij} - \Delta\hat{\hat{x}}_{ij}' \hat{\beta}$ and where

$$\hat{\psi}_{12i} = 2\hat{\hat{\zeta}}_i (d_i - \hat{p}_i) \tag{4.9}$$

with

$$\hat{\hat{\zeta}}_i = \frac{1}{nh_{3n}^2} \sum_{l \neq j} \tau_l \tau_j d_l d_j k'\left(\frac{\hat{p}_l - \hat{p}_j}{h_{3n}} \right) K\left(\frac{w_l - w_i}{h_{1n}} \right) \Delta\hat{\hat{v}}_{lj} \Delta\hat{\hat{x}}_{lj} \times \left[\sum_{l=1}^{n} K\left(\frac{w_l - w_i}{h_{1n}} \right) \right]^{-1}, \tag{4.10}$$

where $k'(\cdot)$ denotes the derivative of the third-stage kernel function.

Finally, to estimate ψ_{2i} we require an estimator of the conditional density of the quantile residuals. Letting $\hat{u}_{2i} \equiv d_i(y_i - \hat{q}_{2i})$ denote the estimated residual of the higher quantile, we propose a kernel estimator of its conditional density function:

$$\hat{f}_{U_2|Z}(0|z_i) = \frac{\frac{1}{\nu_n} \sum_{l=1}^{n} d_l I[z_l \in C_{ni}] \tilde{k}\left(\frac{\hat{u}_{2i}}{\nu_n} \right)}{\sum_{l=1}^{n} d_l I[z_l \in C_{ni}]}, \tag{4.11}$$

where $\tilde{k}(\cdot)$ is a nonnegative kernel function with bounded support that integrates to 1 and is continuously differentiable with bounded derivative. The sequence ν_n satisfies the constraints

$$\nu_n \to 0, \qquad n\nu_n \to \infty, \qquad \frac{(\log n)^{1/2}(nh_{2n}^{k_c})^{-1/2} + h_{2n}^{M_2}}{\nu_n^2} \to 0.$$

A similar estimator can be constructed for the density of the lower quantile residual. With these in hand we can construct the following estimator of $\hat{\psi}_{2i}$:

$$\hat{\psi}_{2i} = -2\tau_i d_i \hat{p}_i (\hat{f}_{U_2|Z}(0|z_i)^{-1}(I[y_i \leq \hat{q}_{2i}] - \alpha_2) - \hat{f}_{U_1|Z}(0|z_i)^{-1}(I[y_i \leq \hat{q}_{1i}] - \alpha_1))\Delta\hat{q}_i^{-1} \times \left(\frac{1}{nh_{3n}} \sum_{j=1}^{n} \tau_j d_j k\left(\frac{\hat{p}_i - \hat{p}_j}{h_{3n}} \right) - \hat{\hat{v}}_j \Delta\hat{\hat{x}}_{ij} \right). \tag{4.12}$$

The following theorem, whose proof is left to the Appendix, establishes the consistency of the estimator $\hat{\Sigma}_{xx}^{-1}\hat{\Omega}_{xx}\hat{\Sigma}_{xx}^{-1}$.

THEOREM 4.2. *Under the conditions used to establish Theorem (4.1),*

$$\hat{\Sigma}_{xx}^{-1}\hat{\Omega}_{xx}\hat{\Sigma}_{xx}^{-1} \xrightarrow{p} \Sigma_{xx}^{-1}\Omega_{xx}\Sigma_{xx}^{-1}. \tag{4.13}$$

5. CONCLUSIONS

This paper introduces a new estimator for a semiparametric sample selection model that is consistent in the presence of multiplicative heteroskedasticity of unknown form. The estimation procedure involves three stages, two of which are analogous to the steps taken in the estimators introduced in Powell (1989), Ahn and Powell (1993), Donald (1995), and Kyriazidou (1997). The new estimator is shown to converge at the parametric rate and has an asymptotic normal distribution.

The work here suggests areas for future research. As was suggested, it would be useful to formally construct a pretest for conditional heteroskedasticity to see if the slope coefficients and the intercept term could be estimated simultaneously. The effect of this pretesting procedure on the limiting distribution theory developed here is also worth exploring. Also, one could consider imposing the multiplicative structure to model heteroskedasticity in other bivariate models, such as the type 3 Tobit model, where the selection equation now involves a censored, instead of binary, variable.

NOTES

1. Kyriazidou (1997) considers estimation of a sample selection model for longitudinal data. Her approach allows for cross-sectional heteroskedasticity and also for individual specific effects. The purpose of our paper is to estimate the parameters of interest when only a cross-sectional data set is available to the econometrician.

2. For example, in his empirical analysis of female labor supply, Mroz (1987) finds the disturbance term in the labor force participation equation to have skewness and kurtotis levels that are inconsistent with the normality assumption.

3. The idea of modeling conditional heteroskedasticity through a multiplicative structure is quite common in the econometrics and statistics literature, in both applied and theoretical work. See Chen and Khan (2000, 2003) and references therein for a list of examples.

4. Here we have required an observation's discrete component to match up exactly for it to fall in a given cell. However, it is often the case that the finite-sample performance of nonparametric procedures can be improved by smoothing across discrete components also (see Li and Racine, 2000).

5. We note that any two quantiles may be used in this stage. One natural choice would be $\alpha_1 = 0.25$, $\alpha_2 = 0.75$ because the interquartile range is typically used as a quantile-based measure of dispersion in practice.

