

# Endogeneity in Quantile Regression Models: A Control Function Approach

Sokbae Lee \*

Centre for Microdata Methods and Practice  
Institute for Fiscal Studies  
and  
Department of Economics  
University College London  
London, WC1E 6BT, UK  
l.simon@ucl.ac.uk

January 2004

## Abstract

This paper considers a linear triangular simultaneous equations model with conditional quantile restrictions. The paper adjusts for endogeneity by adopting a control function approach and presents a simple two-step estimator that exploits the partially linear structure of the model. The first step consists of estimation of the residuals of the reduced-form equation for the endogenous explanatory variable. The second step is series estimation of the primary equation with the reduced-form residual included non-parametrically as an additional explanatory variable. This paper imposes no functional form restrictions on the stochastic relationship between the reduced-form residual and the disturbance term in the primary equation conditional on observable explanatory variables. The paper presents regularity conditions for consistency and asymptotic normality of the two-step estimator. In addition, the paper provides some discussions on related estimation methods in the literature and on possible extensions and limitations of the estimation approach. Finally, the numerical performance and usefulness of the estimator are illustrated by the results of Monte Carlo experiments and two empirical examples, demand for fish and returns to schooling.

*Key words:* Endogeneity, partially linear regression, quantile regression, series estimation.

---

\*I would like to thank David Blau, Richard Blundell, Pedro Carneiro, Andrew Chesher, Joel Horowitz, Hide Ichimura, Francis Kramarz, Roger Koenker, and seminar participants at CREST for helpful comments. Special thanks go to Andrew Chesher for encouraging me to work on this and other related quantile regression projects. I am also grateful to Kathryn Graddy for kindly providing me with her data on demand for fish. Research support from ESRC Grant (RES-000-22-0704) is gratefully acknowledged.

# Endogeneity in Quantile Regression Models: A Control Function Approach

## 1 Introduction

This paper is concerned with estimating a structural quantile regression model. In particular, this paper considers a semiparametric quantile regression version of triangular simultaneous equations models. Structural quantile regression models have been previously considered in Amemiya (1982), Powell (1983), Chernozhukov and Hansen (2001), Abadie, Angrist, and Imbens (2002), Imbens and Newey (2003), Hong and Tamer (2003), Chesher (2003), Ma and Koenker (2003), Chen, Linton, and Van Keilegom (2003), and Honoré and Hu (2003). Amemiya (1982) and Powell (1983) gave a large sample theory for two-stage least-absolute deviations estimators. Chernozhukov and Hansen (2001) and Abadie, Angrist, and Imbens (2002) developed models of quantile treatment effects. Imbens and Newey (2003) investigated identification and estimation of nonseparable, nonparametric triangular simultaneous equations models (including quantile treatment effects as a special case) by extending the ‘control function’ approach of Newey, Powell, and Vella (1999). Hong and Tamer (2003) considered identification and estimation of endogenous linear median regression models with censoring. Chesher (2003) provided important identification results for nonseparable models using conditional quantile restrictions. Ma and Koenker (2003) applied Chesher’s identification results to a nonseparable parametric model and also developed a control function method for the parametric model. Chen, Linton, and Van Keilegom (2003) considered a partially linear median regression model with some endogenous regressors. Honoré and Hu (2003) considered an instrumental variable estimator of linear quantile regression models while assuming directly the identification of the model.

As an alternative to existing methods in the literature, this paper aims to extend the control function approach to structural quantile regression models *semiparametrically*. The model we consider has the form

$$(1) \quad \begin{aligned} Y &= X\beta(\tau) + Z_1'\gamma(\tau) + U, \\ X &= \mu(\alpha) + Z'\pi(\alpha) + V, \end{aligned}$$

where  $Y$  is a real-valued dependent variable,  $X$  is a real-valued, continuously distributed, endogenous explanatory variable,  $Z \equiv (Z_1, Z_2)$  is a  $(d_z \times 1)$  vector of exogenous explanatory variables,  $U$  and  $V$  are real-valued unobserved random variables,  $\beta(\tau)$  and  $\gamma(\tau)$  are unknown

structural parameters of interest, and  $\mu(\alpha)$  is an unknown parameter,  $\pi(\alpha) \equiv [\pi_1(\alpha), \pi_2(\alpha)]$  is a  $(d_z \times 1)$  vector of unknown parameters for some  $\tau$  and  $\alpha$  such that  $0 < \tau < 1$  and  $0 < \alpha < 1$ . For identification it is assumed that there is at least one component of  $Z$  that is not included in  $Z_1$ , and that there is at least one non-zero coefficient for the excluded components of  $Z$ . That is,  $d_{z1} < d_z$  and  $\pi_2(\alpha) \neq 0$ , where  $d_{z1}$  is the dimension of  $Z_1$ . To complete the model, assume that

$$(2) \quad Q_{U|X,Z}(\tau|x, z) = Q_{U|V,Z}(\tau|v, z) = Q_{U|V}(\tau|v) \equiv \lambda_\tau(v) \quad \text{and}$$

$$(3) \quad Q_{V|Z}(\alpha|z) = 0$$

almost surely, where  $\lambda_\tau(\cdot)$  is a real-valued, unknown function of  $V$ ,  $Q_{U|X,Z}(\tau|x, z)$  denotes the  $\tau$ -th quantile of  $U$  conditional on  $X = x$  and  $Z = z$ , and the other expressions are understood similarly. The first equality in (2) holds when  $v$  is the value of  $V$  that satisfies  $v = x - \mu(\alpha) - z'\pi(\alpha)$ . The second equality in (2) assumes a quantile independence of  $U$  on  $Z$  conditional on  $V$ . The model (1)-(3) is a semiparametric quantile regression version of Newey, Powell, and Vella (1999).

To give a specific example of (1)-(3), consider a simple model for log earnings of the following form

$$(4) \quad \log Y = S\beta + U, \quad \text{and} \quad S = Z'\pi + V,$$

where  $Y$  denotes earnings,  $S$  schooling, and  $Z$  a set of observables. In addition, the unobserved random variable  $U$  is an individual-specific random component of the earnings model. Conventionally, the reduced-form schooling residual  $V$  is interpreted as ‘individual ability’ and therefore  $U$  is not assumed to be independent of  $V$ . In this example, making assumption (2) amounts to assuming that the  $\tau$ -th quantile of the individual-specific random component,  $U$ , conditional on schooling and other observables is a smooth function of individual ability,  $V$ , that is schooling residual from the  $\alpha$ -th quantile regression on observables. See, for example, Card (2001, Section 2.C) for discussions on different control function approaches in the schooling context. Section 4 provides more discussions on implications of making assumption (2).

Under assumptions (2)-(3),

$$(5) \quad Q_{Y|X,Z}(\tau|x, z) = x\beta(\tau) + z_1'\gamma(\tau) + \lambda_\tau(v) \quad \text{and}$$

$$(6) \quad Q_{X|Z}(\alpha|z) = \mu(\alpha) + z'\pi(\alpha).$$

This suggests that  $\beta(\tau)$  and  $\gamma(\tau)$  could be estimated by a partially linear ( $\tau$ -th) quantile regression of  $Y$  on  $(X, Z_1, V)$ . In applications,  $V$  is unobserved; however,  $V$  can be estimated consistently by the residual of a linear ( $\alpha$ -th) quantile regression of  $X$  on  $(1, Z)$ . Therefore,  $\beta(\tau)$  and  $\gamma(\tau)$  can be estimated by a two-step procedure. The first step is construction of estimated residuals  $\hat{V}$  from the linear quantile regression of  $X$  on  $(1, Z)$ . The second step is the partially linear regression of  $Y$  on  $X, Z_1$ , and  $\hat{V}$ . This approach corrects for endogeneity by adding estimates of  $V$  as an additional explanatory variable and, therefore, can be viewed as a variant of control function approach (e.g., Newey, Powell, and Vella (1999), Imbens and Newey (2003), and Blundell and Powell (2003b)). A partially linear structure in (5) is motivated by the fact that it is quite difficult to assume the functional form of stochastic relationship between two unobserved variables  $U$  and  $V$  conditional on  $Z$ . It will be shown in Section 3 that the proposed two-step estimator is  $n^{-1/2}$ -consistent and asymptotically normal.

The remainder of this paper is as follows. Section 2 provides an informal description of the two-step estimator. Section 3 gives asymptotic results for the estimator. Section 4 provides some discussions on related estimation methods in the literature. In addition, Section 4 outlines some possible extensions and discusses some limitations of our estimation approach. Section 5 reports results of some Monte Carlo experiments. Section 6 illustrates the estimation method by applying it to data on demand for fish as well as data on the returns to schooling. Section 7 concludes. All the proofs are in the Appendix.

## 2 Estimation

The estimation procedure consists of two steps. The data consist of i.i.d. observations  $\{(Y_i, X_i, Z_i) : i = 1, \dots, n\}$ . The first step is construction of estimated residuals  $\hat{V}_i = X_i - \hat{\mu}(\alpha) - Z_i' \hat{\pi}(\alpha)$  ( $i = 1, \dots, n$ ) by a linear quantile regression of  $X$  on  $(1, Z)$ , where  $(\hat{\mu}(\alpha), \hat{\pi}(\alpha))$  is a solution to

$$(7) \quad \min_{\mu, \pi} n^{-1} \sum_{i=1}^n \rho_\alpha(X_i - \mu - Z_i' \pi),$$

where  $\rho_\alpha(\cdot)$  is the ‘check’ function such that  $\rho_\alpha(u) = |u| + (2\alpha - 1)u$  for  $0 < \alpha < 1$ .

The second step is estimation of a partially linear quantile regression of  $Y$  on  $(X, Z_1, V)$  using the estimated residuals  $\hat{V}_i$  in place of unobserved  $V_i$ 's. In this paper the second step is carried out via series estimation. To describe the second step, let  $W = (X, Z_1', V)'$ ,

$\hat{W} = (X, Z'_1, \hat{V})'$ ,  $W_i = (X_i, Z'_{1i}, V_i)'$ , and  $\hat{W}_i = (X_i, Z'_{1i}, \hat{V}_i)'$ . Also, let  $\{p_k : k = 1, 2, \dots\}$  denote a basis for smooth functions such that a linear combination of  $\{p_k : k = 1, 2, \dots\}$  can approximate  $\lambda_\tau(\cdot)$ . For any positive integer  $\kappa$ , define

$$P_\kappa(w) = [x, z_1, p_1(v), \dots, p_\kappa(v)]'.$$

Let  $d_w$  denote the dimension of  $W$ , and let  $1(\cdot)$  denote the usual indicator function. Define  $t(w) = 1(w \in \mathcal{W}) \equiv \prod_{j=1}^{d_w} 1(\underline{w}^{(j)} \leq w^{(j)} \leq \bar{w}^{(j)})$ , where  $\mathcal{W}$  is a  $(d_w)$ -dimensional rectangle for which  $\underline{w}^{(j)}$  and  $\bar{w}^{(j)}$  are predetermined finite constants and  $w^{(j)}$  is the  $j$ -th component of  $w$ . As in Newey, Powell, and Vella (1999),  $t(w)$  is a trimming function that is useful not only to derive the asymptotic properties of the estimator but also to avoid undue influences of large values of  $W$ .

Let  $\hat{\theta}_{n\kappa}(\tau)$  be a solution to

$$(8) \quad \min_{\theta} S_{n\kappa}(\theta) \equiv n^{-1} \sum_{i=1}^n t(\hat{W}_i) \rho_\tau[Y_i - P_\kappa(\hat{W}_i)' \theta],$$

where  $\rho_\tau(\cdot)$  is again the check function such that  $\rho_\tau(u) = |u| + (2\tau - 1)u$  for  $0 < \tau < 1$ . Let  $d(\kappa)$  denote the dimension of  $P_\kappa(w)$ , that is  $d(\kappa) = 1 + d_{z1} + \kappa$ , and  $A$  denote the  $[(1 + d_{z1}) \times d(\kappa)]$  matrix such that  $A = (I_{1+d_{z1}}, 0_\kappa)$ , where  $I_{1+d_{z1}}$  is the  $(1 + d_{z1})$ -dimensional identity matrix and  $0_\kappa$  is the  $[(1 + d_{z1}) \times \kappa]$  matrix of zeros. Then the estimator of  $(\beta(\tau), \gamma(\tau)')'$  is defined as  $(\hat{\beta}(\tau), \hat{\gamma}(\tau)')' = A\hat{\theta}_{n\kappa}(\tau)$ . That is,  $\hat{\beta}(\tau)$  and  $\hat{\gamma}(\tau)$  are the first  $(1 + d_{z1})$  components of  $\hat{\theta}_{n\kappa}(\tau)$ . This two-step estimator resembles closely the approach of Buchinsky (1998b) in which the sample selection bias is corrected for nonparametrically by a two-stage procedure.

We conclude this section by mentioning computational aspects of the proposed two-step estimator. The second step minimization in (8) has a linear programming representation and, therefore, can be solved easily by computation methods developed for linear quantile regression models.

### 3 Asymptotic Theory

This section gives the asymptotic theory for the estimator described in Section 2. Following Newey (1997) and Newey, Powell, and Vella (1999), regularity conditions for approximating functions are stated below in terms of power series and regression splines. For any matrix

$A$ , let  $\|A\| = [\text{trace}(A'A)]^{1/2}$  be the Euclidean norm. Define  $\zeta_0(\kappa) = \sup_{w \in \mathcal{W}} \|P_\kappa(w)\|$  and  $\zeta_1(\kappa) = \sup_{w \in \mathcal{W}} \|\partial P_\kappa(w)/\partial v\|$ , where  $\partial P_\kappa(w)/\partial v$  denotes a vector of partial derivatives of  $P_\kappa(w)$  with respect to  $v$ . It is well known (see, for example, Newey (1997)) that for power series  $\zeta_0(\kappa) \leq C\kappa$  and  $\zeta_1(\kappa) \leq C\kappa^3$  and for splines  $\zeta_0(\kappa) \leq C\kappa^{1/2}$  and  $\zeta_1(\kappa) \leq C\kappa^{3/2}$ , where  $C$  is a generic positive constant.

To describe asymptotic results for the estimator, let  $m_\tau(w) = x\beta(\tau) + z_1'\gamma(\tau) + \lambda_\tau(v)$  and  $\varepsilon_\tau = Y - m_\tau(W)$ . Also, let  $F_{\varepsilon_\tau}(\cdot|x, z)$  and  $f_{\varepsilon_\tau}(\cdot|x, z)$ , respectively, denote the cumulative distribution function and probability density function of  $\varepsilon_\tau$  conditional on  $X = x$  and  $Z = z$ . Define

$$q(w) = \left[ \left( x - \frac{E[t(W)f_{\varepsilon_\tau}(0|X, Z)X|v]}{E[t(W)f_{\varepsilon_\tau}(0|X, Z)|v]} \right), \left( z_1 - \frac{E[t(W)f_{\varepsilon_\tau}(0|X, Z)Z_1|v]}{E[t(W)f_{\varepsilon_\tau}(0|X, Z)|v]} \right) \right]' \text{ and}$$

$$\varphi(w) = (E[t(W)f_{\varepsilon_\tau}(0|X, Z)q(W)q(W)'])^{-1}q(w),$$

assuming the inverse exists. As in Newey (1997) and Newey, Powell, and Vella (1999), it is useful to represent  $(\beta(\tau), \gamma(\tau)')$  as an expected product form. Specifically, it can be shown that

$$(\beta(\tau), \gamma(\tau)')' = E[t(W)f_{\varepsilon_\tau}(0|X, Z)\varphi(W)m_\tau(W)] \quad \text{and}$$

$$(\hat{\beta}(\tau), \hat{\gamma}(\tau)')' = A\hat{\theta}_{n\kappa}(\tau),$$

where  $A = (I_{1+d_{z_1}}, 0_\kappa) = E[t(W)f_{\varepsilon_\tau}(0|X, Z)\varphi(W)P_\kappa(W)']$ .

To establish the  $n^{-1/2}$ -consistency and asymptotic normality of  $(\hat{\beta}(\tau), \hat{\gamma}(\tau)')$ , we make the following assumptions:

**Assumption 3.1.** *The data  $\{(Y_i, X_i, Z_i) : i = 1, \dots, n\}$  are i.i.d.*

**Assumption 3.2.** *The dimension  $d_z$  of  $Z$  is larger than the dimension  $d_{z_1}$  of  $Z_1$ . Also,  $\pi_2(\alpha) \neq 0$ .*

These familiar exclusion and inclusion conditions are necessary for identification.