6. This result is based on the assumption of strict monotonicity of the c.d.f. of the random variable ν_i, the disturbance in the selection equation.

7. This result was first used in the work of Ahn and Powell (1993), where the authors assumed that the selection equation disturbance term was additively separable and distributed independently

of the regressors. Although we relax their homoskedasticity assumption, we require additive separability and multiplicative heteroskedasticity to express the distribution function as a function of the propensity score.

8. We note that any of the existing estimators for the partially linear model, such as in Robinson (1988) and Andrews (1991, 1994), could be used in this stage.

9. This type of procedure is also used to estimate a heteroskedastic nonlinear regression model in Delgado (1992) and Hildago (1992).

10. In this section we assume that β_0 does include an intercept term to be estimated. As discussed in Remark 2.2, the intercept cannot be consistently estimated under conditional homoskedasticity, though the slope coefficients can be. Thus we are implicitly assuming that either conditional heteroskedasticity is present or that the parameter of interest is the last k components of β_0.

11. We note that this result is in contrast to the results found in Carroll (1982) and Robinson (1987) for the heteroskedastic linear regression model and in Andrews (1994) for the (partially) heteroskedastic partially linear model, where asymptotic equivalence is established for their GLS-type procedures.

REFERENCES

Ahn, H. & J.L. Powell (1993) Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics* 58, 3–29.

Amemiya, T. (1985) *Advanced Econometrics*. Cambridge: Harvard University Press.

Andrews, D.W.K. (1991) Asymptotic normality for nonparametric and semiparametric regression models. *Econometrica* 59, 307–345.

Andrews, D.W.K. (1994) Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* 62, 43–72.

Andrews, D.W.K. & M.M.A. Schafgans (1998) Semiparametric estimation of the intercept of a sample selection model. *Review of Economic Studies* 65, 497–517.

Bierens, H.J. (1987) Kernel estimators of regression functions. In T.F. Bewley (ed.), *Advances in Econometrics, Fifth World Congress*, vol. 1, pp. 99–144. Cambridge: Cambridge University Press.

Buchinsky, M. (1998) Recent advances in quantile regression models: A practical guideline for empirical research. *Journal of Human Resources* 33, 88–126.

Carroll, R.J. (1982) Adapting for heteroskedasticity in linear models. *Annals of Statistics* 10, 1224–1233.

Chaudhuri, P. (1991a) Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics* 19, 760–777.

Chaudhuri, P. (1991b) Global nonparametric estimation of conditional quantiles and their derivatives. *Journal of Multivariate Analysis* 39, 246–269.

Chaudhuri, P., K. Doksum, & A. Samarov (1997) On average derivative quantile regression. *Annals of Statistics* 25, 715–744.

Chen, S. (1999) Semiparametric estimation of heteroskedastic binary choice sample selection models under symmetry. Manuscript, Hong Kong University of Science and Technology.

Chen, S. & S. Khan (2000) Estimating censored regression models in the presence of nonparametric multiplicative heteroskedasticity. *Journal of Econometrics* 98, 283–316.

Chen, S. & S. Khan (2001) Semiparametric estimation of a partially linear censored regression model. *Econometric Theory* 17, 567–590.

Chen, S. & S. Khan (2003) Rates of convergence for estimating regression coefficients in heteroskedastic discrete response models. *Journal of Econometrics*, in press.

Choi, K. (1990) Semiparametric Estimation of the Sample Selection Model Using Series Expansion and the Propensity Score. Manuscript, University of Chicago.

Delgado, M. (1992) Semiparametric generalized least squares in the multivariate nonlinear regression model. *Econometric Theory* 8, 203–222.

Donald, S.G. (1995) Two-step estimation of heteroscedastic sample selection models. *Journal of Econometrics* 65, 347–380.

Gronau, R. (1974) Wage comparisons: A selectivity bias. *Journal of Political Economy* 82, 1119–1144.

Heckman, J.J. (1976) The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator of such models. *Annals of Economic and Social Measurement* 15, 475–492.

Heckman, J.J. (1979) Sample selection bias as a specification error. *Econometrica* 47, 153–161.

Heckman, J.J. (1990) Varieties of selection bias. *American Economic Review* 80, 313–318.

Hildago, J. (1992) Adaptive estimation in time series models with heteroskedasticity of unknown form. *Econometric Theory* 8, 161–187.

Horowitz, J.L. (1992) A smoothed maximum score estimator for the binary response model. *Econometrica* 60, 505–531.

Khan, S. (2001) Two stage rank estimation of quantile index models. *Journal of Econometrics* 100, 319–355.

Khan, S. & J.L. Powell (2001) Two-step quantile estimation of semiparametric censored regression models. *Journal of Econometrics* 103, 73–110.

Koenker, R. & G.S. Bassett, Jr. (1978) Regression quantiles. *Econometrica* 46, 33–50.

Kyriazidou, E. (1997) Estimation of a panel data sample selection model. *Econometrica* 65, 1335–1364.

Li, Q. & J. Racine (2000) Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data. Manuscript, Texas A&M University.

Mroz, T.A. (1987) The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica* 55, 765–799.

Powell, J.L. (1989) Semiparametric Estimation of Censored Selection Models. Manuscript, University of Wisconsin.

Robinson, P.M. (1987) Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica* 56, 875–891.

Robinson, P.M. (1988) Root-*N*-consistent semiparametric regression. *Econometrica* 56, 931–954.

Rosenbaum, P.R. & D.B. Rubin (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41–55.

Sherman, R.P. (1994) Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Annals of Statistics* 22, 439–459.

APPENDIX

To keep expressions notationally simple, in this section we let q_{1i}, q_{2i} denote $q_{\alpha_1}(z_i)$ and $q_{\alpha_2}(z_i)$, respectively. We denote estimated values by $\hat{q}_{1i}, \hat{q}_{2i}$. For the second-stage nonparametric procedure, we let C_{ni} denote $C_n(z_i)$. Also, for any matrix A with elements a_{ij}, we let $\|A\|$ denote $(\sum_{i,j} a_{ij}^2)^{1/2}$.

We begin by stating two uniform convergence results, one for each of the nonparametric estimators used in the first two stages of the estimation procedure. These uniform rates can be found in Ahn and Powell (1993) and Chaudhuri et al. (1997), respectively.

LEMMA A.1. (From Ahn and Powell, 1993, Lemma A.1). *Under Assumptions RD2, PS, K1, H1, and H3*

$$\sup_{w \in \mathcal{W}} |\hat{p}(w) - p(w)| = O_p(n^{-(1/3+\delta)}). \tag{A.1}$$

LEMMA A.2. (From Chaudhuri et al., 1997, Lemma 4.1). *Under Assumptions RD1, S2, H2, and CED,*

$$\sup_{z\in\mathcal{Z}}|\Delta\hat{q}(z)-\Delta q(z)| = O_p((\log n)^{1/2}(nh_{2n}^{k_c})^{-1/2}+h_{2n}^{M_2}). \tag{A.2}$$

Before proceeding to the arguments used in the proof, we first state rates of convergence for various terms that will arise in the proof. These rates arise directly from the preceding two lemmas and also from Assumptions H1–H3. We adopt the notation that $\|\cdot\|_\infty$ denotes the supremum over the set in question.

$$h_{3n}^{-2}\|\hat{q}_\alpha - q_\alpha\|_\infty\|\hat{p}-p\|_\infty = o_p(n^{-1/2}), \tag{A.3}$$

$$h_{3n}^{-1}(nh_{2n}^{k_c})^{-3/4} = o_p(n^{-1/2}), \tag{A.4}$$

$$h_{3n}^{-1}h_{2n}^{-k_c} = o(n), \tag{A.5}$$

$$h_{3n}^{-1}h_{2n}^{M_2} = o(n^{-1/2}), \tag{A.6}$$

$$h_{3n}^{-1}\|\hat{q}_\alpha - q_\alpha\|_\infty^2 = o_p(n^{-1/2}). \tag{A.7}$$

The proof of Theorem 4.1 is based on an asymptotically linear representation for $\hat{\beta}-\beta_0$. By (2.22), we have

$$\hat{\beta}-\beta_0 = \hat{S}_{xx}^{-1}\hat{S}_{xy}, \tag{A.8}$$

where

$$\hat{S}_{xx} = \frac{1}{n(n-1)}\sum_{i\neq j} h_{3n}^{-1}\tau_i\tau_j d_i d_j k\left(\frac{\hat{p}_i-\hat{p}_j}{h_{3n}}\right)\Delta\hat{\hat{x}}_{ij}\Delta\hat{\hat{x}}_{ij}' \tag{A.9}$$

and

$$\hat{S}_{xy} = \frac{1}{n(n-1)}\sum_{i\neq j} h_{3n}^{-1}\tau_i\tau_j d_i d_j k\left(\frac{\hat{p}_i-\hat{p}_j}{h_{3n}}\right)\Delta\hat{\hat{x}}_{ij}(\Delta\hat{\hat{y}}_{ij}-\Delta\hat{\hat{x}}_{ij}'\beta_0). \tag{A.10}$$

The following lemma establishes the probability limit of A.9.

LEMMA A.3. *Under Assumptions I, K3.1, K3.2, H3, RD1, S2, H2, CED, RD2, PS, K1, and H1,*

$$\hat{S}_{xx} \xrightarrow{p} \Sigma_{xx}. \tag{A.11}$$

Proof. Define

$$S_{xx} = \sum_{i\neq j}\tau_i\tau_j d_i d_j h_{3n}^{-1}k\left(\frac{\hat{p}_i-\hat{p}_j}{h_{3n}}\right)\Delta\tilde{x}_{ij}\Delta\tilde{x}_{ij}'.$$

We first show that

$$\hat{S}_{xx} - S_{xx} = o_p(1). \tag{A.12}$$

Let $r_{ni} = \tilde{x}_i - \hat{\tilde{x}}_i$ and $\Delta r_{nij} = r_{nj} - r_{ni}$. We thus have

$$\Delta\hat{\tilde{x}}_{ij}\Delta\hat{\tilde{x}}'_{ij} = \Delta\tilde{x}_{ij}\Delta\tilde{x}'_{ij} + \Delta\tilde{x}_{ij}\Delta r'_{nij} + \Delta r_{nij}\Delta\tilde{x}'_{ij} + \Delta r_{nij}\Delta r'_{nij}.$$

We note that Δq_i is bounded away from 0 on the support of τ_i. Noting that $\hat{\tilde{x}}_i - \tilde{x}_i = x_i(1/\Delta\hat{q}_i - 1/\Delta q_i)$ we have by the bounds on $k(\cdot)$ and the support of τ_i and Assumption H2 that

$$\max_{1\le i\le n}\|r_{ni}\| = O_p(\|\hat{q}_1 - q_1\|_\infty + \|\hat{q}_2 - q_2\|_\infty).$$

It thus follows by the bound on $k(\cdot)$ that

$$\hat{S}_{xx} = S_{xx} + O_p(h_{3n}^{-1}(\|\hat{q}_1 - q_1\|_\infty + \|\hat{q}_2 - q_2\|_\infty)) = S_{xx} + o_p(1),$$

where the second equality follows by (A.7). This establishes (A.12). By the same arguments of Lemma 3.1 and Theorem 3.1 in Ahn and Powell (1993) it follows that

$$S_{xx} - \Sigma_{xx} = o_p(1),$$

which concludes the proof. ■

We next derive a linear representation for $\hat{S}_{xy}$. The following lemma establishes the difference between $\hat{S}_{xy}$ and the analogous expression involving the true values of the conditional interquartile range.