**Assumption 3.3.** *The rectangle  $\mathcal{W}$  has a nonempty interior and is contained in the interior of the support of  $W$ .*

Let  $\mathcal{V} = 1(\underline{w}^{(d_w)} \leq w^{(d_w)} \leq \overline{w}^{(d_w)})$ . That is,  $\mathcal{V}$  is a trimming function on a compact interval corresponding to  $V$ .

**Assumption 3.4.** *The distribution of  $V$  is absolutely continuous with respect to Lebesgue measure. The density of  $V$  is bounded away from zero on  $\mathcal{V}$ .*

This restriction requires that the endogenous variable  $X$  be continuously distributed.

**Assumption 3.5.**  $\lambda_\tau(v)$  is  $r$ -times continuously differentiable on  $\mathcal{V}$ .

Assumptions 3.4 and 3.5 imply that for both power series and splines, there exists  $\theta_{\kappa 0}(\tau) \in \mathbf{R}^{d(\kappa)}$  such that (a) the first  $(1+d_{z1})$  components of  $\theta_{\kappa 0}(\tau)$  are equal to  $(\beta(\tau), \gamma(\tau))'$ , so that  $(\beta(\tau), \gamma(\tau))' = A\theta_{\kappa 0}(\tau)$ , (b)  $\sup_{w \in \mathcal{W}} |x\beta(\tau) + z_1'\gamma(\tau) + \lambda_\tau(v) - P_\kappa(w)'\theta_{\kappa 0}(\tau)| = O(\kappa^{-r})$ , and (c)  $\sup_{w \in \mathcal{W}} |\partial\lambda_\tau(v)/\partial v - [\partial P_\kappa(w)/\partial v]'\theta_{\kappa 0}(\tau)| = O[\kappa^{-(r-1)}]$  (See, for example, Newey (1997)).

**Assumption 3.6.** For almost every  $x$  and  $z$ ,  $F_{\varepsilon_\tau}(0|x, z) = \tau$ . There is a positive constant  $c_1 < \infty$  such that  $|f_{\varepsilon_\tau}(\varepsilon_1|x, z) - f_{\varepsilon_\tau}(\varepsilon_2|x, z)| \leq c_1|\varepsilon_1 - \varepsilon_2|$  for all  $\varepsilon_1$  and  $\varepsilon_2$  in a neighborhood of zero and for all  $x$  and  $z$ . Also, there are constants  $c_2 > 0$  and  $c_3 < \infty$  such that  $c_2 \leq f_{\varepsilon_\tau}(\varepsilon|x, z) \leq c_3$  for all  $\varepsilon$  in a neighborhood of zero and for all  $x$  and  $z$ .

Among other things, Assumption 3.6 requires that  $f_{\varepsilon_\tau}(\cdot|x, z)$  be bounded away from zero in a neighborhood of zero uniformly over  $x$  and  $z$ .

**Assumption 3.7.** Let  $\Phi_\kappa = E[t(W)f_\varepsilon(0|X, Z)P_\kappa(W)P_\kappa(W)']$ . The smallest eigenvalue of  $\Phi_\kappa$  is bounded away from zero for all  $\kappa$ , and the largest eigenvalue of  $\Phi_\kappa$  is bounded for all  $\kappa$ .

**Assumption 3.8.** The matrix  $E[t(W)f_{\varepsilon_\tau}(0|X, Z)q(W)q(W)']$  is nonsingular.

Assumptions 3.7 and 3.8 insure the non-singularity of the second moment matrix of the estimator.

**Assumption 3.9.** As functions of  $v$ ,  $E[t(W)f_{\varepsilon_\tau}(0|X, Z)|v]$ ,  $E[t(W)f_{\varepsilon_\tau}(0|X, Z)X|v]$ , and  $E[t(W)f_{\varepsilon_\tau}(0|X, Z)Z_1|v]$  are continuously differentiable.

Assumption 3.9 implies that for both power series and splines, there exists a sequence of  $[(1+d_{z1}) \times d(\kappa)]$  matrices  $\tilde{\Theta}_\kappa$  such that

$$(9) \quad E \left[ t(W)f_{\varepsilon_\tau}(0|X, Z) \left\| \varphi(W) - \tilde{\Theta}_\kappa P_\kappa(W) \right\|^2 \right] \rightarrow 0$$

as  $\kappa \rightarrow \infty$ .

**Assumption 3.10.** There is a  $[(d_z + 1) \times 1]$ -vector-valued function  $\Delta_{\mu, \pi}(x, z)$  such that

$$(a) \quad E[\Delta_{\mu, \pi}(X, Z)] = 0,$$

(b) the components of  $\Sigma_{\mu,\pi} \equiv E[\Delta_{\mu,\pi}(X, Z)\Delta_{\mu,\pi}(X, Z)']$  are finite, and

(c) as  $n \rightarrow \infty$ ,

$$\begin{pmatrix} \hat{\mu}(\alpha) - \mu(\alpha) \\ \hat{\pi}(\alpha) - \pi(\alpha) \end{pmatrix} = n^{-1} \sum_{i=1}^n \Delta_{\mu,\pi}(X_i, Z_i) + o_p(n^{-1/2}).$$

Assumption 3.10 imposes regularity conditions for the first step estimation. These conditions are satisfied by the linear quantile regression estimator (see, e.g., Koenker and Bassett (1978)). The  $n^{-1/2}$  consistency of  $\hat{\mu}(\alpha)$  and  $\hat{\pi}(\alpha)$  implies that the estimated residuals satisfy

$$(10) \quad \max_{1 \leq i \leq n} t(W_i) |\hat{V}_i - V_i| = O_p(n^{-1/2}).$$

**Assumption 3.11.** For power series  $\kappa = C_1 n^{\nu_1}$  for some constants  $C_1$  satisfying  $0 < C_1 < \infty$  and some  $\nu_1$  satisfying  $1/(2r) < \nu_1 < 1/8$ , and for splines  $\kappa = C_2 n^{\nu_2}$  for some constants  $C_2$  satisfying  $0 < C_2 < \infty$  and some  $\nu_2$  satisfying  $1/(2r) < \nu_2 < 1/5$ .

This condition restricts the growth rate of  $\kappa$ . For power series the necessary smoothness condition is that  $r \geq 5$ , and for splines the condition is that  $r \geq 3$ . Define

$$\begin{aligned} \Omega &= \tau(1 - \tau)E[t(W)\varphi(W)\varphi(W)'] \\ &+ E \left[ t(W) f_{\varepsilon_\tau}(0|X, Z) \frac{d\lambda(V)}{dv} \varphi(W)(1, Z') \right] \Sigma_{\mu,\pi} E \left[ t(W) f_{\varepsilon_\tau}(0|X, Z) \frac{d\lambda(V)}{dv} \begin{pmatrix} 1 \\ Z \end{pmatrix} \varphi(W)' \right]. \end{aligned}$$

The following theorem gives the main result of this paper.

**Theorem 3.1.** Let Assumptions 3.1 - 3.11 hold. Then as  $n \rightarrow \infty$ ,

$$\sqrt{n} \begin{pmatrix} \hat{\beta}(\tau) - \beta(\tau) \\ \hat{\gamma}(\tau) - \gamma(\tau) \end{pmatrix} \rightarrow_d \mathbf{N}(0, \Omega).$$

Theorem 3.1 states that the two-step estimator is  $n^{-1/2}$ -consistent and asymptotically normal with mean zero and variance  $\Omega$ . The second component of  $\Omega$  is nonnegative definite and, therefore,  $\Omega$  is in general larger than the first component  $\tau(1 - \tau)E[t(W)\varphi(W)\varphi(W)']$ , which would be the asymptotic variance matrix if  $V_i$  were observed.

To carry out asymptotic inference based on Theorem 3.1, it is necessary to obtain a consistent estimator of  $\Omega$ . As in Powell (1984,1986) and Buchinsky (1998a),  $\Omega$  can be estimated by a sample analog estimator using the kernel method. Let  $K(\cdot)$  denote a kernel function and  $h_n$  a sequence of bandwidths. Also, define  $\hat{\varepsilon}_{\tau,i} = Y_i - X_i \hat{\beta}(\tau) - Z'_{1i} \hat{\gamma}(\tau) - \hat{\lambda}_\tau(\hat{V}_i)$ ,



where  $\hat{\lambda}_\tau(v)$  is a series estimator of  $\lambda_\tau(v)$ , that is  $\hat{\lambda}_\tau(v)$  is the product of  $[p_1(v), \dots, p_\kappa(v)]$  with the appropriate components of  $\hat{\theta}_{n\kappa}(\tau)$ . Then one can estimate  $\Omega$  by  $\hat{\Omega}_{n\kappa}$ :

$$(11) \quad \hat{\Omega}_{n\kappa} = A\hat{\Phi}_{n\kappa}^{-1} \left( \hat{\Sigma}_{n\kappa} + \hat{\Gamma}_{n\kappa} \hat{\Sigma}_{\mu,\pi} \hat{\Gamma}'_{n\kappa} \right) \hat{\Phi}_{n\kappa}^{-1} A',$$

where

$$\begin{aligned} \hat{\Phi}_{n\kappa} &= (nh_n)^{-1} \sum_{i=1}^n t(\hat{W}_i) K \left( \frac{\hat{\varepsilon}_{\tau,i}}{h_n} \right) \hat{P}_{\kappa i} \hat{P}'_{\kappa i}, \\ \hat{\Sigma}_{n\kappa} &= \tau(1-\tau)n^{-1} \sum_{i=1}^n t(\hat{W}_i) \hat{P}_{\kappa i} \hat{P}'_{\kappa i}, \\ \hat{\Gamma}_{n\kappa} &= (nh_n)^{-1} \sum_{i=1}^n t(\hat{W}_i) K \left( \frac{\hat{\varepsilon}_{\tau,i}}{h_n} \right) \frac{d\hat{\lambda}_\tau(\hat{V}_i)}{dv} \hat{P}_{\kappa i} (1, Z'_i), \end{aligned}$$

and  $\hat{\Sigma}_{\mu,\pi}$  is a consistent estimator of  $\Sigma_{\mu,\pi}$ . It is useful to make additional assumptions to establish the consistency of  $\hat{\Omega}_{n\kappa}$ .

**Assumption 3.12.**  $f_{\varepsilon_\tau}(\varepsilon|x, z)$  is twice continuously differentiable with respect to  $\varepsilon$  in a neighborhood of zero and for all  $x$  and  $z$ .

The smoothness assumption on  $f_{\varepsilon_\tau}(\varepsilon|x, z)$  is necessary to estimate the second component of  $\Omega$  consistently.

**Assumption 3.13.** The kernel function  $K$  has support  $[-1, 1]$ , is bounded and symmetrical about 0, and satisfies  $\int_{-1}^1 K(u)du = 1$ ,  $\int_{-1}^1 uK(u)du = 0$ , and  $\int_{-1}^1 u^2K(u)du < \infty$ .

These are standard restrictions on the kernel function.

**Assumption 3.14.** (1) For power series  $\kappa = C_1 n^{\nu_1}$  for some constants  $C_1$  satisfying  $0 < C_1 < \infty$  and some  $\nu_1$  satisfying  $1/(2r) < \nu_1 < 1/9$ , and for splines  $\kappa = C_2 n^{\nu_2}$  for some constants  $C_2$  satisfying  $0 < C_2 < \infty$  and some  $\nu_2$  satisfying  $1/(2r) < \nu_2 < 1/6$ .

(2)  $h_n = C_h n^{-\mu}$  for some positive finite constant  $C_h$  and some  $\mu$  satisfying  $1/(2\nu_1) < \mu < (1 - 4\nu_1)/4$  for power series or satisfying  $1/(2\nu_2) < \mu < (1 - 4\nu_2)/4$  for splines.

Compared to Assumption 3.11, more stringent restrictions are needed to estimate  $\Omega$ . This is because estimation of  $\Omega$  involves series estimation of the derivative of  $\lambda_\tau$ .

**Assumption 3.15.** There is an estimator of  $\Sigma_{\mu,\pi}$  such that  $\hat{\Sigma}_{\mu,\pi} \rightarrow_p \Sigma_{\mu,\pi}$ .

One may use a kernel estimator of  $\Sigma_{\mu,\pi}$  (e.g., Powell (1986)). See Buchinsky (1998b) for detailed discussions on estimation of  $\Sigma_{\mu,\pi}$ .

The following theorem establishes the consistency of  $\hat{\Omega}_{n\kappa}$ .

**Theorem 3.2.** *Let Assumptions 3.1 - 3.15 hold. Then  $\hat{\Omega}_{n\kappa} \rightarrow_p \Omega$  as  $n \rightarrow \infty$ .*

We conclude this section by considering estimation of  $\lambda_\tau(v)$ . As is noted above,  $\lambda_\tau(v)$  can be estimated by a series estimator  $\hat{\lambda}_\tau(v)$ . However, it is difficult to carry out standard inference for  $\lambda_\tau(v)$  using asymptotic results on the series estimator of  $\lambda_\tau(v)$ . This is because it is difficult to obtain the asymptotic distribution of the series estimator of  $\lambda_\tau(v)$ . One simple alternative is to estimate  $\lambda_\tau(v)$  by carrying out a local polynomial quantile regression of  $Y - X\hat{\beta}(\tau) - Z_1'\hat{\gamma}(\tau)$  on  $\hat{V}$ . Since  $\hat{\beta}(\tau)$ ,  $\hat{\gamma}(\tau)$ , and  $\hat{V}$  are estimated with rates of  $n^{-1/2}$ , the resulting estimator is asymptotically as efficient as an estimator obtained from an infeasible local polynomial quantile regression of  $Y - X\beta(\tau) - Z_1'\gamma(\tau)$  on  $V$  for which the asymptotic distribution is well known (see, for example, Chaudhuri (1991)).

## 4 Alternative Approaches in the Literature

This section compares the two-step estimator with alternative approaches in the literature. In addition, this section outlines some possible extensions and discusses some limitations of our estimation approach.

As was discussed by Blundell and Powell (2003a), there are two major alternative approaches to structural regression models, namely the ‘instrumental variables’ (IV) approach and the ‘fitted value’ approach. First, an IV approach in the context of quantile regression may be referred to a regression model for which it is only assumed that  $Q_{U|Z}(\tau|z)$  is independent of  $z$  (see, e.g., Hong and Tamer (2003), Chen, Linton, and Van Keilegom (2003), and Honoré and Hu (2003)). In this section, we consider a linear regression model of the following form

$$(12) \quad Y = X\beta(\tau) + Z_1'\gamma(\tau) + U, \quad Q_{U|Z}(\tau|z) = \mu_{IV}(\tau),$$

where  $\mu_{IV}(\tau)$  is a constant. This model has the same regression function as that of (1) but imposes a different assumption on  $U$ .

Second, the ‘fitted value’ approach, which is developed by Amemiya (1982) and Powell (1983), replaces  $X$  with the fitted value of  $\mu(\alpha) + Z'\pi(\alpha)$  in (1). To see how the fitted value

approach works, consider the reduced-form equation for  $Y$

$$Y = \beta(\tau)[\mu(\alpha) + Z'\pi(\alpha)] + Z_1'\gamma(\tau) + \eta,$$

where  $\eta = U + \beta(\tau)V$ . In order to estimate  $\beta(\tau)$  and  $\gamma(\tau)$  consistently, the fitted value approach requires that  $Q_{\eta|Z}(\tau|z)$  be independent of  $z$ .

In addition, Chesher (2003) has recently provided important identification results for structural quantile regression models using weaker restrictions than those assumed in this paper. The approach taken by Chesher (2003) can be viewed as a ‘local, nonseparable, nonparametric’ control function approach, whereas our approach is a ‘global, separable, semiparametric’ control function approach. When the assumptions in (1) are palatable and the dimension of  $Z$  is large (including over-identified cases), our approach can be regarded as an alternative to a minimum distance estimator of Chesher (2003). A very recent paper by Ma and Koenker (2003) developed a ‘global, nonseparable, parametric’ control function method.