LEMMA A.4. *Under Assumptions K3.1, K3.2, H3, H2, RD1, S2, and CED,*

$$\hat{S}_{xy} = S_{xy} + \frac{1}{n}\sum_{i=1}^{n}\psi_{2i},$$

where

$$S_{xy} = \frac{1}{n(n-1)}\sum_{i\neq j}\tau_i\tau_j d_i d_j h_{3n}^{-1} k\left(\frac{\hat{p}_i - \hat{p}_j}{h_{3n}}\right)\Delta\tilde{x}_{ij}(\Delta\tilde{y}_{ij} - \Delta\tilde{x}'_{ij}\beta_0)$$

and

$$\psi_{2i} = -2\tau_i d_i(f_{U_2|Z}(0|z_i)^{-1}(I[y_i \le q_{2i}] - \alpha_2) - (f_{U_1|Z}(0|z_i)^{-1}(I[y_i \le q_{1i}] - \alpha_1))$$

$$\times\Delta q_i^{-1} f_i\tilde{\lambda}(p_i)p_i^2(\tilde{x}_i\mu_{\tau i} - \mu_{\tau x i}).$$

Proof. We replace the first $\Delta\hat{\tilde{x}}_{ij}$ in (A.10) with $\Delta\tilde{x}_{ij}$ and derive a linear representation for

$$\frac{1}{n(n-1)}\sum_{i\neq j}\tau_i\tau_j d_i d_j h_{3n}^{-1}k\left(\frac{\hat{p}_i-\hat{p}_j}{h_{3n}}\right)\Delta\tilde{x}_{ij}[(\Delta\hat{\tilde{y}}_{ij}-\Delta\hat{\tilde{x}}_{ij}'\beta_0)-(\Delta\tilde{y}_{ij}-\Delta\tilde{x}_{ij}'\beta_0)]. \quad \textbf{(A.13)}$$

We note the preceding expression can be expressed as

$$\frac{1}{n(n-1)}\sum_{i\neq j}\tau_i\tau_j d_i d_j h_{3n}^{-1}k\left(\frac{\hat{p}_i-\hat{p}_j}{h_{3n}}\right)\Delta\tilde{x}_{ij}\left(\frac{1}{\Delta\hat{q}_j}-\frac{1}{\Delta q_j}\right)(y_j-x_j'\beta_0) \quad \textbf{(A.14)}$$

$$-\frac{1}{n(n-1)}\sum_{i\neq j}\tau_i\tau_j d_i d_j h_{3n}^{-1}k\left(\frac{\hat{p}_i-\hat{p}_j}{h_{3n}}\right)\Delta\tilde{x}_{ij}\left(\frac{1}{\Delta\hat{q}_i}-\frac{1}{\Delta q_i}\right)(y_i-x_i'\beta_0). \quad \textbf{(A.15)}$$

We next linearize $1/\Delta\hat{q}_i-1/\Delta q_i$ as

$$\frac{1}{\Delta\hat{q}_i}-\frac{1}{\Delta q_i}=\frac{-1}{(\Delta q_i)^2}(\Delta\hat{q}_i-\Delta q_i)+\frac{1}{\Delta\hat{q}_i\Delta q_i}(\Delta\hat{q}_i-\Delta q_i)^2.$$

It follows that the quadratic terms in this expansion are $o_p(n^{-1/2})$ by (A.7). Plugging the linear term of the expansion into (A.14) and (A.15) yields the expression

$$\frac{1}{n(n-1)}\sum_{i\neq j}\tau_i\tau_j d_i d_j h_{3n}^{-1}k\left(\frac{\hat{p}_i-\hat{p}_j}{h_{3n}}\right)\Delta\tilde{x}_{ij}\Delta q_i^{-1}(\Delta\hat{q}_i-\Delta q_i)(\tilde{y}_i-\tilde{x}_i'\beta_0) \quad \textbf{(A.16)}$$

$$-\frac{1}{n(n-1)}\sum_{i\neq j}\tau_i\tau_j d_i d_j h_{3n}^{-1}k\left(\frac{\hat{p}_i-\hat{p}_j}{h_{3n}}\right)\Delta\tilde{x}_{ij}\Delta q_j^{-1}(\Delta\hat{q}_j-\Delta q_j)(\tilde{y}_j-\tilde{x}_j'\beta_0). \quad \textbf{(A.17)}$$

Writing Δq_i as $q_{2i}-q_{1i}$, we have that (A.16) and (A.17) can be expressed as

$$\frac{1}{n(n-1)}\sum_{i\neq j}\tau_i\tau_j d_i d_j h_{3n}^{-1}k\left(\frac{\hat{p}_i-\hat{p}_j}{h_{3n}}\right)\Delta\tilde{x}_{ij}\Delta q_i^{-1}(\hat{q}_{2i}-q_{2i})(\tilde{y}_i-\tilde{x}_i'\beta_0) \quad \textbf{(A.18)}$$

$$-\frac{1}{n(n-1)}\sum_{i\neq j}\tau_i\tau_j d_i d_j h_{3n}^{-1}k\left(\frac{\hat{p}_i-\hat{p}_j}{h_{3n}}\right)\Delta\tilde{x}_{ij}\Delta q_j^{-1}(\hat{q}_{2j}-q_{2j})(\tilde{y}_j-\tilde{x}_j'\beta_0) \quad \textbf{(A.19)}$$

plus analogous terms involving $\hat{q}_{1i}$, q_{1i}, $\hat{q}_{1j}$, and q_{1j}. We focus our attention on (A.18).

By a mean value expansion of $h_{3n}^{-1}k\big((\hat{p}_i-\hat{p}_j)/h_{3n}\big)$ around $h_{3n}^{-1}k\big((p_i-p_j)/h_{3n}\big)$ and the bound on the second moments of $\tau_i\tau_j d_i d_j\Delta\tilde{x}_{ij}(\tilde{y}_i-\tilde{x}_i'\beta_0)$ it follows that $h_{3n}^{-1}k\big((\hat{p}_i-\hat{p}_j)/h_{3n}\big)$ can be replaced with $h_{3n}^{-1}k\big((p_i-p_j)/h_{3n}\big)$ in (A.18) with a resulting remainder term that is

$$O_p\Big(h_{3n}^{-2}\max_{1\le i\le n}|\hat{p}_i-p_i||\hat{q}_{2i}-q_{2i}|\Big)=o_p(n^{-1/2}),$$

where the equality follows from (A.3). Noting that we can replace $d_i(\tilde{y}_i-\tilde{x}_i'\beta_0)$ with $d_i(\tilde{\lambda}(p_i)+\tilde{u}_i)$ in (A.18), where we recall that the $\tilde{u}_i$ is a mean 0 residual term, we will first show that

$$\frac{1}{n(n-1)}\sum_{i\neq j}\tau_i\tau_j d_i d_j h_{3n}^{-1}k\left(\frac{p_i-p_j}{h_{3n}}\right)\Delta\tilde{x}_{ij}\Delta q_i^{-1}(\hat{q}_{2i}-q_{2i})\tilde{u}_i = o_p(n^{-1/2}), \tag{A.20}$$

and we will then derive a linear representation for

$$\frac{1}{n(n-1)}\sum_{i\neq j}\tau_i\tau_j d_i d_j h_{3n}^{-1}k\left(\frac{p_i-p_j}{h_{3n}}\right)\Delta\tilde{x}_{ij}\Delta q_i^{-1}(\hat{q}_{2i}-q_{2i})\tilde{\lambda}(p_i). \tag{A.21}$$

To show (A.20), we use the local Bahadur representations for the local polynomial estimator established in Chaudhuri (1991a) and Chaudhuri et al. (1997). In our context this can be expressed as

$$\hat{q}_{2i}-q_{2i} = \frac{1}{nh_{2n}^{k_c}}f_{U_2,Z}(0,z_i)^{(-1)}\sum_{k\neq i}d_k I[z_k\in C_{ni}](I[y_k\le q^*_{2(k,i)}]-\alpha_2)+R_{ni}, \tag{A.22}$$

where $q^*_{2(k,i)}$ denotes the Taylor polynomial approximation of q_{2k} for z_k close to z_i, $f_{U_2,Z}(\cdot,\cdot)$ denotes the joint density function of the quantile residual and z_k; R_{ni} is a remainder term converging to 0 at a rate depending on the bandwidth h_{2n}. Using the same arguments as in Chen and Khan (2000, 2001), it can be shown that under Assumptions RD1, S2, H2, and H3 that

$$\sup_{z_i\in\mathcal{Z}} h_{3n}^{-1}\|R_{ni}\| = O_p(h_{3n}^{-1}(nh_{2n}^{k_c})^{-3/4}) = o_p(n^{-1/2}),$$

where the second equality follows from (A.4). Thus to show (A.20), it will suffice to show the asymptotic negligibility of the following third-order U-statistic:

$$\frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k}\tau_i\tau_j d_i d_j h_{3n}^{-1}k\left(\frac{p_i-p_j}{h_{3n}}\right)$$
$$\times(I[y_k\le q^*_{2(k,i)}]-\alpha_2)f_{U_2,Z}(0,z_i)^{-1}h_{2n}^{-k_c}\Delta\tilde{x}_{ij}d_k I[z_k\in C_{ni}]\tilde{u}_i. \tag{A.23}$$

We next replace $q^*_{2(k,i)}$ with q_{2k}. By Assumption S2.1, this replacement results in a remainder term that is $O_p(h_{3n}^{-1}h_{2n}^{M_2})$, which is $o_p(n^{-1/2})$ by (A.6). Let $\chi_i = (y_i,z_i',q_{2i},p_i,u_{2i},\tilde{u}_i,\tau_i,d_i)'$. Let $\mathcal{F}_n(\chi_i,\chi_j,\chi_k)$ denote the expression in the preceding triple summation after this replacement. It follows after a change of variables that

$$E[\|\mathcal{F}_n(\cdot,\cdot,\cdot)\|^2] = O(h_{3n}^{-1}h_{2n}^{-k_c}) = o(n),$$

where the second equality follows from (A.5). We can thus apply the projection theorem for U-statistics in Ahn and Powell (1993). We note that

$$E[\mathcal{F}_n(\cdot,\cdot,\cdot)] = E[\mathcal{F}_n(\chi_i,\cdot,\cdot)] = E[\mathcal{F}_n(\cdot,\chi_j,\cdot)] = E[\mathcal{F}_n(\cdot,\cdot,\chi_k)] = 0,$$

where the last three terms denote expectations conditional on first, second, and third arguments, respectively, and the last term is 0 because of the presence of $\tilde{u}_i$. We thus have by Lemma A.3 in Ahn and Powell (1993) that

$$\frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k}\mathcal{F}_n(\chi_i,\chi_j,\chi_k) = o_p(n^{-1/2}),$$

establishing (A.20).