In what follows, we will consider several possible stochastic relationships between  $U$ ,  $V$ ,  $X$ , and  $Z$ . It will be shown that structural models based on three major approaches are, in general, non-nested. In particular, it will be demonstrated that different types of heteroskedasticity are allowed between alternative approaches. First, we consider a homoskedastic case.

**Case 1:  $U$  and  $V$  are jointly independent of  $Z$ .**

If  $U$  and  $V$  are jointly independent of  $Z$ , then for any  $\tau$ , (a)  $Q_{U|Z}(\tau|z)$  is independent of  $z$ , (b)  $Q_{\eta|Z}(\tau|z)$  is independent of  $z$ , and (c)  $Q_{U|V,Z}(\tau|v, z) = Q_{U|V}(\tau|v)$ . Therefore, under the rather strong independence assumption,  $\beta(\tau)$  and  $\gamma(\tau)$  are constant over  $\tau$  and can be estimated consistently by any of three approaches. Furthermore, in this case, the IV method may be preferred because it does not require specification of a reduced-form equation for  $X$ .

**Case 2: A heteroskedastic partially linear quantile regression model**

As an example, suppose that

$$\begin{aligned} Y &= X\beta + Z_1'\gamma + U, \quad U = \phi_1(V) + [X\varphi_1 + Z_1'\varphi_2 + \phi_2(V)]\tilde{U}, \\ X &= \mu + Z'\pi + V, \quad Q_{V|Z}(\alpha|z) = 0, \end{aligned}$$

where  $\phi_1$  and  $\phi_2$  are real-valued functions of  $V$ , and  $\tilde{U}$  is independent of  $(Z, V)$  with a strictly increasing distribution function  $F_{\tilde{U}}$ . Then, in general,  $Q_{U|Z}(\tau|z)$  and  $Q_{\eta|Z}(\tau|z)$  depend on  $z$  (especially when  $\tau \neq 1/2$ ), implying that neither the IV approach (equation (12)) nor the fitted value approach is applicable. However, note that

$$Q_{Y|X, Z_1, V}(\tau|x, z_1, v) = x(\beta + \varphi_1 F^{-1}(\tau)) + z'_1(\gamma + \varphi_2 F^{-1}(\tau)) + \phi_1(v) + \phi_2(v)F^{-1}(\tau)$$

and

$$Q_{X|Z}(\alpha|z) = \mu + z'\pi.$$

so that the control function approach developed in previous sections can be applied to estimate structural parameters  $\beta(\tau) \equiv \beta + \varphi_1 F^{-1}(\tau)$  and  $\gamma(\tau) \equiv \gamma + \varphi_2 F^{-1}(\tau)$ . The identification strategy developed by Chesher (2003) can also be applied. To do so, rewrite

$$\begin{aligned} Q_{Y|X, Z}(\tau|x, z) &= x(\beta + \varphi_1 F^{-1}(\tau)) + z'_1(\gamma + \varphi_2 F^{-1}(\tau)) \\ &\quad + \phi_1(x - \mu - z'\pi) + \phi_2(x - \mu - z'\pi)F^{-1}(\tau) \end{aligned}$$

and

$$Q_{X|Z}(\alpha|z) = \mu + z'\pi.$$

Let  $\mathcal{Z}_2$  denote a set of components of  $z$  such that if a scalar component  $z_2 \in \mathcal{Z}_2$ , then  $z_2$  is not in  $z_1$  and  $\partial Q_{X|Z}(\alpha|z)/\partial z_2 \neq 0$ . By applying the quantile identification formula of Chesher (2003),

$$(13) \quad \beta(\tau) = \frac{\partial}{\partial x} Q_{Y|X, Z}(\tau|x, z) + \frac{\partial Q_{Y|X, Z}(\tau|x, z)/\partial z_2}{\partial Q_{X|Z}(\alpha|z)/\partial z_2}$$

for any  $(z, x)$  and  $z_2 \in \mathcal{Z}_2$ . Therefore,  $\beta(\tau)$  is over-identified. To develop a sample analog estimator, write

$$(14) \quad \beta(\tau) = \int w_{x,z}(x, z) \sum_{z_{2j} \in \mathcal{Z}_2} w_j \left[ \frac{\partial}{\partial x} Q_{Y|X, Z}(\tau|x, z) + \frac{\partial Q_{Y|X, Z}(\tau|x, z)/\partial z_{2j}}{\partial Q_{X|Z}(\alpha|z)/\partial z_{2j}} \right] dx dz,$$

where  $w_{x,z}(x, z)$  is a weight function that integrates to one and  $w_j$ 's are weights that sum to one. An estimator of  $\beta(\tau)$  can be proposed by replacing  $Q_{Y|X, Z}(\tau|x, z, x)$  with a non-parametric quantile regression estimator and  $Q_{X|Z}(\alpha|z)$  with a linear quantile regression estimator. Under suitable regularity conditions, the resulting estimator is expected to converge in probability to  $\beta(\tau)$  at a rate of  $n^{-1/2}$ . Similarly, one can identify and estimate  $\gamma(\tau)$ . The two-step estimator in Section 2 may be viewed as a convenient alternative to the weighted-average-derivative-type estimator of Chesher (2003).

### Case 3: A heteroskedastic linear median regression model

Consider a simultaneous equations model used in the Monte Carlo experiments of Hong and Tamer (2003):

$$\begin{aligned} Y &= X\beta + Z_1'\gamma + U, \quad U = \phi(Z_1)\tilde{U}, \\ X &= \mu + Z_1'\pi + V, \end{aligned}$$

where  $\phi(z_1)$  is a function of  $z_1$ ,  $\tilde{U}$  and  $V$  are independent of  $Z$  and are drawn from a standard bivariate normal distribution with covariance  $\rho$ . Then (a)  $Q_{U|Z}(1/2|z) = 0$ , (b)  $Q_{\eta|Z}(1/2|z) = 0$ , and (c)  $Q_{U|V,Z}(1/2|v, z) = \phi(z_1)Q_{\tilde{U}|V}(1/2|v) = \phi(z_1)\rho v$  since the conditional distribution of  $\tilde{U}$  given  $V = v$  is normal with mean  $\rho v$  and variance  $1 - \rho^2$ . Thus, both the IV and fitted value approaches can be used to estimate  $\beta$  and  $\gamma$ . As was already mentioned, the IV approach does not require the specification of the second equation. Since  $Q_{U|V,Z}(1/2|v, z)$  depends on  $z_1$ , the two-step estimator in Section 2 may be inconsistent. However, there is a straightforward extension to handle this case. In place of (2), now assume that

$$(15) \quad Q_{U|V,Z}(1/2|v, z) = \phi(z_1)\lambda_{1/2}(v)$$

almost surely, where  $\phi(z_1)$  is a real-valued, unknown function of  $Z_1$ . It can be shown that up to certain location normalization, the product of  $\phi(z_1)$  and  $\lambda_{1/2}(v)$  is identified nonparametrically, thereby implying that  $\beta(1/2)$  and  $\gamma(1/2)$  are identified. Then one can carry out the second step with a partially linear regression with multiplicative nonparametric components  $\phi(z_1)$  and  $\lambda_\tau(v)$ . It is interesting to note that the relationship (14) holds exactly here when  $\tau = 1/2$ . Therefore, one may use the identical weighted-average-derivative-type estimator to estimate  $\beta(1/2)$ . Estimation of  $\gamma(1/2)$  can be also carried out via the quantile identification formula of Chesher (2003).

### Case 4: A random coefficients model

The simple model for log earnings in (4) allows for a random intercept  $U$  that may be correlated with schooling residual  $V$ . However, it does not allow for ‘random slopes’. Modern labor economics of returns to schooling emphasizes the importance of heterogeneity in individual returns. See, for instance, Wooldridge (1997), Heckman and Vytlačil (1998), Card (2001), and references therein. Consider the following ‘random coefficients’ model that is

used as an example in Chesher (2003):

$$(16) \quad Y = S\beta(V) + U, \quad Q_{U|V,Z}(\tau|v, z) = \lambda(v),$$

$$(17) \quad S = Z'\pi + V, \quad Q_{V|Z}(\alpha|z) = 0,$$

where  $\beta(v)$  and  $\lambda(v)$  are unknown functions of individual ability  $V$ ,  $Y$  denotes the log earnings,  $S$  schooling, and  $Z$  a set of instruments. Chesher (2003) provided an elegant identification result on his parameter of interest, that is the returns to schooling  $\beta[Q_{V|Z}(\alpha_0|z_0)]$  for an individual with  $V = Q_{V|Z}(\alpha_0|z_0)$  for given  $\alpha_0$  and  $z_0$ . As an alternative, one can estimate  $\beta(v)$  globally by extending the control function approach along the lines of Newey, Powell, and Vella (1999) and Imbens and Newey (2003). To do so, replace (5) with the following restriction

$$(18) \quad Q_{Y|S,Z}(\tau|s, z) = Q_{Y|S,V,Z}(\tau|s, v, z) = s\beta(v) + \lambda(v)$$

almost surely. This suggests that the second step is now carried out by a nonparametric separable quantile regression with series approximations of  $\beta(v)$  and  $\lambda(v)$  while  $V_i$  being estimated in the first step. Finally, it is unclear whether the IV method based on (12) or the fitted value approach can be modified to estimate a random coefficients model such as (16).

We end this section by mentioning that the control function approach can be extended to deal with nonlinear specifications of regression functions (1) in a straightforward way. For example, regression functions  $X\beta(\tau) + Z_1'\gamma(\tau)$  and  $\mu(\alpha) + Z'\pi(\alpha)$  in (1) can be replaced with a nonparametric, additive, or single-index specification.

## 5 Monte Carlo Experiments

This section reports the results of a small set of Monte Carlo experiments to investigate the finite sample performance of the two-step estimator. In all experiments  $\tau = 0.9$  and  $\alpha = 0.5$ . For each  $n \in \{100, 400, 900, 1600\}$ , we considered the following model:

$$Y_i = X_i\beta + Z_{1i}\gamma + U_i, \quad U_i = V_i + \phi(V_i) + 0.5[\tilde{U}_i - F_{\tilde{U}}^{-1}(\tau)],$$

$$X_i = \mu + Z_{1i}\pi_1 + Z_{2i}\pi_2 + V_i, \quad V_i = \exp(Z_{2i}/2)\tilde{V}_i, \quad i = 1, \dots, n,$$

where  $Z_{1i}$ ,  $Z_{2i}$ ,  $\tilde{V}_i$ , and  $\tilde{U}_i$  are independently drawn from the standard normal distribution,  $\phi(v) = 4 \exp[-(v-1)^2]$ , and  $F_{\tilde{U}}$  is the cumulative distribution function of  $\tilde{U}$ . The function

$\phi(v)$  has a bell-shaped hump around one and represents a nonlinear component of  $\lambda_\tau(v) = v + \phi(v)$ . We set the parameter values  $(\beta, \gamma, \mu, \pi_1, \pi_2) = (1, 1, 1, 3, 1)$ .

The first step was carried out by a linear median regression of  $X$  on  $(1, Z_1, Z_2)$ . The second step requires the choice of basis functions and the number of approximating functions  $\kappa$ . In the experiments, we considered polynomial approximations from the first order polynomial to the eighth order polynomial. Asymptotic theory in Section 3 provides only qualitative restrictions on  $\kappa$  in terms of asymptotic rates, so it is the main purpose of the experiments to check the sensitivity of the two-step estimator to the choice of  $\kappa$ . The trimming function was set to be  $t(\hat{W}_i) = 1(|X_i| \leq 10)1(|Z_{1i}| \leq 3)1(|\hat{V}_i| \leq 5)$ . We compared the two-step estimator with a ( $\tau$ -th) linear quantile regression estimator (ignoring endogeneity) and a fitted-value quantile regression estimator for which  $X_i$  is replaced with  $\hat{\mu} + Z_{1i}\hat{\pi}_1 + Z_{2i}\hat{\pi}_2$ . There were 1,000 replications in each experiment. The computations were carried out in GAUSS with GAUSS pseudo-random number generators.

Tables 1 and 2 show results of the experiments for  $\beta$  and  $\gamma$ , respectively. Both the linear quantile regression (QR) estimator and fitted value (FV) estimator have large biases for all sample sizes. This is expected since they are inconsistent. The two-step estimator with a first order polynomial (TS(1)) has nonnegligible biases that result from the misspecification of  $\lambda_\tau$ . The two-step estimators with flexible polynomial approximations perform quite well. The biases are rather negligible compared to the size of standard deviations and the root mean square errors (RMSE) shrink to zero roughly at a rate of  $n^{-1/2}$ . Furthermore, it can be seen that the estimator is not very sensitive to the choice of the order of polynomial approximations.

## 6 Empirical Examples

In this section, the estimation method is illustrated by applying it to two empirical applications, demand for fish and the returns to schooling.

### 6.1 Demand for Whiting at the Fulton Fish Market

This section presents estimation results for the first empirical application. This application consists of using the data on demand for whiting at the Fulton fish market to estimate price elasticities of demand for fish. The data were used previously in Graddy (1995),

Angrist, Graddy, and Imbens (2000), and Chernozhukov and Hansen (2001). In particular, Chernozhukov and Hansen (2001) estimated quantile treatment effects (price elasticities at different quantiles of demand level) using wind speed as the instrument variable (without covariates).

The dependent variable  $Y$  is the logarithm of the total quantity sold by a single dealer on each day and the endogenous explanatory variable  $X$  is the logarithm of the average daily price. The exogenous explanatory variables  $Z_1$  are indicators for days of the week (Monday, Tuesday, Wednesday, and Thursday) and for weather conditions on shore (Rainy on shore and Cold on shore). The instrument variables are indicators for weather conditions at sea (Stormy and Mixed). The exogenous explanatory variables and instruments were those used by Angrist, Graddy, and Imbens (2000). See Angrist, Graddy, and Imbens (2000) for details about the data and variables. The sample size is 111.

The first step was carried out by a linear median regression of  $X$  on a constant term,  $Z_1$  and instruments. In the second step, a third order polynomial approximation was used to estimate price elasticities at different values of  $\tau$ . There was no trimming of the data (that is,  $t(\hat{W}_i) = 1$  for all  $i$ ). The standard errors of the estimates were calculated by (11) with the standard normal density as the kernel function and  $h_n = \hat{\sigma}_\varepsilon^2 n^{-3/20}$  as a bandwidth, where  $\hat{\sigma}_\varepsilon^2$  is the empirical standard deviation of  $\hat{\varepsilon}_\tau$ . The estimation results were not very sensitive to the choices of the order of polynomials and  $h_n$ . To compare estimation results with those of Chernozhukov and Hansen (2001), we also estimated the model using the wind speed as the instrument without  $Z_1$ .

The estimation results are summarized in Figure 1. The price elasticities are shown in the absolute value. The left-hand panel of the figure shows price elasticities without the adjustment for endogeneity (dotted lines) and those with the adjustment for endogeneity (solid lines) for the model with  $Z_1$  using binary instruments (Stormy and Mixed). The price elasticities without the adjustment for endogeneity were estimated by linear quantile regressions. To show the accuracy of the adjusted estimates, 90% pointwise confidence intervals (dashed lines) of adjusted elasticities are superimposed in the figure. The right-hand panel shows the price elasticities for the model without  $Z_1$  using the wind speed as the instrument.

In both panels, adjusted elasticities are quite different from unadjusted elasticities, especially at lower quantiles. This is consistent with previous findings of Chernozhukov and Hansen (2001). In fact, results in the right-hand panel are quite comparable to those in



Figure 1 of Chernozhukov and Hansen (2001). However, adjusted elasticities increase at higher quantiles in the left-hand panel, whereas those decrease in the right-hand panel. This difference may yield quite contradictory interpretations, but in view of rather small sample size, more careful analysis is needed to decide whether or not it is just an artifact of random sampling error.

## 6.2 Returns to Schooling Using Quarter of Birth as Instrument

This section presents estimation results for the returns to schooling. This empirical example consists of using the data of Angrist and Krueger (1991) to estimate returns to schooling at different quantiles. Angrist and Krueger (1991) estimated effects of compulsory schooling on earnings using quarter of birth as an instrument for schooling. We used a sample of 329,509 men born 1930-1939 from the 1980 census. This data set was used previously by Angrist, Imbens, and Krueger (1999) and is available at the Journal of Applied Econometrics web site.