It thus remains to derive a linear representation for (A.21). To do so we again plug in the linear representation for the conditional quantile estimator; this yields another mean 0 third-order U-statistic after showing that the bias term is asymptotically negligible. We will let $\mathcal{F}_{3n}(\chi_i, \chi_j, \chi_k)$ denote the term in the triple summation of this U-statistic. As before, we have $E[\mathcal{F}_{3n}(\cdot,\cdot,\cdot)] = E[\mathcal{F}_{3n}(\chi_i,\cdot,\cdot)] = E[\mathcal{F}_{3n}(\cdot,\chi_j,\cdot)] = 0$, and will derive a linear representation for

$$\frac{1}{n}\sum_{k=1}^{n} E[\mathcal{F}_{3n}(\cdot,\cdot,\chi_k)]. \tag{A.24}$$

Once again, we ignore the bias component of the nonparametric estimator, as it converges to 0 at the parametric rate. We will thus derive a linear representation for

$$\frac{1}{n}\sum_{k=1}^{n} d_k(I[y_k \le q_{2k}] - \alpha_2)\Xi_{1n}(z_k), \tag{A.25}$$

where

$$\Xi_{1n}(z_k) = \int \tau_i \tau_j \Delta\tilde{x}_{ij} h_{2n}^{-k_c} I[z_k \in C_{ni}]\Delta q_i^{-1} f_{U_2,Z}(0, z_i)^{-1}\, dF(z_i|p_i)\, dF(z_j|p_j)$$
$$\times\ \tilde{\lambda}(p_i)p_i p_j f(p_i) f(p_j)\, dp_i\, dp_j.$$

To derive an expression for (A.25), we will again use Chebyshev's inequality. Since each term in the summation has mean 0, we will show that

$$E[\Xi_{1n}(z_k)] \to -f_{U_2|Z}(0|z_k)\Delta q_k^{-1} f_k \tau_k \tilde{\lambda}(p_k)p_k^2(\tilde{x}_k\mu_{\tau k} - \mu_{\tau x k}). \tag{A.26}$$

To do so, we make a change of variables in the definition of $\Xi_{1n}(z_k)$. Using the change of variables $u_i = (z_k^{(c)} - z_i^{(c)})/h_{2n}$ and $v_i = (p_i - p_j)/h_{3n}$, by the dominated convergence theorem,

$$\Xi_{1n}(z_k) \to \int \tau_k \tau_j(\tilde{x}_j - \tilde{x}_k)\Delta q_k^{-1} f_{U_2,Z}(0, z_k)^{-1}\, dF(z_k|p_j)\, dF(z_j|p_i)\tilde{\lambda}(p_i)p_i^2 f(p_i)^2\, dp_i.$$

Expression (A.26) follows by noting that p_i is a function of z_i. Thus we have shown that (A.21) has the following linear representation:

$$\frac{1}{n}\sum_{i=1}^{n} -\tau_i f_{U_2|Z}(0|z_i)^{-1} d_i(I[y_i \le q_{2i}] - \alpha_2)\Delta q_i^{-1} f_i \tilde{\lambda}(p_i)p_i^2(\tilde{x}_i\mu_{\tau i} - \mu_{\tau x i}) + o_p(n^{-1/2}), \tag{A.27}$$

and identical arguments can be used to show that double summations involving $(\hat{q}_{2j} - q_{2j})$, $(\hat{q}_{1i} - q_{1i})$, and $(\hat{q}_{1j} - q_{1j})$ have analogous linear representations. Combining these results, we get that (A.13) has the following linear representation:

$$\frac{1}{n}\sum_{i=1}^{n} -2\tau_i d_i(f_{U_2|Z}(0|z_i)^{-1}(I[y_i \le q_{2i}] - \alpha_2) - f_{U_1|Z}(0|z_i)^{-1}(I[y_i \le q_{1i}] - \alpha_1))$$
$$\times\ \Delta q_i^{-1} f_i \tilde{\lambda}(p_i)p_i^2(\tilde{x}_i\mu_{\tau i} - \mu_{\tau x i}) + o_p(n^{-1/2}). \tag{A.28}$$

The remaining step involves showing that

$$\frac{1}{n(n-1)}\sum_{i\neq j}\tau_i\tau_j d_i d_j h_{3n}^{-1}k\left(\frac{\hat{p}_i-\hat{p}_j}{h_{3n}}\right)(\Delta\hat{\tilde{x}}_{ij}-\Delta\tilde{x}_{ij})(\Delta\hat{\tilde{y}}_{ij}-\Delta\hat{\tilde{x}}_{ij}'\beta_0)=o_p(n^{-1/2}). \quad \textbf{(A.29)}$$