The dependent variable  $Y$  is the log weekly wage and the endogenous explanatory variable  $X$  is years of schooling. The exogenous explanatory variables  $Z_1$  are 10 indicator variables of year of birth. The instrument variables are 30 indicator variables of quarter of birth interacted with year of birth. This simple version of the Angrist-Krueger model is used here in an attempt to mitigate the ‘weak instruments’ bias. See, for example, Bound, Jaeger, and Baker (1995) and Staiger and Stock (1997) for the problem of weak instruments.

It is quite plausible that quarter and year of birth is independent of taste and ability factors ( $Z$  is independent of  $V$  using the notation in previous sections). In that case, schooling residual  $V$  can be estimated (up to location) by any quantile regression or even by mean regression as well since years of schooling is bounded. Although schooling is conceptually continuous, years of schooling has only finite number of distinct values. To avoid this problem, the first step was carried out by a linear mean regression of  $X$  on  $Z_1$  and instruments. In the second step, a fifth order polynomial approximation was used to estimate returns to schooling at different values of  $\tau$ . There was no trimming of the data. The standard error of the estimate was calculated similarly as in Section 6.1. Qualitative estimation results were not very sensitive to the choice of the order of polynomials.

The estimation results are summarized in Figure 2. Following Buchinsky (1994), the return to schooling at each  $\tau$  is defined as the derivative of the primary equation with respect to schooling, that is  $\beta(\tau)$ . If there is endogeneity, then returns to schooling obtained by

standard quantile regressions may be misleading. As in Figure 1, Figure 2 shows returns to schooling without the adjustment for endogeneity (dotted lines) and those with the adjustment for endogeneity (solid lines) along with 90% pointwise confidence intervals (dashed lines) of adjusted estimates.

It can be seen that unadjusted returns to schooling are between 0.7 and 0.8 and roughly constant over the range of quantiles. In contrast, adjusted returns to schooling are quite different from unadjusted returns at lowest quantiles. As was discussed in Angrist and Krueger (1991), the adjusted returns to schooling in this example may be interpreted as the structural effects of compulsory school attendance. Under this interpretation, our estimation results may suggest that effects of compulsory schooling are much stronger at lowest quantiles of earnings. The mean return to schooling estimated by the conventional two-stage-least-squares (2SLS) was 0.089 with standard error of 0.016, and the OLS estimate was 0.071 with standard error of 0.0003. Thus, as far as the mean effect is concerned, the bias due to endogeneity is small. However, it seems that the endogeneity bias is nonnegligible at the bottom of the distribution of earnings ( $\tau = 0.05$ ). Similar results have been presented previously by Andrew Chesher on October 2, 2003 at a lecture to inaugurate the academic year of the International Doctorate in Economic Analysis (IDEA) at Univeristad Autònoma Barcelona under the title ‘Identification of the distribution of policy impacts’. Using the same data set and a different method, Honoré and Hu (2003) also found that quantile effects are larger at the lower end of the distribution of earnings.

## 7 Conclusions

This paper has presented the method for estimating quantile structural effects based on the control function approach. The paper has also provided empirical examples for which the new method has revealed some important features of endogeneity that could not easily be detected using standard methods like 2SLS.

The success of a control function approach depends crucially on the plausibility of assumptions about the stochastic relationship between the unobserved components  $U$  and  $V$  and observed variables  $X$  and  $Z$ . Therefore, it would be useful to extend the basic model (1)-(3) to more complex situations such as some examples discussed in Section 4. It would be also useful to consider nonparametric estimation of the primary and reduced-form equations like Newey, Powell, and Vella (1999). These are topics for future research.

Another interesting topic is estimation of an intercept term of the primary equation. For applications considered in the paper, it is unnecessary to know the intercept term. However, in some cases such as estimating within-group wage inequality (Buchinsky (1994, 1998a, 1998b)), it is important to know the intercept term. Obviously, estimation of the intercept term requires an additional restriction. For example, one may use a condition like  $Q_U(\tau) = 0$  to develop an estimator of the intercept term. To do so, rewrite the primary equation in (1) as

$$Y = X\beta(\tau) + Z_1'\gamma(\tau) + \psi(\tau) + U,$$

where  $\psi(\tau)$  is the intercept term. Under the assumption that  $Q_U(\tau) = 0$ ,  $\psi(\tau)$  can be estimated by a  $\tau$ -th sample quantile of  $Y_i - X_i\hat{\beta}(\tau) - Z_{1i}'\hat{\gamma}(\tau)$ . Under some regularity conditions, it would be straightforward to establish  $n^{-1/2}$ -consistency and asymptotic normality of this estimator. However, it is beyond the scope of this paper to provide full details.

## A Appendix: Proofs

Throughout the Appendix, let  $C$  denote a generic positive constant that may be different in different uses. Let  $\lambda_{\min}(A)$  and  $\lambda_{\max}(A)$  denote minimum and maximum eigenvalues of a symmetric matrix  $A$ . For notational simplicity, we will suppress dependence on  $\tau$  and  $\alpha$ . As shorthand notation, let  $P_{\kappa i} = P_{\kappa}(W_i)$ ,  $\hat{P}_{\kappa i} = P_{\kappa}(\hat{W}_i)$ ,  $t_i = t(W_i)$ ,  $\hat{t}_i = t(\hat{W}_i)$ ,  $m_i = X_i\beta + Z'_{1i}\gamma + \lambda(V_i)$ ,  $f_i = f_{\varepsilon}(0|X_i, Z_i)$ , and  $b_{\kappa i} = P'_{\kappa i}\theta_{\kappa 0} - m_i$ . Define

$$\begin{aligned}\tilde{\Phi}_{n\kappa} &= n^{-1} \sum_{i=1}^n \hat{t}_i f_i \hat{P}_{\kappa i} \hat{P}'_{\kappa i} \quad \text{and} \\ \Phi_{n\kappa} &= n^{-1} \sum_{i=1}^n t_i f_i P_{\kappa i} P'_{\kappa i}.\end{aligned}$$

**Lemma A.1.** *As  $n \rightarrow \infty$ ,*

$$\begin{aligned}(a) \quad & \max_{1 \leq i \leq n} t_i \left\| \hat{P}_{\kappa i} - P_{\kappa i} \right\| = O_p \left[ \zeta_1(\kappa)/n^{1/2} \right] = o_p(1). \\ (b) \quad & \|\Phi_{n\kappa} - \Phi_{\kappa}\| = O_p \left[ \zeta_0(\kappa)^2 \kappa/n \right] = o_p(1). \\ (c) \quad & \left\| \tilde{\Phi}_{n\kappa} - \Phi_{n\kappa} \right\| = O_p \left[ \kappa^{1/2} \zeta_1(\kappa)/n^{1/2} \right] = o_p(1).\end{aligned}$$

*Proof.* To prove part (a), notice that

$$\begin{aligned}\max_{1 \leq i \leq n} t_i \left\| \hat{P}_{\kappa i} - P_{\kappa i} \right\|^2 &= \max_{1 \leq i \leq n} t_i \sum_{k=1}^{\kappa} [p_k(\hat{V}_i) - p_k(V_i)]^2 \\ &= \max_{1 \leq i \leq n} t_i \sum_{k=1}^{\kappa} \left[ \frac{\partial p_k(\tilde{V}_i)}{\partial v} (\hat{V}_i - V_i) \right]^2 \\ &= \max_{1 \leq i \leq n} t_i (\hat{V}_i - V_i)^2 \sum_{k=1}^{\kappa} \left[ \frac{\partial p_k(\tilde{V}_i)}{\partial v} \right]^2 \\ &= O_p \left[ \zeta_1(\kappa)^2/n \right],\end{aligned}$$

where  $\tilde{V}_i$  is between  $\hat{V}_i$  and  $V_i$ . Part (b) can be proved as in the proof of Theorem 1 of Newey (1997).

Now consider part (c). Notice that using arguments similar to those used in the proof of Lemma A3 of Newey, Powell, and Vella (1999),

$$(19) \quad n^{-1} \sum_{i=1}^n t_i |\hat{t}_i - t_i| = O_p \left( \max_i t_i |\hat{V}_i - V_i| \right) = O_p(n^{-1/2}),$$

where the last equality follows from (10). Now, as in (A.5) of Newey, Powell, and Vella (1999),

$$\begin{aligned}
\left\| \tilde{\Phi}_{n\kappa} - \Phi_{n\kappa} \right\| &\leq \left\| n^{-1} \sum_{i=1}^n t_i \hat{t}_i f_i [\hat{P}_{\kappa i} \hat{P}'_{\kappa i} - P_{\kappa i} P'_{\kappa i}] \right\| + \left\| n^{-1} \sum_{i=1}^n (t_i \hat{t}_i - t_i) f_i \hat{P}_{\kappa i} \hat{P}'_{\kappa i} \right\| \\
&\quad + \left\| n^{-1} \sum_{i=1}^n (t_i \hat{t}_i - t_i) f_i P_{\kappa i} P'_{\kappa i} \right\| \\
&\leq C n^{-1} \sum_{i=1}^n t_i \hat{t}_i \left( \left\| \hat{P}_{\kappa i} - P_{\kappa i} \right\|^2 + 2 \left\| \hat{P}_{\kappa i} - P_{\kappa i} \right\| \left\| P_{\kappa i} \right\| \right) + C \zeta_0(\kappa)^2 n^{-1} \sum_{i=1}^n t_i |\hat{t}_i - t_i| \\
&\leq C n^{-1} \sum_{i=1}^n t_i \left\| \hat{P}_{\kappa i} - P_{\kappa i} \right\|^2 \\
&\quad + C \left( n^{-1} \sum_{i=1}^n t_i \left\| P_{\kappa i} \right\|^2 \right)^{1/2} \left( n^{-1} \sum_{i=1}^n t_i \left\| \hat{P}_{\kappa i} - P_{\kappa i} \right\|^2 \right)^{1/2} \\
&\quad + C \zeta_0(\kappa)^2 n^{-1} \sum_{i=1}^n t_i |\hat{t}_i - t_i| \\
&= O_p \left[ \zeta_1(\kappa)^2 / n \right] + O_p \left[ \kappa^{1/2} \zeta_1(\kappa) / n^{1/2} \right] + O_p \left[ \zeta_0(\kappa)^2 / n^{1/2} \right].
\end{aligned}$$

Part (c) now follows from the fact that  $\zeta_1(\kappa)/n^{1/2} \rightarrow 0$  and  $\zeta_0(\kappa) \geq C\kappa^{1/2}$ .  $\square$

**Lemma A.2.** *As  $n \rightarrow \infty$ ,*

$$\max_{1 \leq i \leq n} t_i \left| (\hat{P}_{\kappa i} - P_{\kappa i})' \theta_{\kappa 0} - \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) \right| = o_p \left( n^{-1/2} \right).$$

*Proof.* Let  $\theta_{\kappa 0}^{(j)}$  denote the  $j$ -th component of  $\theta_{\kappa 0}$ . A Taylor series expansion gives

$$\begin{aligned}
t_i (\hat{P}_{\kappa i} - P_{\kappa i})' \theta_{\kappa 0} &= t_i \sum_{k=1}^{\kappa} \left[ p_k(\hat{V}_i) - p_k(V_i) \right] \theta_{\kappa 0}^{(1+d_{z_1+k})} \\
&= t_i \sum_{k=1}^{\kappa} \frac{\partial p_k(\tilde{V}_i)}{\partial v} \theta_{\kappa 0}^{(1+d_{z_1+k})} (\hat{V}_i - V_i) \\
&= t_i \left[ \sum_{k=1}^{\kappa} \frac{\partial p_k(\tilde{V}_i)}{\partial v} \theta_{\kappa 0}^{(1+d_{z_1+k})} - \frac{d\lambda(\tilde{V}_i)}{dv} \right] (\hat{V}_i - V_i) \\
&\quad + t_i \left[ \frac{d\lambda(\tilde{V}_i)}{dv} - \frac{d\lambda(V_i)}{dv} \right] (\hat{V}_i - V_i) + t_i \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i),
\end{aligned}$$

where  $\tilde{V}_i$  is between  $\hat{V}_i$  and  $V_i$ . The lemma now follows from the fact that  $\max_i t_i |\hat{V}_i - V_i| = O_p(n^{-1/2})$ ,  $\sup_{w \in \mathcal{W}} |\partial \lambda(w) / \partial v - [\partial P_\kappa(w) / \partial v]' \theta_{\kappa 0}| = O(\kappa^{-(r-1)})$ , and  $d\lambda(v) / dv$  is continuously differentiable on  $\mathcal{W}$ .  $\square$

Let  $F(\cdot | x, z)$  denote the cumulative distribution function of  $Y$  conditional on  $X = x$  and  $Z = z$ . Define

$$\begin{aligned}\hat{G}_{n\kappa}(\theta) &= n^{-1} \tilde{\Phi}_{n\kappa}^{-1} \sum_{i=1}^n \hat{t}_i \left\{ \tau - 1 [Y_i \leq \hat{P}'_{\kappa i}(\theta - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0}] \right\} \hat{P}_{\kappa i}, \\ G_{n\kappa}(\theta) &= n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1}^n t_i \left\{ \tau - 1 [Y_i \leq P'_{\kappa i}(\theta - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0}] \right\} P_{\kappa i}, \\ \hat{G}_{n\kappa}^*(\theta) &= n^{-1} \tilde{\Phi}_{n\kappa}^{-1} \sum_{i=1}^n \hat{t}_i \left\{ \tau - F[\hat{P}'_{\kappa i}(\theta - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0} | X_i, Z_i] \right\} \hat{P}_{\kappa i}, \\ G_{n\kappa}^*(\theta) &= n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1}^n t_i \left\{ \tau - F[P'_{\kappa i}(\theta - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} | X_i, Z_i] \right\} P_{\kappa i}, \\ \hat{H}_{n\kappa}(\theta) &= \hat{G}_{n\kappa}(\theta) - \hat{G}_{n\kappa}^*(\theta),\end{aligned}$$

and  $H_{n\kappa}(\theta) = G_{n\kappa}(\theta) - G_{n\kappa}^*(\theta)$ . Let  $1_n$  be the indicator function such that  $1_n = 1\{\lambda_{\min}(\tilde{\Phi}_{n\kappa}) \geq \lambda_{\min}(\Phi_\kappa)/2 \text{ and } \lambda_{\min}(\Phi_{n\kappa}) \geq \lambda_{\min}(\Phi_\kappa)/2\}$ . By Lemma A.1 (b) and (c),  $\|\tilde{\Phi}_{n\kappa} - \Phi_\kappa\| = o_p(1)$  and  $\|\Phi_{n\kappa} - \Phi_\kappa\| = o_p(1)$ . Thus,  $\Pr(1_n = 1) \rightarrow 1$  as  $n \rightarrow \infty$ .

**Lemma A.3.** *As  $n \rightarrow \infty$ ,*

$$1_n \left\| \hat{G}_{n\kappa}(\hat{\theta}_{n\kappa}) \right\| = O_p[\zeta_0(\kappa)\kappa/n] = o_p(n^{-1/2}).$$

*Proof.* To prove the lemma, it is useful to introduce some additional notation that is used in Koenker and Bassett (1978) and Chaudhuri (1991). Let  $\mathcal{N} = \{1, \dots, n\}$  and  $\mathcal{H}_\kappa$  denote the collection of all  $d(\kappa)$ -element subsets of  $\mathcal{N}$ . Also, let  $B(h)$  denote the submatrix (subvector) of a matrix (vector)  $B$  with rows (components) that are indexed by the elements of  $h \in \mathcal{H}_\kappa$ . In particular, let  $\hat{\mathbf{P}}_\kappa(h)$  denote the  $d(\kappa) \times d(\kappa)$  matrix, whose rows are the vectors  $\hat{P}'_{\kappa i}$  such that  $i \in h$ , and let  $\mathbf{Y}_\kappa(h)$  denote the  $d(\kappa) \times 1$  vector, whose elements are  $Y_i$  such that  $i \in h$ . In addition, let  $\hat{\mathbf{P}}_\kappa$  denote the  $n \times d(\kappa)$  matrix, whose rows are the vectors  $\hat{P}'_{\kappa i}$  for  $i = 1, \dots, n$ . The matrix  $\hat{\mathbf{P}}_\kappa$  has rank  $= d(\kappa)$  almost surely for all sufficiently large  $n$ .

By Theorem 3.1 of Koenker and Bassett (1978), there exists an index set  $h_\kappa \in \mathcal{H}_\kappa$  such that the problem (8) has at least one solution of the form  $\hat{\theta}_{n\kappa} = \hat{\mathbf{P}}_\kappa(h_\kappa)^{-1} \mathbf{Y}_\kappa(h_\kappa)$ . Now

write  $1_n \hat{G}_{n\kappa}(\hat{\theta}_{n\kappa}) = 1_n \hat{G}_{n\kappa 1}(\hat{\theta}_{n\kappa}) + 1_n \hat{G}_{n\kappa 2}(\hat{\theta}_{n\kappa})$ , where

$$G_{n\kappa 1}(\hat{\theta}_{n\kappa}) = n^{-1} \tilde{\Phi}_{n\kappa}^{-1} \sum_{i=1, i \in h_\kappa}^n \hat{t}_i \left\{ \tau - 1 [Y_i \leq \hat{P}'_{\kappa i} \hat{\theta}_{n\kappa}] \right\} \hat{P}_{\kappa i},$$

and

$$G_{n\kappa 2}(\hat{\theta}_{n\kappa}) = n^{-1} \tilde{\Phi}_{n\kappa}^{-1} \sum_{i=1, i \in h_\kappa^c}^n \hat{t}_i \left\{ \tau - 1 [Y_i \leq \hat{P}'_{\kappa i} \hat{\theta}_{n\kappa}] \right\} \hat{P}_{\kappa i}.$$

Notice that  $\max_{1 \leq i \leq n} 1_n \hat{t}_i \left\| \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} \right\| = O_p[\zeta_0(\kappa)]$  by Lemma A.1 and the fact that the smallest eigenvalue of  $\tilde{\Phi}_{n\kappa}$  is bounded away from zero (when  $1_n = 1$ ). Thus, we have  $1_n \left\| \hat{G}_{n\kappa 1}(\hat{\theta}_{n\kappa}) \right\| = O_p[\zeta_0(\kappa)d(\kappa)/n]$ .

Now consider  $\hat{G}_{n\kappa 2}(\hat{\theta}_{n\kappa})$ . Define

$$\bar{G}_{n\kappa 2} = n G_{n\kappa 2}(\hat{\theta}_{n\kappa})' \tilde{\Phi}_{n\kappa} \hat{\mathbf{P}}_\kappa(h_\kappa)^{-1}.$$

By Theorem 3.3 of Koenker and Bassett (1978), each component in  $\bar{G}_{n\kappa 2}$  is between  $\tau - 1$  and  $\tau$ . Thus,  $\left\| \bar{G}_{n\kappa 2}(\hat{\theta}_{n\kappa}) \right\| \leq d(\kappa)^{1/2}$ . Since the smallest eigenvalue of  $\Phi_{n\kappa}$  is bounded away from zero (when  $1_n = 1$ ), we can find a constant  $C < \infty$  (independent of  $\kappa$ ) such that

$$1_n \left\| \hat{\mathbf{P}}_\kappa(h_\kappa) \tilde{\Phi}_{n\kappa}^{-1} \right\| \leq C \left\| \hat{\mathbf{P}}_\kappa(h_\kappa) \right\|.$$

Also notice that

$$\begin{aligned} \left\| \hat{\mathbf{P}}_\kappa(h_\kappa) \right\|^2 &= \text{trace}[\hat{\mathbf{P}}_\kappa(h_\kappa)' \hat{\mathbf{P}}_\kappa(h_\kappa)] = \text{trace}[\hat{\mathbf{P}}_\kappa(h_\kappa) \hat{\mathbf{P}}_\kappa(h_\kappa)'] \\ &= \sum_{i \in h_\kappa} \left\| \hat{P}_{\kappa i} \right\|^2 \leq C \zeta_0(\kappa)^2 d(\kappa). \end{aligned}$$

Hence,  $\left\| \hat{\mathbf{P}}_\kappa(h_\kappa) \right\| \leq C \zeta_0(\kappa) d(\kappa)^{1/2}$ . Therefore,

$$1_n \left\| G_{n\kappa 2}(\hat{\theta}_{n\kappa}) \right\| \leq n^{-1} \left\| \bar{G}_{n\kappa 2}(\hat{\theta}_{n\kappa}) \right\| 1_n \left\| \hat{\mathbf{P}}_\kappa(h_\kappa) \tilde{\Phi}_{n\kappa}^{-1} \right\| \leq C \zeta_0(\kappa) d(\kappa)/n.$$

Since arguments used in this proof hold uniformly over  $h_\kappa$ , the lemma follows immediately.  $\square$

**Lemma A.4.** *As  $n \rightarrow \infty$ ,*

$$(a) \quad 1_n \left\| H_{n\kappa}(\theta_{\kappa 0}) \right\| = O_p \left[ (\kappa/n)^{1/2} \right].$$

$$(b) \ 1_n \|AH_{n\kappa}(\theta_{\kappa 0})\| = O_p(n^{-1/2}).$$

*Proof.* First, we will prove part (b). Notice that since the data are i.i.d.,  $f_\varepsilon(\cdot|x, z)$  is bounded away from zero in a neighborhood of zero for all  $x$  and  $z$ ,  $t_i^2 = t_i$ , and the smallest eigenvalue of  $\Phi_{n\kappa}$  is bounded away from zero (when  $1_n = 1$ ),

$$\begin{aligned} & E \left[ 1_n \|H_{n\kappa}(\theta_{\kappa 0})\|^2 \mid X_1, \dots, X_n, Z_1, \dots, Z_n \right] \\ & \leq 1_n n^{-2} \sum_{i=1}^n t_i \left[ E \left\{ \left[ F[P'_{\kappa i} \theta_{\kappa 0} \mid X_i, Z_i] - 1[Y_i \leq P'_{\kappa i} \theta_{\kappa 0}] \right]^2 \mid X_i, Z_i \right\} P'_{\kappa i} \Phi_{n\kappa}^{-1} A' A \Phi_{n\kappa}^{-1} P_{\kappa i} \right] \\ & \leq C 1_n n^{-2} \sum_{i=1}^n \text{trace} \left[ t_i P'_{\kappa i} \Phi_{n\kappa}^{-1} A' A \Phi_{n\kappa}^{-1} P_{\kappa i} \right] \\ & \leq C 1_n n^{-2} \sum_{i=1}^n \left[ (\min_i f_i)^{-1} f_i \right] \text{trace} \left[ t_i A \Phi_{n\kappa}^{-1} P_{\kappa i} P'_{\kappa i} \Phi_{n\kappa}^{-1} A' \right] \\ & \leq C 1_n n^{-1} \text{trace} \left\{ A \Phi_{n\kappa}^{-1} \left[ n^{-1} \sum_{i=1}^n t_i f_i P_{\kappa i} P'_{\kappa i} \right] \Phi_{n\kappa}^{-1} A' \right\} \\ & = C 1_n n^{-1} \text{trace} \left( A \Phi_{n\kappa}^{-1} A' \right) \\ & \leq C n^{-1} (1 + d_{z1}). \end{aligned}$$

Therefore, part (b) of the lemma follows from Markov's inequality. Part (a) follows by repeating the same arguments with  $A$  replaced by an identity matrix.  $\square$

**Lemma A.5.** (a) As  $n \rightarrow \infty$ ,

$$1_n \hat{G}_{n\kappa}^*(\theta) = -1_n(\theta - \theta_{\kappa 0}) - 1_n n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1}^n t_i f_i \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) P_{\kappa i} + R_{n\kappa}^*(\theta),$$

$$\text{where } \|R_{n\kappa}^*(\theta)\| = o_p[\kappa^{1/2} n^{-1/2}] + O_p[\kappa^{-r} + \zeta_0(\kappa) \|\theta - \theta_{\kappa 0}\|^2 + \kappa^{1/2} n^{-1} + \kappa^{1/2} \kappa^{-2r}].$$

(b) As  $n \rightarrow \infty$ ,

$$1_n A \hat{G}_{n\kappa}^*(\theta) = -1_n A(\theta - \theta_{\kappa 0}) - 1_n n^{-1} A \Phi_{n\kappa}^{-1} \sum_{i=1}^n t_i f_i \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) P_{\kappa i} + AR_{n\kappa}^*(\theta),$$

$$\text{where } \|AR_{n\kappa}^*(\theta)\| = o_p[n^{-1/2}] + O_p[\kappa^{-r} + \zeta_0(\kappa) \|\theta - \theta_{\kappa 0}\|^2 + n^{-1} + \kappa^{-2r}].$$

*Proof.* First, we will prove part (a). Define

$$1_n \tilde{G}_{n\kappa}^*(\theta) = -1_n(\theta - \theta_{\kappa 0}) - 1_n n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1}^n t_i f_i \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) P_{\kappa i}$$



and

$$\begin{aligned} 1_n \bar{G}_{n\kappa}^*(\theta) &= -1_n(\theta - \theta_{\kappa 0}) - 1_n n^{-1} \tilde{\Phi}_{n\kappa}^{-1} \sum_{i=1}^n \hat{t}_i f_i (\hat{P}_{\kappa i} - P_{\kappa i})' \theta_{\kappa 0} \hat{P}_{\kappa i} \\ &\quad - 1_n n^{-1} \tilde{\Phi}_{n\kappa}^{-1} \sum_{i=1}^n \hat{t}_i f_i b_{\kappa i} \hat{P}_{\kappa i}. \end{aligned}$$

Write  $1_n \hat{G}_{n\kappa}^*(\theta) = 1_n \tilde{G}_{n\kappa}^*(\theta) + R_{n\kappa 1}^*(\theta) + R_{n\kappa 2}^*(\theta)$ , where

$$R_{n\kappa 1}^*(\theta) = 1_n \bar{G}_{n\kappa}^*(\theta) - 1_n \tilde{G}_{n\kappa}^*(\theta) \quad \text{and} \quad R_{n\kappa 2}^*(\theta) = 1_n \hat{G}_{n\kappa}^*(\theta) - 1_n \bar{G}_{n\kappa}^*(\theta).$$

Notice that

$$\begin{aligned} 1_n n^{-1} \sum_{i=1}^n \left\| \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} \right\|^2 &= 1_n n^{-1} \sum_{i=1}^n \text{trace} \left( \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} \hat{P}_{\kappa i}' \tilde{\Phi}_{n\kappa}^{-1} \right) \\ (20) \quad &\leq C 1_n \text{trace} \left( \tilde{\Phi}_{n\kappa}^{-1} \right) = O_p(\kappa). \end{aligned}$$

By Lemmas A.1 and A.2, equation (20), and the Cauchy-Schwartz inequality,

$$\begin{aligned} &\left\| 1_n n^{-1} \tilde{\Phi}_{n\kappa}^{-1} \sum_{i=1}^n \hat{t}_i f_i (\hat{P}_{\kappa i} - P_{\kappa i})' \theta_{\kappa 0} \hat{P}_{\kappa i} - 1_n n^{-1} \tilde{\Phi}_{n\kappa}^{-1} \sum_{i=1}^n \hat{t}_i f_i \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) \hat{P}_{\kappa i} \right\| \\ &\leq 1_n n^{-1} \sum_{i=1}^n \left\| \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} \right\| \left| \hat{t}_i f_i \left| (\hat{P}_{\kappa i} - P_{\kappa i})' \theta_{\kappa 0} - \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) \right| \right| \\ &\leq 1_n \left( n^{-1} \sum_{i=1}^n \left\| \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} \right\|^2 \right)^{1/2} \left( n^{-1} \sum_{i=1}^n \hat{t}_i f_i^2 \left| (\hat{P}_{\kappa i} - P_{\kappa i})' \theta_{\kappa 0} - \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) \right|^2 \right)^{1/2} \\ (21) \quad &= o_p \left( \kappa^{1/2} n^{-1/2} \right). \end{aligned}$$

Also, notice that

$$\begin{aligned} &1_n n^{-1} \sum_{i=1}^n \left\| \hat{t}_i \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} - t_i \Phi_{n\kappa}^{-1} P_{\kappa i} \right\|^2 \\ &\leq C 1_n n^{-1} \sum_{i=1}^n |\hat{t}_i - t_i| \left[ \left\| \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} \right\|^2 + \left\| \Phi_{n\kappa}^{-1} P_{\kappa i} \right\|^2 \right] \\ &\quad + C 1_n n^{-1} \sum_{i=1}^n \left\| \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} - \tilde{\Phi}_{n\kappa}^{-1} P_{\kappa i} \right\|^2 + C 1_n n^{-1} \sum_{i=1}^n \left\| \tilde{\Phi}_{n\kappa}^{-1} P_{\kappa i} - \Phi_{n\kappa}^{-1} P_{\kappa i} \right\|^2 \\ &\leq C 1_n \max_{1 \leq i \leq n} \left[ \left\| \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} \right\|^2 + \left\| \Phi_{n\kappa}^{-1} P_{\kappa i} \right\|^2 \right] n^{-1} \sum_{i=1}^n |\hat{t}_i - t_i| \\ &\quad + O_p(1) n^{-1} \sum_{i=1}^n \left\| \hat{P}_{\kappa i} - P_{\kappa i} \right\|^2 + O_p(1) \left\| \tilde{\Phi}_{n\kappa} - \Phi_{n\kappa} \right\|^2 n^{-1} \sum_{i=1}^n \left\| P_{\kappa i} \right\|^2. \end{aligned}$$

Thus, by Lemmas A.1 and A.2 and equation (19),

$$(22) \quad 1_n n^{-1} \sum_{i=1}^n \left\| \hat{t}_i \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} - t_i \Phi_{n\kappa}^{-1} P_{\kappa i} \right\|^2 = O_p \left[ \zeta_0(\kappa)^2/n^{1/2} + \kappa^2 \zeta_1(\kappa)^2/n \right] = o_p(1).$$

By (22) and the Cauchy-Schwartz inequality, we have

$$(23) \quad \begin{aligned} & \left\| 1_n n^{-1} \tilde{\Phi}_{n\kappa}^{-1} \sum_{i=1}^n \hat{t}_i f_i \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) \hat{P}_{\kappa i} - 1_n n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1}^n t_i f_i \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) P_{\kappa i} \right\| \\ & \leq 1_n n^{-1} \sum_{i=1}^n f_i \left| \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) \right| \left\| \hat{t}_i \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} - t_i \Phi_{n\kappa}^{-1} P_{\kappa i} \right\| \\ & \leq C 1_n \left( n^{-1} \sum_{i=1}^n \left| \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) \right|^2 \right)^{1/2} \left( n^{-1} \sum_{i=1}^n \left\| \hat{t}_i \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} - t_i \Phi_{n\kappa}^{-1} P_{\kappa i} \right\|^2 \right)^{1/2} \\ & \leq O_p(n^{-1/2}) o_p(1) = o_p(n^{-1/2}). \end{aligned}$$

In addition, it is not difficult to show that

$$(24) \quad 1_n \left\| n^{-1} \tilde{\Phi}_{n\kappa}^{-1} \sum_{i=1}^n \hat{t}_i f_i b_{\kappa i} \hat{P}_{\kappa i} \right\| = O_p \left[ \max_i b_{\kappa i} \right] = O_p(\kappa^{-r}).$$

Thus, by combining (21), (23), and (24),

$$(25) \quad \|R_{n\kappa 1}^*(\theta)\| = o_p \left[ \kappa^{1/2} n^{-1/2} \right] + O_p(\kappa^{-r}).$$

Now, using equation (20), a first-order Taylor series expansion and the assumption that  $F(m_i|X_i, Z_i) = \tau$  for each  $i$ , we have

$$(26) \quad \begin{aligned} \|R_{n\kappa 2}^*(\theta)\| & \leq C 1_n \left[ n^{-1} \sum_{i=1}^n \hat{t}_i \left\| \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} \right\| \left\{ \hat{P}'_{\kappa i}(\theta - \theta_{\kappa 0}) + (\hat{P}_{\kappa i} - P_{\kappa i})' \theta_{\kappa 0} + b_{\kappa i} \right\}^2 \right] \\ & \leq C 1_n \max_{1 \leq i \leq n} \left\| \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} \right\| (\theta - \theta_{\kappa 0})' \left\{ \sum_{i=1}^n \hat{t}_i \hat{P}_{\kappa i} \hat{P}'_{\kappa i} \right\} (\theta - \theta_{\kappa 0}) \\ & \quad + C 1_n \left\{ n^{-1} \sum_{i=1}^n \hat{t}_i \left\| \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} \right\| \right\} \max_{1 \leq i \leq n} \left[ \left| (\hat{P}_{\kappa i} - P_{\kappa i})' \theta_{\kappa 0} \right|^2 + b_{\kappa 0}(X_i)^2 \right] \\ & \leq C \zeta_0(\kappa) \lambda_{\max}(\tilde{\Phi}_{n\kappa}) (\theta - \theta_{\kappa 0})' (\theta - \theta_{\kappa 0}) + C \kappa^{1/2} O_p(n^{-1} + \kappa^{-2r}) \\ & = O \left[ \zeta_0(\kappa) \|\theta - \theta_{\kappa 0}\|^2 \right] + O \left[ \kappa^{1/2} n^{-1} + \kappa^{1/2} \kappa^{-2r} \right]. \end{aligned}$$

Part (a) now follows by combining (26) with (25).