This can be done by using similar arguments as before, so we only sketch the details. First, we note that we can replace $h_{3n}^{-1}k\big((\hat{p}_i-\hat{p}_j)/h_{3n}\big)$ with $h_{3n}^{-1}k\big((p_i-p_j)/h_{3n}\big)$, and the remainder term is asymptotically negligible as shown previously. We can also replace $(\Delta\hat{\tilde{y}}_{ij}-\Delta\hat{\tilde{x}}_{ij}'\beta_0)$ with $(\Delta\tilde{y}_{ij}-\Delta\tilde{x}_{ij}'\beta_0)$, and the remainder terms are of order $(\Delta\hat{q}_i-\Delta q_i)^2$ and $(\Delta\hat{q}_j-\Delta q_j)^2$, which are (uniformly) of order $o_p(n^{-1/2})$. Thus we will show

$$\frac{1}{n(n-1)}\sum_{i\neq j}\tau_i\tau_j d_i d_j h_{3n}^{-1}k\left(\frac{p_i-p_j}{h_{3n}}\right)(\Delta\hat{\tilde{x}}_{ij}-\Delta\tilde{x}_{ij})(\Delta\tilde{y}_{ij}-\Delta\tilde{x}_{ij}'\beta_0)=o_p(n^{-1/2}). \quad \textbf{(A.30)}$$

Note that we can replace $(\Delta\tilde{y}_{ij}-\Delta\tilde{x}_{ij}'\beta_0)$ with $\tilde{\lambda}(p_j)-\tilde{\lambda}(p_i)+\Delta\tilde{u}_{ij}$. We also note that we need only show negligibility of the term involving $\hat{\tilde{x}}_j-\tilde{x}_j$ because the same argument can be used for $\hat{\tilde{x}}_i-\tilde{x}_i$. We can write $\hat{\tilde{x}}_j-\tilde{x}_j$ as $x_j(\hat{q}_j^{-1}-q_j^{-1})$ and as before linearize the difference and plug in the linear representation of the conditional quantile estimator, yielding a centered third-order U-statistic plus a remainder term that is $o_p(n^{-1/2})$. The term in the U-statistic involving $\Delta\tilde{u}_{ij}$ is negligible by the same arguments as used in showing (A.23). The term in the U-statistic involving $\tilde{\lambda}(p_j)-\tilde{\lambda}(p_i)$ is also negligible, as a result of presence of $h_{3n}^{-1}k\big((p_i-p_j)/h_{3n}\big)$, the smoothness assumption on $\tilde{\lambda}(\cdot)$ (Assumption S1.1), and the rate condition on h_{3n} (Assumption H3). This shows (A.30) and hence (A.29).

This proves the lemma and establishes the asymptotic difference between the feasible and infeasible estimators, as stated in Lemma 4.1. ∎

The final step in deriving the limiting distribution of the estimator is to establish a linear representation for S_{xy}. This is done in the following lemma, whose proof is omitted because it follows from identical arguments used in the proof of Theorem 3.1(ii) of Ahn and Powell (1993).

LEMMA A.5. *Under Assumptions K3, H3, S1, RD2, PS, K1, and H1,*

$$S_{xy}-\frac{1}{n}\sum_{i=1}^{n}\psi_{1i}=o_p(n^{-1/2}). \quad \textbf{(A.31)}$$

The limiting distribution in Theorem 4.1 now easily follows.

Proof of Theorem 4.2. We note that the consistency of $\hat{\Sigma}_{xx}$ follows easily from Lemma A.3. To show consistency of $\hat{\Omega}_{xx}$ we first define $\tilde{\zeta}_i$ as

$$\tilde{\zeta}_i=2\tau_i p_i f_i(\tilde{x}_i\mu_{\tau i}-\mu_{\tau xi})\cdot p_i\tilde{\lambda}'(p_i).$$

We show the following lemma.

LEMMA A.6. *Under the assumptions,*

$$\sup_{\tau_i>0}\|\tilde{\zeta}_i-\hat{\tilde{\zeta}}_i\|=o_p(1).$$

Proof. First we note that by standard results from kernel estimation (e.g., Bierens, 1987) the expression

$$\frac{1}{n}\sum_{i=1}^{n} h_{1n}^{-w_c} K\left(\frac{w_l - w_i}{h_{1n}}\right)$$

converges uniformly (in w_i) in probability to the density of w_i, $f_W(w_i)$, so we focus on the expression

$$\frac{1}{n^2 h_{3n}^2 h_{1n}^{w_c}} \sum_{l \neq j} \tau_l \tau_j d_l d_j k'\left(\frac{\hat{p}_l - \hat{p}_j}{h_{3n}}\right) K\left(\frac{w_l - w_i}{h_{1n}}\right) \Delta\hat{\tilde{v}}_{lj} \Delta\hat{\tilde{x}}_{lj}.$$

We decompose $\Delta\hat{\tilde{v}}_{lj} \Delta\hat{\tilde{x}}_{lj}$ as

$$\Delta\hat{\tilde{y}}_{lj} \Delta\hat{\tilde{x}}_{lj} - \Delta\hat{\tilde{x}}_{lj}' \hat{\beta}_0 \Delta\hat{\tilde{x}}_{lj}$$