Now consider part (b). Notice that

$$(27) \quad \begin{aligned} 1_n n^{-1} \sum_{i=1}^n \left\| A \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} \right\|^2 &= 1_n n^{-1} \sum_{i=1}^n \text{trace} \left( A \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} \hat{P}'_{\kappa i} \tilde{\Phi}_{n\kappa}^{-1} A' \right) \\ &\leq C 1_n \text{trace} \left( A \tilde{\Phi}_{n\kappa}^{-1} A' \right) = O_p(1). \end{aligned}$$

Then part (b) can be proved by combining (27) with arguments identical to those used to prove part (a).  $\square$

The next lemma is based on the elegant argument of Welsh (1989).

**Lemma A.6.** *As  $n \rightarrow \infty$ ,*

$$(a) \quad \sup_{\|\theta - \theta_{\kappa 0}\| \leq C(\kappa/n)^{1/2}} 1_n \|H_{n\kappa}(\theta) - H_{n\kappa}(\theta_{\kappa 0})\| = O_p \left[ (\kappa^2/n)^{3/4} (\log n)^{1/2} \right] = o_p \left[ (\kappa/n)^{1/2} \right].$$

$$(b) \quad \sup_{\|\theta - \theta_{\kappa 0}\| \leq C(\kappa/n)^{1/2}} 1_n \|AH_{n\kappa}(\theta) - AH_{n\kappa}(\theta_{\kappa 0})\| = O_p \left[ \kappa/n^{3/4} (\log n)^{1/2} \right] = o_p \left[ n^{-1/2} \right].$$

*Proof.* First, we will prove part (a). Let  $B_n = \{\theta : \|\theta - \theta_{\kappa 0}\| \leq C(d(\kappa)/n)^{1/2}\}$ . As in the proof of Theorem 3.1 of Welsh (1989), cover the ball  $B_n$  with cubes  $\mathcal{C} = \{\mathcal{C}(\theta_l)\}$ , where  $\mathcal{C}(\theta_l)$  is a cube containing  $(\theta_l - \theta_{\kappa 0})$  with sides of  $C(d(\kappa)/n^5)^{1/2}$  such that  $\theta_l \in B_n$ . Then the number of the cubes covering the ball  $B_n$  is  $L = (2n^2)^{d(\kappa)}$ . Also, for each  $l = 1, \dots, L$ , we have that  $\|(\theta - \theta_{\kappa 0}) - (\theta_l - \theta_{\kappa 0})\| \leq C(d(\kappa)/n^5)^{1/2}$  for any  $(\theta - \theta_{\kappa 0}) \in \mathcal{C}(\theta_l)$ .

First note that

$$(28) \quad \begin{aligned} &\sup_{\theta \in B_n} 1_n \|H_{n\kappa}(\theta) - H_{n\kappa}(\theta_{\kappa 0})\| \\ &\leq \max_{1 \leq l \leq L} \sup_{(\theta - \theta_{\kappa 0}) \in \mathcal{C}(\theta_l)} 1_n \|H_{n\kappa}(\theta) - H_{n\kappa}(\theta_l)\| + \max_{1 \leq l \leq L} 1_n \|H_{n\kappa}(\theta_l) - H_{n\kappa}(\theta_{\kappa 0})\|. \end{aligned}$$

Define  $\eta_n = C(d(\kappa)/n^5)^{1/2}$ . Now using the fact that  $1[Y_i \leq \cdot]$  and  $F[\cdot | X_i, Z_i]$  are monotone

increasing functions for each  $i$ , we have

$$\begin{aligned}
& \sup_{(\theta - \theta_{\kappa 0}) \in \mathcal{C}(\theta_l)} 1_n \|H_{n\kappa}(\theta) - H_{n\kappa}(\theta_l)\| \\
& \leq \sup_{(\theta - \theta_{\kappa 0}) \in \mathcal{C}(\theta_l)} n^{-1} \sum_{i=1}^n 1_n t_i \|\Phi_{n\kappa}^{-1} P_{\kappa i}\| \\
& \times \left| \left\{ 1[Y_i \leq P'_{\kappa i}(\theta - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0}] - F[P'_{\kappa i}(\theta - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} | X_i, Z_i] \right\} \right. \\
& \left. - \left\{ 1[Y_i \leq P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0}] - F[P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} | X_i, Z_i] \right\} \right| \\
& \leq n^{-1} \sum_{i=1}^n 1_n t_i \|\Phi_{n\kappa}^{-1} P_{\kappa i}\| \\
& \times \left\{ 1[Y_i \leq P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} + \|P_{\kappa i}\| \eta_n] - F[P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} + \|P_{\kappa i}\| \eta_n | X_i, Z_i] \right\} \\
& - \left\{ 1[Y_i \leq P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0}] - F[P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} | X_i, Z_i] \right\} \\
& \leq \left| n^{-1} \sum_{i=1}^n 1_n t_i \|\Phi_{n\kappa}^{-1} P_{\kappa i}\| \right. \\
& \times \left\{ 1[Y_i \leq P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} + \|P_{\kappa i}\| \eta_n] - F[P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} + \|P_{\kappa i}\| \eta_n | X_i, Z_i] \right\} \\
& \left. - \left\{ 1[Y_i \leq P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0}] - F[P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} | X_i, Z_i] \right\} \right| \\
& + n^{-1} \sum_{i=1}^n 1_n t_i \|\Phi_{n\kappa}^{-1} P_{\kappa i}\| \left\{ F[P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} + \|P_{\kappa i}\| \eta_n | X_i, Z_i] \right. \\
& \left. - F[P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} - \|P_{\kappa i}\| \eta_n | X_i, Z_i] \right\}. \tag{29}
\end{aligned}$$

Consider the second term in (29). Notice that

$$\begin{aligned}
& \max_{1 \leq l \leq L} n^{-1} \sum_{i=1}^n 1_n t_i \|\Phi_{n\kappa}^{-1} P_{\kappa i}\| \left\{ F[P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} + \|P_{\kappa i}\| \eta_n | X_i, Z_i] \right. \\
& \quad \left. - F[P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} - \|P_{\kappa i}\| \eta_n | X_i, Z_i] \right\} \\
& \leq C \eta_m \max_{1 \leq i \leq n} \|P_{\kappa i}\| n^{-1} \sum_{i=1}^n 1_n t_i \|\Phi_{n\kappa}^{-1} P_{\kappa i}\| \\
& \leq C(d(\kappa)/n^{5/2}) \zeta_0(\kappa) d(\kappa)^{1/2} = o_p(n^{-1/2}). \tag{30}
\end{aligned}$$

Now consider the second term in (28), that is  $\max_{1 \leq l \leq L} 1_n \|H_{n\kappa}(\theta_l) - H_{n\kappa}(\theta_{\kappa 0})\|$ . Let  $\Delta_{H_{n\kappa}}^{(j)}(\theta_l)$  denote the  $j$ -th element of  $[H_{n\kappa}(\theta_l) - H_{n\kappa}(\theta_{\kappa 0})]$ . Then we have

$$1_n \Delta_{H_{n\kappa}}^{(j)}(\theta_l) = 1_n e'_{(j)} [H_{n\kappa}(\theta_l) - H_{n\kappa}(\theta_{\kappa 0})],$$

where  $e_{(j)}$  is a unit vector whose components are all zero except for the  $j$ -th component being one. Notice that conditional on  $\{X_1, \dots, X_n, Z_1, \dots, Z_n\}$ , the summands in  $1_n \Delta_{H_{n\kappa}}^{(j)}(\theta_l)$  are independently distributed with mean 0 and that the summands in  $1_n \Delta_{H_{n\kappa}}^{(j)}(\theta_l)$  are bounded uniformly (over  $j$  and  $l$ ) by  $n^{-1}C\zeta_0(\kappa)$  for all sufficiently large  $n$ . Furthermore, the variance of  $1_n \Delta_{H_{n\kappa}}^{(j)}(\theta_l)$  conditional on  $\{X_1, \dots, X_n, Z_1, \dots, Z_n\}$  is bounded by  $1_n n^{-2} \sum_{i=1}^n t_i |e'_{(j)} \Phi_{n\kappa}^{-1} P_{\kappa i}|^2 |P'_{\kappa i}(\theta_l - \theta_{\kappa 0})|$ . Moreover, using the fact that  $f_\varepsilon(0|x, z)$  is bounded away from zero,  $t_i^2 = t_i$ ,  $\max_{1 \leq j \leq d(\kappa)} \max_{1 \leq i \leq n} |e'_{(j)} \Phi_{n\kappa}^{-1} P_{\kappa i}| \leq Cd(\kappa)^{1/2}$ , and that the smallest eigenvalue of  $\Phi_{n\kappa}^{-1}$  is bounded away from zero (when  $1_n = 1$ ) for all  $\kappa$ ,

$$\begin{aligned}
& 1_n n^{-1} \sum_{i=1}^n t_i \left| e'_{(j)} \Phi_{n\kappa}^{-1} P_{\kappa i} \right|^2 |P'_{\kappa i}(\theta_l - \theta_{\kappa 0})| \\
& \leq 1_n n^{-1} \sum_{i=1}^n t_i [\min_i f_i]^{-1} f_i \left| e'_{(j)} \Phi_{n\kappa}^{-1} P_{\kappa i} \right|^2 |P'_{\kappa i}(\theta_l - \theta_{\kappa 0})| \\
& \leq C 1_n \max_{1 \leq i \leq n} \left| e'_{(j)} \Phi_{n\kappa}^{-1} P_{\kappa i} \right| n^{-1} \sum_{i=1}^n t_i f_i \left| e'_{(j)} \Phi_{n\kappa}^{-1} P_{\kappa i} \right| |P'_{\kappa i}(\theta_l - \theta_{\kappa 0})| \\
& \leq C 1_n \max_{1 \leq i \leq n} \left| e'_{(j)} \Phi_{n\kappa}^{-1} P_{\kappa i} \right| \left( n^{-1} \sum_{i=1}^n t_i f_i \left| e'_{(j)} \Phi_{n\kappa}^{-1} P_{\kappa i} \right|^2 \right)^{1/2} \left( n^{-1} \sum_{i=1}^n |P'_{\kappa i}(\theta_l - \theta_{\kappa 0})|^2 \right)^{1/2} \\
& \leq C 1_n d(\kappa)^{1/2} \|\theta_l - \theta_{\kappa 0}\| \left( e'_{(j)} \Phi_{n\kappa}^{-1} \left[ n^{-1} \sum_{i=1}^n t_i f_i P_{\kappa i} P'_{\kappa i} \right] \Phi_{n\kappa}^{-1} e_{(j)} \right)^{1/2} \\
& \leq C 1_n d(\kappa)^{1/2} (d(\kappa)/n)^{1/2} [\lambda_{\max}(\Phi_{n\kappa}^{-1})]^{1/2} \\
& \leq Cd(\kappa)/n^{1/2}
\end{aligned}$$

uniformly (over  $j$  and  $l$ ) for all sufficiently large  $n$ . Therefore, the conditional variance of  $1_n \Delta_{H_{n\kappa}}^{(j)}(\theta_l)$  is bounded uniformly (over  $j$  and  $l$ ) by  $Cn^{-3/2}d(\kappa)$  for all sufficiently large  $n$ . Let  $\varepsilon_n = (d(\kappa)^2/n)^{3/4}(\log n)^{1/2}$ . An application of Bernstein's inequality (see, for example, van der Vaart and Wellner (1996, p.102)) to the sum  $\Delta_{H_{n\kappa}}^{(j)}(\theta_l)$  gives

$$\begin{aligned}
& \Pr \left( \max_{1 \leq l \leq L} 1_n \|H_{n\kappa}(\theta_l) - H_{n\kappa}(\theta_{\kappa 0})\| > C\varepsilon_n \mid X_1, \dots, X_n, Z_1, \dots, Z_n \right) \\
& \leq \sum_{l=1}^L \Pr \left( 1_n \|H_{n\kappa}(\theta_l) - H_{n\kappa}(\theta_{\kappa 0})\| > C\varepsilon_n \mid X_1, \dots, X_n, Z_1, \dots, Z_n \right) \\
& \leq \sum_{l=1}^L \sum_{j=1}^{d(\kappa)} \Pr \left( 1_n \left| \Delta_{H_{n\kappa}}^{(j)}(\theta_l) \right| > C\varepsilon_n d(\kappa)^{-1/2} \mid X_1, \dots, X_n, Z_1, \dots, Z_n \right) \\
& \leq 2(2n^2)^{d(\kappa)} d(\kappa) \exp \left[ -C \frac{\varepsilon_n^2 d(\kappa)^{-1}}{n^{-3/2}d(\kappa) + n^{-1}\zeta_0(\kappa)\varepsilon_n d(\kappa)^{-1/2}} \right]
\end{aligned}$$

$$\begin{aligned}
&\leq C \exp [2d(\kappa) \log(2n) + \log d(\kappa) - Cd(\kappa) \log n] \\
(31) \quad &\leq C \exp [-Cd(\kappa) \log n]
\end{aligned}$$

for all sufficiently large  $n$ . In particular, it is required here that  $(\zeta_0(\kappa)^4/n)(\log n)^2 \rightarrow 0$ .

Now consider the first term in (29). Let  $\tilde{T}_{n\kappa}(\theta_l)$  denote the expression inside  $|\cdot|$  in the first term in (29). Notice that conditional on  $\{X_1, \dots, X_n, Z_1, \dots, Z_n\}$ , the summands in  $\tilde{T}_n(\theta_l)$  are independently distributed with mean 0 and with range bounded by  $n^{-1}C\zeta_0(\kappa)$  and that the variance of  $\tilde{T}_{n\kappa}(\theta_l)$  conditional on  $\{X_1, \dots, X_n, Z_1, \dots, Z_n\}$  is bounded by  $Cn^{-1}d(\kappa)\zeta_0(\kappa)\eta_n = Cn^{-1}(d(\kappa)^2/n^{5/2})\zeta_0(\kappa)$  uniformly over  $l$  for all sufficiently large  $n$ . Another application of Bernstein's inequality to  $\tilde{T}_n(\theta_l)$  gives

$$\begin{aligned}
\Pr(\max_{1 \leq l \leq L} |\tilde{T}_{n\kappa}(\theta_l)| > C\varepsilon_n | X_1, \dots, X_n) &\leq \sum_{l=1}^L \Pr(|\tilde{T}_{n\kappa}(\theta_l)| > C\varepsilon_n | X_1, \dots, X_n) \\
&\leq 2(2n^2)^{d(\kappa)} \exp \left[ -C \frac{\varepsilon_n^2}{n^{-1}(d(\kappa)^2/n^{5/2})\zeta_0(\kappa) + n^{-1}\zeta_0(\kappa)\varepsilon_n} \right] \\
&\leq 2(2n^2)^{d(\kappa)} \exp [-Cn\varepsilon_n/\zeta_0(\kappa)] \\
(32) \quad &\leq C \exp [2d(\kappa) \log(2n) - C[d(\kappa)^{3/2}/\zeta_0(\kappa)]n^{1/4}(\log n)^{1/2}]
\end{aligned}$$

for all sufficiently large  $n$ . Now part (a) of the lemma follows by combining (30), (31), and (32).

To prove part (b), notice that the dimension of  $AH_{n\kappa}(\theta)$  is fixed. Then the desired result follows by repeating arguments identical to those used to prove part (a). In particular,  $\varepsilon_n$  in (31) and (32) is replaced with  $d(\kappa)/n^{3/4}(\log n)^{1/2}$  and  $C\varepsilon_n$  is substituted for  $C\varepsilon_n d(\kappa)^{-1/2}$  in the third line of (31).  $\square$

**Lemma A.7.** *As  $n \rightarrow \infty$ ,*

$$\sup_{\|\theta - \theta_{\kappa 0}\| \leq C(\kappa/n)^{1/2}} 1_n \left\| \hat{H}_{n\kappa}(\theta) - H_{n\kappa}(\theta) \right\| = o_p \left( n^{-1/2} \right).$$

*Proof.* The proof of Lemma A.7 is analogous to that of Lemma A.6. As in the proof of Lemma A.6, let  $B_n = \{\theta : \|\theta - \theta_{\kappa 0}\| \leq C(d(\kappa)/n)^{1/2}\}$  and cover the ball  $B_n$  with cubes  $\mathcal{C} = \{\mathcal{C}(\theta_l)\}$ , where  $\mathcal{C}(\theta_l)$  is a cube containing  $(\theta_l - \theta_{\kappa 0})$  with sides of  $C(d(\kappa)/n^5)^{1/2}$  such that  $\theta_l \in B_n$ . Recall that the number of the cubes covering the ball  $B_n$  is  $L = (2n^2)^{d(\kappa)}$ . and that for each  $l = 1, \dots, L$ ,  $\|(\theta - \theta_{\kappa 0}) - (\theta_l - \theta_{\kappa 0})\| \leq C(d(\kappa)/n^{5/2}) \equiv \eta_n$  for any  $(\theta - \theta_{\kappa 0}) \in \mathcal{C}(\theta_l)$ .

Now define

$$\begin{aligned} \tilde{H}_{n\kappa}(\theta) = & \\ & - n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1}^n t_i \left\{ 1[Y_i \leq \hat{P}'_{\kappa i}(\theta - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0}] - F[\hat{P}'_{\kappa i}(\theta - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0} | X_i, Z_i] \right\} P_{\kappa i}. \end{aligned}$$

Then

$$(33) \quad \begin{aligned} & \sup_{\theta \in B_n} 1_n \left\| \hat{H}_{n\kappa}(\theta) - H_{n\kappa}(\theta) \right\| \\ & \leq \max_{1 \leq l \leq L} \sup_{(\theta - \theta_{\kappa 0}) \in \mathcal{C}(\theta_l)} \left[ 1_n \left\| \hat{H}_{n\kappa}(\theta) - \tilde{H}_{n\kappa}(\theta) \right\| + 1_n \left\| \tilde{H}_{n\kappa}(\theta) - H_{n\kappa}(\theta) \right\| \right]. \end{aligned}$$

Using the fact that  $1[Y_i \leq \cdot]$  and  $F[\cdot | X_i, Z_i]$  are monotone increasing functions, write the first term in (33) as

$$(34) \quad \begin{aligned} & \max_{1 \leq l \leq L} \sup_{(\theta - \theta_{\kappa 0}) \in \mathcal{C}(\theta_l)} 1_n \left\| \hat{H}_{n\kappa}(\theta) - \tilde{H}_{n\kappa}(\theta) \right\| \\ & \leq \max_{1 \leq l \leq L} \sup_{(\theta - \theta_{\kappa 0}) \in \mathcal{C}(\theta_l)} 1_n n^{-1} \sum_{i=1}^n \left\| \hat{t}_i \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} - t_i \Phi_{n\kappa}^{-1} P_{\kappa i} \right\| \\ & \quad \times \left| \left\{ 1[Y_i \leq \hat{P}'_{\kappa i}(\theta - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0}] - F[\hat{P}'_{\kappa i}(\theta - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0} | X_i, Z_i] \right\} \right| \\ & \leq \left| \max_{1 \leq l \leq L} 1_n n^{-1} \sum_{i=1}^n \left\| \hat{t}_i \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} - t_i \Phi_{n\kappa}^{-1} P_{\kappa i} \right\| \right| \\ & \quad \times \left\{ 1[Y_i \leq \hat{P}'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0}] + \left\| \hat{P}_{\kappa i} \right\| \eta_n \right\} - F[\hat{P}'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0} + \left\| \hat{P}_{\kappa i} \right\| \eta_n | X_i, Z_i] \left| \right\} \\ & \quad + \max_{1 \leq l \leq L} 1_n n^{-1} \sum_{i=1}^n \left\| \hat{t}_i \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} - t_i \Phi_{n\kappa}^{-1} P_{\kappa i} \right\| \\ & \quad \times \left\{ F[\hat{P}'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0} + \left\| \hat{P}_{\kappa i} \right\| \eta_n | X_i, Z_i] - F[\hat{P}'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0} - \left\| \hat{P}_{\kappa i} \right\| \eta_n | X_i, Z_i] \right\}. \end{aligned}$$

As in (30), write the second term in (34) further as

$$(35) \quad \begin{aligned} & \max_{1 \leq l \leq L} 1_n n^{-1} \sum_{i=1}^n \left\| \hat{t}_i \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} - t_i \Phi_{n\kappa}^{-1} P_{\kappa i} \right\| \\ & \quad \times \left\{ F[\hat{P}'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0} + \left\| \hat{P}_{\kappa i} \right\| \eta_n | X_i, Z_i] - F[\hat{P}'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0} - \left\| \hat{P}_{\kappa i} \right\| \eta_n | X_i, Z_i] \right\} \\ & \leq C \eta_n \max_{1 \leq i \leq n} \left\| \hat{P}_{\kappa i} \right\| 1_n n^{-1} \sum_{i=1}^n \left\| \hat{t}_i \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} - t_i \Phi_{n\kappa}^{-1} P_{\kappa i} \right\| \\ & \leq \eta_n \zeta_0(\kappa) o_p(1), \end{aligned}$$

where the last inequality follows from (22). Therefore, the second term in (34) is of order  $o_p(n^{-1/2})$ .

Now consider the first term in (34). Notice that  $\hat{t}_i$ ,  $\tilde{\Phi}_{n\kappa}^{-1}$ , and  $\hat{P}_{\kappa i}$  depend only on  $\{(X_i, Z_i) : i = 1, \dots, n\}$ . Hence, conditional on  $\{X_1, \dots, X_n, Z_1, \dots, Z_n\}$ , the summands in the expression inside  $|\cdot|$  in the first term in (34) are independently distributed with mean 0. Also, the sum of the variances of the summands conditional on  $\{X_1, \dots, X_n, Z_1, \dots, Z_n\}$  is bounded by  $C1_n n^{-2} \sum_{i=1}^n \left\| \hat{t}_i \tilde{\Phi}_{n\kappa}^{-1} \hat{P}_{\kappa i} - t_i \Phi_{n\kappa}^{-1} P_{\kappa i} \right\|^2$ . Notice that (22) implies that this bound is of order  $O_p(n^{-1-c^*})$ , where  $c^*$  is a positive constant. Then as in (31) and (32), applying Bernstein's inequality yields that the first term in (34) is of order  $o_p(n^{-1/2})$ .

Again using the fact that  $1[Y_i \leq \cdot]$  and  $F[\cdot | X_i, Z_i]$  are monotone increasing functions, write the second term in (33) as

$$\begin{aligned}
& \max_{1 \leq l \leq L} \sup_{(\theta - \theta_{\kappa 0}) \in \mathcal{C}(\theta_l)} 1_n \left\| \tilde{H}_{n\kappa}(\theta) - H_{n\kappa}(\theta) \right\| \\
& \leq \max_{1 \leq l \leq L} \sup_{(\theta - \theta_{\kappa 0}) \in \mathcal{C}(\theta_l)} 1_n n^{-1} \sum_{i=1}^n t_i \left\| \Phi_{n\kappa}^{-1} P_{\kappa i} \right\| \\
& \times \left| \left\{ 1[Y_i \leq \hat{P}'_{\kappa i}(\theta - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0}] - F[\hat{P}'_{\kappa i}(\theta - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0} | X_i, Z_i] \right\} \right. \\
& \left. - \left\{ 1[Y_i \leq P'_{\kappa i}(\theta - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0}] - F[P'_{\kappa i}(\theta - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} | X_i, Z_i] \right\} \right| \\
& \leq \left| \max_{1 \leq l \leq L} 1_n n^{-1} \sum_{i=1}^n t_i \left\| \Phi_{n\kappa}^{-1} P_{\kappa i} \right\| \right. \\
& \times \left[ \left\{ 1[Y_i \leq \hat{P}'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0} + \left\| \hat{P}_{\kappa i} \right\| \eta_n] - F[\hat{P}'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0} + \left\| \hat{P}_{\kappa i} \right\| \eta_n | X_i, Z_i] \right\} \right. \\
& \left. - \left\{ 1[Y_i \leq P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} - \left\| P_{\kappa i} \right\| \eta_n] - F[P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} - \left\| P_{\kappa i} \right\| \eta_n | X_i, Z_i] \right\} \right| \\
& + \max_{1 \leq l \leq L} 1_n n^{-1} \sum_{i=1}^n t_i \left\| \Phi_{n\kappa}^{-1} P_{\kappa i} \right\| \\
& \times \left[ \left\{ F[\hat{P}'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0} + \left\| \hat{P}_{\kappa i} \right\| \eta_n | X_i, Z_i] - F[\hat{P}'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + \hat{P}'_{\kappa i} \theta_{\kappa 0} - \left\| \hat{P}_{\kappa i} \right\| \eta_n | X_i, Z_i] \right\} \right. \\
& \left. + \left\{ F[P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} + \left\| P_{\kappa i} \right\| \eta_n | X_i, Z_i] - F[P'_{\kappa i}(\theta_l - \theta_{\kappa 0}) + P'_{\kappa i} \theta_{\kappa 0} - \left\| P_{\kappa i} \right\| \eta_n | X_i, Z_i] \right\} \right]
\end{aligned}$$

Recall that  $\hat{t}_i$ ,  $\tilde{\Phi}_{n\kappa}^{-1}$ , and  $\hat{P}_{\kappa i}$  depend only on  $\{(X_i, Z_i) : i = 1, \dots, n\}$ . Hence, conditional on  $\{X_1, \dots, X_n, Z_1, \dots, Z_n\}$ , the summands in the expression inside  $|\cdot|$  in the first term above are independently distributed with mean 0. The sum of the variances of the summands conditional on  $\{X_1, \dots, X_n, Z_1, \dots, Z_n\}$  is bounded by



$Cn^{-1}d(\kappa) \max_{1 \leq i \leq n} t_i \left[ (\hat{P}_{\kappa i} - P_{\kappa i})'(\theta_l - \theta_{\kappa 0}) + (\hat{P}_{\kappa i} - P_{\kappa i})'\theta_{\kappa 0} + 2 \left\| \hat{P}_{\kappa i} \right\| \eta_n \right]$ , which is of order  $O_p(n^{-1-c^*})$ , where  $c^*$  is a positive constant. Then as in (31) and (32), applying Bernstein's inequality yields that the first term above is of order  $o_p(n^{-1/2})$ . Also, arguments identical to those used in (30) show that the second term above is of order  $O_p \left[ (d(\kappa)/n^{5/2})\zeta_0(\kappa)d(\kappa)^{1/2} \right] = o_p(n^{-1/2})$ . Now the lemma follows by combining all the results obtained above.  $\square$

**Lemma A.8.** *As  $n \rightarrow \infty$ ,*

$$\left\| \hat{\theta}_{n\kappa} - \theta_{\kappa 0} \right\| = O_p \left[ (\kappa/n)^{1/2} \right].$$

*Proof.* Write

$$(36) \quad \begin{aligned} 1_n \hat{G}_{n\kappa}(\hat{\theta}_{n\kappa}) &= 1_n H_{n\kappa}(\theta_{\kappa 0}) + 1_n [H_{n\kappa}(\hat{\theta}_{n\kappa}) - H_{n\kappa}(\theta_{\kappa 0})] + 1_n [\hat{H}_{n\kappa}(\hat{\theta}_{n\kappa}) - H_{n\kappa}(\hat{\theta}_{n\kappa})] \\ &\quad + 1_n \hat{G}_{n\kappa}^*(\hat{\theta}_{n\kappa}). \end{aligned}$$

By part (a) of Lemma A.5, (36) can be rewritten as

$$(37) \quad \begin{aligned} 1_n(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) &= -1_n \hat{G}_{n\kappa}(\hat{\theta}_{n\kappa}) + 1_n H_{n\kappa}(\theta_{\kappa 0}) \\ &\quad + 1_n [H_{n\kappa}(\hat{\theta}_{n\kappa}) - H_{n\kappa}(\theta_{\kappa 0})] + 1_n [\hat{H}_{n\kappa}(\hat{\theta}_{n\kappa}) - H_{n\kappa}(\hat{\theta}_{n\kappa})] \\ &\quad - 1_n n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1}^n t_i f_i \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) P_{\kappa i} + R_{n\kappa}^*(\hat{\theta}_{n\kappa}), \end{aligned}$$

where  $\left\| R_{n\kappa}^*(\hat{\theta}_{n\kappa}) \right\| = o_p \left[ (\kappa/n)^{1/2} \right] + O_p \left[ \zeta_0(\kappa) \left\| \hat{\theta}_{n\kappa} - \theta_{\kappa 0} \right\|^2 \right]$ .

Now suppose that  $\left\| \hat{\theta}_{n\kappa} - \theta_{\kappa 0} \right\| \leq C(\kappa/n)^{1/2}$  for any constant  $C > 0$ . It is not difficult to show that

$$(38) \quad 1_n \left\| n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1}^n t_i f_i \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) P_{\kappa i} \right\| = O_p(n^{-1/2}).$$

Then, by applying equation (38), Lemmas A.3, A.4 (a), A.6 (a), and A.7 to equation (37), we have

$$(39) \quad \begin{aligned} 1_n \left\| \hat{\theta}_{n\kappa} - \theta_{\kappa 0} \right\| &\leq +1_n \left\| \hat{G}_{n\kappa}(\hat{\theta}_{n\kappa}) \right\| + 1_n \left\| H_{n\kappa}(\theta_{\kappa 0}) \right\| \\ &\quad + 1_n \left\| H_{n\kappa}(\hat{\theta}_{n\kappa}) - H_{n\kappa}(\theta_{\kappa 0}) \right\| + 1_n \left\| \hat{H}_{n\kappa}(\hat{\theta}_{n\kappa}) - H_{n\kappa}(\hat{\theta}_{n\kappa}) \right\| \\ &\quad + 1_n \left\| n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1}^n t_i f_i \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) P_{\kappa i} \right\| + \left\| R_{n\kappa}^*(\hat{\theta}_{n\kappa}) \right\| \\ &= O_p \left[ (\kappa/n)^{1/2} \right]. \end{aligned}$$

Therefore, the right-hand side of (39) is less than  $C(\kappa/n)^{1/2}$  with probability approaching one (w.p.a.1). This implies that w.p.a.1,

$$1_n \left\| \hat{\theta}_{n\kappa} - \theta_{\kappa 0} \right\| \leq C(\kappa/n)^{1/2},$$

which in turn implies that w.p.a.1,

$$\left\| \hat{\theta}_{n\kappa} - \theta_{\kappa 0} \right\| \leq C(\kappa/n)^{1/2}$$

since  $\Pr(1_n = 1) \rightarrow 1$  as  $n \rightarrow \infty$ . Therefore, the lemma is proved.  $\square$

*Proof of Theorem 3.1.* Define

$$\begin{aligned} \Sigma_\kappa &= \tau(1 - \tau)E[t(W)P_\kappa(W)P_\kappa(W)'] \\ \Gamma_\kappa &= E \left[ t(W)f_\varepsilon(0|X, Z) \frac{d\lambda(V)}{dv} P_\kappa(W)(1, Z') \right] \\ \Omega_\kappa &= A\Phi_\kappa^{-1} (\Sigma_\kappa + \Gamma_\kappa \Sigma_{\mu, \pi} \Gamma_\kappa') \Phi_\kappa^{-1} A' \end{aligned}$$

Let  $\varphi_\kappa(w) = A\Phi_\kappa^{-1}P_\kappa(w)$ , so that

$$\varphi_\kappa(w) = E[t(W)f_\varepsilon(0|X, Z)\varphi(W)P_\kappa(W)'] E[t(W)f_\varepsilon(0|X, Z)P_\kappa(W)P_\kappa(W)']^{-1} P_\kappa(w).$$

Then

$$\begin{aligned} \Omega_\kappa &= \tau(1 - \tau)E[t(W)\varphi_\kappa(W)\varphi_\kappa(W)'] \\ &\quad + E[t(W)f_\varepsilon(0|X, Z) \frac{d\lambda(V)}{dv} \varphi_\kappa(W)(1, Z')] \Sigma_{\mu, \pi} E[t(W)f_\varepsilon(0|X, Z) \frac{d\lambda(V)}{dv} \begin{pmatrix} 1 \\ Z \end{pmatrix} \varphi_\kappa(W)']. \end{aligned}$$

Notice that  $\varphi_\kappa(w)$  is the  $(t(w)f_\varepsilon(0|x, z)$ -weighted) mean square projection of  $\varphi(w)$  on the approximating functions. Also, using equation (9) and the fact that  $f_\varepsilon(0|x, z)$  is bounded away from zero, we have

$$\begin{aligned} E[t(W) \|\varphi(W) - \varphi_\kappa(W)\|^2] &\leq CE[t(W)f_\varepsilon(0|X, Z) \|\varphi(W) - \varphi_\kappa(W)\|^2] \\ &\leq E \left[ t(W)f_\varepsilon(0|X, Z) \left\| \varphi(W) - \tilde{\Theta}_\kappa P_\kappa(W) \right\|^2 \right] \\ &\rightarrow 0 \end{aligned}$$

as  $\kappa \rightarrow \infty$ . This implies that  $\|\Omega_\kappa - \Omega\| \rightarrow 0$ , which in turn implies that  $\|\Omega_\kappa^{-1/2}\|$  is bounded.