and decompose $\hat{\tilde{\zeta}}_i$ accordingly. Concentrating on the first term after this decomposition, we note that by the bounds on $k'(\cdot), K(\cdot)$, the trimming functions, and x_l, x_j in the support of the trimming functions, and also by the moment conditions on y_l, y_j, we can replace $\Delta\hat{\tilde{y}}_{lj} \Delta\hat{\tilde{x}}_{lj}$ with $\Delta\tilde{y}_{lj} \Delta\tilde{x}_{lj}$, and the resulting remainder term (uniformly in i) is $O_p(h_{3n}^{-2} h_{1n}^{-w_c} \|\hat{q}_\alpha - q_\alpha\|_\infty)$, which by (A.3) and Assumptions H1 and H3, is $o_p(1)$. Similarly, exploiting the result that $\hat{\beta} - \beta_0 = O_p(n^{-1/2})$ the second term (involving $\Delta\hat{\tilde{x}}_{lj}' \hat{\beta}_0 \Delta\hat{\tilde{x}}_{lj}$) can be replaced with $\Delta\tilde{x}_{lj}' \beta_0 \Delta\tilde{x}_{lj}$ with a resulting remainder term that is uniformly $o_p(1)$. We can thus turn our attention to the expression

$$\frac{1}{n^2 h_{3n}^2 h_{1n}^{w_c}} \sum_{l \neq j} \tau_l \tau_j d_l d_j k'\left(\frac{\hat{p}_l - \hat{p}_j}{h_{3n}}\right) K\left(\frac{w_l - w_i}{h_{1n}}\right) \Delta\tilde{v}_{lj} \Delta\tilde{x}_{lj}.$$

We replace $\hat{p}_l - \hat{p}_j$ with $p_l - p_j$. Again exploiting boundedness (this time involving the second derivative of $k(\cdot)$) and moment conditions, this time the remainder term is uniformly $O_p(\|\hat{p} - p\|_\infty h_{3n}^{-3} h_{1n}^{-w_c})$, which also is $o_p(1)$ by Assumptions H1 and H3, and Lemma A.1. We are thus left with the expression

$$\frac{1}{n^2 h_{3n}^2 h_{1n}^{w_c}} \sum_{l \neq j} \tau_l \tau_j d_l d_j k'\left(\frac{p_l - p_j}{h_{3n}}\right) K\left(\frac{w_l - w_i}{h_{1n}}\right) \Delta\tilde{v}_{lj} \Delta\tilde{x}_{lj}.$$

We treat w_i as fixed and multiply the preceding expression by $n/(n-1)$; we have a second-order U-statistic. We note that the expectation of the squared norm of this U-statistic is $O(h_{3n}^{-3} h_{1n}^{-w_c})$, which is $o(n)$ by Assumptions H1 and H3. Thus by Lemma A.3(i) in Ahn and Powell (1993), the U-statistic converges in probability to the expected value of the term in the double summation divided by $h_{3n}^2 h_{1n}^{w_c}$. Working with this sequence of expected values, the usual change of variables, and the bounded convergence theorem, the sequence of expected values converges to $\tilde{\zeta}_i f_W(w_i)$. Furthermore, it follows by uniform laws of large numbers for U-statistics (see, e.g., Sherman, 1994) that this convergence is uniform in w_i. This establishes the lemma. ■

Next we define

$$\psi_{11i} = 2\tau_i p_i f_i(\tilde{x}_i \mu_{\tau i} - \mu_{\tau x i})\cdot(d_i \tilde{u}_i)$$

and

$$\psi_{12i} = 2\tau_i p_i f_i(\tilde{x}_i \mu_{\tau i} - \mu_{\tau x i})\cdot(p_i \tilde{\lambda}'(p_i)(d_i - p_i)).$$

An immediate consequence of the lemma, using the uniform convergence of $\hat{p}_i$ to p_i, is that $\hat{\psi}_{12i}$ converges uniformly in i to ψ_{12i}. We note that similar, though simpler, arguments used in Lemma A.6 can be used to establish the uniform convergence of $\hat{\psi}_{11i}$ to ψ_{1i}. As a final step we show uniform convergence of $\hat{\psi}_{2i}$ to ψ_{2i}. We first show uniform convergence of $\hat{f}_{U_2|Z}(z_i)$ to $f_{U_2|Z}(z_i)$. A mean value expansion of $\tilde{k}(\hat{u}_{2i}/\nu_n)$ around $\tilde{k}(u_{2i}/\nu_n)$ yields a remainder term of order $\nu_n^{-2}\|\hat{q}_2 - q_2\|_\infty$, which is $o_p(1)$ by the conditions on ν_n and Lemma A.2.

Also, similar, though simpler, arguments than used in Lemma A.6 can be used to establish uniform convergence of

$$\left(\frac{1}{nh_{3n}}\sum_{j=1}^{n}\tau_j d_j k\left(\frac{\hat{p}_i - \hat{p}_j}{h_{3n}}\right)\hat{\hat{v}}_j \Delta\hat{\hat{x}}_{ij}\right)$$

to $f_i \tilde{\lambda}(p_i)p_i(\tilde{x}_i \mu_{\tau i} - \mu_{\tau x i})$. The uniform convergence of $\hat{\psi}_{2i}$ to ψ_{2i} then follows from the uniform convergence of $\hat{p}_i, \hat{q}_{1i}, \hat{q}_{2i}$ to p_i, q_{1i}, q_{2i}, respectively.

It immediately follows that

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\psi}_{1i} + \hat{\psi}_{2i})(\hat{\psi}_{1i} + \hat{\psi}_{2i})' = \frac{1}{n}\sum_{i=1}^{n}(\psi_{1i} + \psi_{2i})(\psi_{1i} + \psi_{2i})' + o_p(1).$$

Thus by the law of large numbers, we have $\hat{\Omega}_{xx} \xrightarrow{p} \Omega_{xx}$. Slutsky's theorem then implies Theorem 4.2. ∎