Now write

$$\begin{aligned} (40) \quad 1_n \Omega_\kappa^{-1/2} A \hat{G}_{n\kappa}(\hat{\theta}_{n\kappa}) &= 1_n \Omega_\kappa^{-1/2} A H_{n\kappa}(\theta_{\kappa 0}) + 1_n \Omega_\kappa^{-1/2} [A H_{n\kappa}(\hat{\theta}_{n\kappa}) - A H_{n\kappa}(\theta_{\kappa 0})] \\ &\quad + 1_n \Omega_\kappa^{-1/2} [A \hat{H}_{n\kappa}(\hat{\theta}_{n\kappa}) - A H_{n\kappa}(\hat{\theta}_{n\kappa})] + 1_n \Omega_\kappa^{-1/2} A \hat{G}_{n\kappa}^*(\hat{\theta}_{n\kappa}). \end{aligned}$$

By Lemma A.8 and part (b) of Lemma A.5, (40) can be rewritten as

$$\begin{aligned}
1_n \Omega_\kappa^{-1/2} A(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) &= -1_n \Omega_\kappa^{-1/2} A \hat{G}_{n\kappa}(\hat{\theta}_{n\kappa}) + 1_n \Omega_\kappa^{-1/2} A H_{n\kappa}(\theta_{\kappa 0}) \\
&\quad + 1_n \Omega_\kappa^{-1/2} [A H_{n\kappa}(\hat{\theta}_{n\kappa}) - A H_{n\kappa}(\theta_{\kappa 0})] \\
&\quad + 1_n \Omega_\kappa^{-1/2} [A \hat{H}_{n\kappa}(\hat{\theta}_{n\kappa}) - A H_{n\kappa}(\hat{\theta}_{n\kappa})] \\
&\quad - 1_n n^{-1} \Omega_\kappa^{-1/2} A \Phi_{n\kappa}^{-1} \sum_{i=1}^n t_i f_i \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) P_{\kappa i} + \Omega_\kappa^{-1/2} A R_{n\kappa}^*(\hat{\theta}_{n\kappa}),
\end{aligned}$$

where  $\left\| \Omega_\kappa^{-1/2} A R_{n\kappa}^*(\hat{\theta}_{n\kappa}) \right\| = o_p [n^{-1/2}]$ . By combining Lemmas A.3, A.6 (b), and A.7 with the fact that  $\left\| \Omega_\kappa^{-1/2} \right\|$  is bounded,

$$\begin{aligned}
(41) \quad 1_n \Omega_\kappa^{-1/2} A(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) &= 1_n \Omega_\kappa^{-1/2} A H_{n\kappa}(\theta_{\kappa 0}) \\
&\quad - 1_n n^{-1} \Omega_\kappa^{-1/2} A \Phi_{n\kappa}^{-1} \sum_{i=1}^n t_i f_i \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) P_{\kappa i} + R_{n\kappa},
\end{aligned}$$

where the remainder  $R_{n\kappa}$  satisfies  $\|R_{n\kappa}\| = o_p [n^{-1/2}]$ .

Define

$$\bar{H}_{n\kappa}(\theta_{\kappa 0}) = n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1}^n t_i \left\{ \tau - 1 [Y_i \leq m_i] \right\} P_{\kappa i}.$$

By arguments similar to those used in the proof of Lemma A.4, we have

$$E \left[ 1_n \left\| \Omega_\kappa^{-1/2} A H_{n\kappa}(\theta_{\kappa 0}) - \Omega_\kappa^{-1/2} A \bar{H}_{n\kappa}(\theta_{\kappa 0}) \right\|^2 \middle| X_1, \dots, X_n, Z_1, \dots, Z_n \right] \leq C n^{-1} \sup_i |b_{\kappa i}|.$$

Therefore, by Markov's inequality,

$$(42) \quad 1_n \left\| \Omega_\kappa^{-1/2} A H_{n\kappa}(\theta_{\kappa 0}) - \Omega_\kappa^{-1/2} A \bar{H}_{n\kappa}(\theta_{\kappa 0}) \right\| = o_p(n^{-1/2}).$$

It follows from part (b) of Lemma A.1 that

$$(43) \quad 1_n \left\| \Omega_\kappa^{-1/2} A \bar{H}_{n\kappa}(\theta_{\kappa 0}) - \Omega_\kappa^{-1/2} n^{-1} A \Phi_\kappa^{-1} \sum_{i=1}^n t_i \left\{ \tau - 1 [Y_i \leq m_i] \right\} P_{\kappa i} \right\| = o_p(n^{-1/2}).$$

Also, notice that

$$\begin{aligned}
& -1_n n^{-1} \Omega_\kappa^{-1/2} A \Phi_{n\kappa}^{-1} \sum_{i=1}^n t_i f_i \frac{d\lambda(V_i)}{dv} (\hat{V}_i - V_i) P_{\kappa i} \\
&= 1_n \left[ n^{-1} \Omega_\kappa^{-1/2} A \Phi_{n\kappa}^{-1} \sum_{i=1}^n t_i f_i \frac{d\lambda(V_i)}{dv} P_{\kappa i}(1, Z_i') \right] \begin{pmatrix} \hat{\mu}(\alpha) - \mu(\alpha) \\ \hat{\pi}(\alpha) - \pi(\alpha) \end{pmatrix} + R_{n\kappa,1} \\
&= 1_n \Omega_\kappa^{-1/2} A \Phi_\kappa^{-1} \Gamma_\kappa \begin{pmatrix} \hat{\mu}(\alpha) - \mu(\alpha) \\ \hat{\pi}(\alpha) - \pi(\alpha) \end{pmatrix} + R_{n\kappa,2} \\
(44) \quad &= 1_n \Omega_\kappa^{-1/2} A \Phi_\kappa^{-1} \Gamma_\kappa n^{-1} \sum_{i=1}^n \Delta_{\mu,\pi}(X_i, Z_i) + R_{n\kappa,3},
\end{aligned}$$

where  $\|R_{n\kappa,j}\| = o_p[n^{-1/2}]$  for  $j = 1, 2, 3$ . It follows from (41), (42), (43), and (44) that

$$\begin{aligned}
(45) \quad & 1_n \Omega_\kappa^{-1/2} A(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) = 1_n \Omega_\kappa^{-1/2} n^{-1} A \Phi_\kappa^{-1} \sum_{i=1}^n t_i \left\{ \tau - 1[Y_i \leq m_i] \right\} P_{\kappa i} \\
& \quad + 1_n \Omega_\kappa^{-1/2} A \Phi_\kappa^{-1} \Gamma_\kappa n^{-1} \sum_{i=1}^n \Delta_{\mu,\pi}(X_i, Z_i) + \tilde{R}_{n\kappa},
\end{aligned}$$

where  $\|\tilde{R}_{n\kappa}\| = o_p[n^{-1/2}]$ .

For any  $[(1 + d_{z_1}) \times 1]$  vector  $c$  with  $\|c\| = 1$ , let

$$\nu_{in} = n^{-1/2} c' \Omega_\kappa^{-1/2} A \Phi_\kappa^{-1} \left[ t_i \left\{ \tau - 1[Y_i \leq m_i] \right\} P_{\kappa i} + \Gamma_\kappa \Delta_{\mu,\pi}(X_i, Z_i) \right].$$

Using the fact that  $\Pr(1_n = 1) \rightarrow 1$  as  $n \rightarrow \infty$ ,

$$(46) \quad n^{1/2} c' \Omega_\kappa^{-1/2} A(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) = \sum_{i=1}^n \nu_{in} + o_p(1).$$

As in the proof of Lemma A2 of Newey, Powell, and Vella (1999), it can be shown that  $\nu_{in}$  is i.i.d. for each  $n$ ,  $E[\nu_{in}] = 0$ ,  $E[\nu_{in}^2] = 1/n$ , and for any  $\delta > 0$ ,

$$nE[1(|\nu_{in}| > \delta) |\nu_{in}|^2] \leq n\delta^{-2} E[|\nu_{in}|^4] = O[\zeta_0(\kappa)^2 \kappa/n] = o(1).$$

Therefore, by the Cramér-Wold device and the Lindeberg-Feller central limit theorem,

$$n^{1/2} \Omega_\kappa^{-1/2} A(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) \rightarrow_d \mathbf{N}(0, I).$$

Now the desired result follows from the fact that  $\|\Omega_\kappa - \Omega\| \rightarrow 0$ . □

The following lemmas are useful to prove theorem 3.2.

**Lemma A.9.**  $\left\| \hat{\Phi}_{n\kappa} - \Phi_\kappa \right\| = o_p(1)$ .

*Proof.* In view of Lemma A.1 (b) and (c), it suffices to show that  $\left\| \hat{\Phi}_{n\kappa} - \tilde{\Phi}_{n\kappa} \right\| = o_p(1)$ .

First, define

$$\bar{\Phi}_{n\kappa} = (nh_n)^{-1} \sum_{i=1}^n \hat{t}_i K\left(\frac{\varepsilon_i}{h_n}\right) \hat{P}_{\kappa i} \hat{P}'_{\kappa i}.$$

Then by  $\max_{1 \leq i \leq n} \hat{t}_i |\hat{\varepsilon}_i - \varepsilon_i| = O_p(\kappa/n^{1/2})$ ,

$$\begin{aligned} \left\| \hat{\Phi}_{n\kappa} - \bar{\Phi}_{n\kappa} \right\| &\leq h_n^{-1} \max_{1 \leq i \leq n} \hat{t}_i \left| K\left(\frac{\hat{\varepsilon}_i}{h_n}\right) - K\left(\frac{\varepsilon_i}{h_n}\right) \right| n^{-1} \sum_{i=1}^n \hat{t}_i \left\| \hat{P}_{\kappa i} \right\|^2 \\ &\leq Ch_n^{-2} \max_{1 \leq i \leq n} \hat{t}_i |\hat{\varepsilon}_i - \varepsilon_i| O_p(\kappa) \\ &= O_p\left(h_n^{-2} \kappa^2 n^{-1/2}\right) = o_p(1), \end{aligned}$$

where the last equality follows from Assumption 3.14. By the triangle inequality, it remains to show that

$$(47) \quad \left\| \bar{\Phi}_{n\kappa} - \tilde{\Phi}_{n\kappa} \right\| = o_p(1).$$

To do so, define

$$\Xi_{n\kappa 1} = (nh_n)^{-1} \sum_{i=1}^n \hat{t}_i \left\{ K\left(\frac{\varepsilon_i}{h_n}\right) - E\left[K\left(\frac{\varepsilon_i}{h_n}\right) \middle| X_i, Z_i\right] \right\} \hat{P}_{\kappa i} \hat{P}'_{\kappa i}$$

and

$$\Xi_{n\kappa 2} = n^{-1} \sum_{i=1}^n \hat{t}_i \left\{ E\left[h_n^{-1} K\left(\frac{\varepsilon_i}{h_n}\right) \middle| X_i, Z_i\right] - f_\varepsilon(0|X_i, Z_i) \right\} \hat{P}_{\kappa i} \hat{P}'_{\kappa i}.$$

Let  $\Xi_{n\kappa 1}^{(j,k)}$  denote the  $(j, k)$  element of  $\Xi_{n\kappa 1}$ . Then

$$\Xi_{n\kappa 1}^{(j,k)} = (nh_n)^{-1} \sum_{i=1}^n \hat{t}_i \left\{ K\left(\frac{\varepsilon_i}{h_n}\right) - E\left[K\left(\frac{\varepsilon_i}{h_n}\right) \middle| X_i, Z_i\right] \right\} e'_{(j)} \hat{P}_{\kappa i} \hat{P}'_{\kappa i} e_{(k)}.$$

Notice that conditional on  $\{X_1, \dots, X_n, Z_1, \dots, Z_n\}$ , the summands in  $\Xi_{n\kappa 1}^{(j,k)}$  are i.i.d. with mean zero. Also, the conditional variance of  $\Xi_{n\kappa 1}^{(j,k)}$  is bounded by  $Cn^{-1}h_n^{-1}d(\kappa)$ , where  $C$  can be chosen uniformly over  $j$  and  $k$ . Then by Bernstein's inequality,

$$\begin{aligned} &\Pr\left(\left\|\Xi_{n\kappa 1}\right\| > C\left(\frac{d(\kappa)^3 \log n}{nh_n}\right)^{1/2} \middle| X_1, \dots, X_n, Z_1, \dots, Z_n\right) \\ &\leq \sum_{j=1}^{d(\kappa)} \sum_{k=1}^{d(\kappa)} \Pr\left(\left|\Xi_{n\kappa 1}^{(j,k)}\right| > C\left(\frac{d(\kappa) \log n}{nh_n}\right)^{1/2} \middle| X_1, \dots, X_n, Z_1, \dots, Z_n\right) \\ &\leq Cd(\kappa)^2 \exp(-C \log n) \end{aligned}$$

for all sufficiently large  $n$ . This implies that  $\|\Xi_{n\kappa 1}\| = o_p(1)$  because  $(\kappa^3 \log n)/(nh_n) \rightarrow 0$  by Assumption 3.14.

Now consider  $\Xi_{n\kappa 2}$ . Then by Assumption 3.12,

$$\max_{1 \leq i \leq n} \left| E \left[ h_n^{-1} K \left( \frac{\varepsilon_i}{h_n} \right) \middle| X_i, Z_i \right] - f_\varepsilon(0|X_i, Z_i) \right| = O_p(h_n^2),$$

which implies that

$$\|\Xi_{n\kappa 2}\| = O_p(h_n^2 \kappa) = o_p(1),$$

where the last equality follows from Assumption 3.14. Finally, (47) follows by the triangle inequality.  $\square$

**Lemma A.10.**  $\|\hat{\Sigma}_{n\kappa} - \Sigma_\kappa\| = o_p(1)$ .

*Proof.* This can be proved using the arguments identical to those used to prove Lemma A.1 (b) and (c).  $\square$

**Lemma A.11.**  $\|\hat{\Gamma}_{n\kappa} - \Gamma_\kappa\| = o_p(1)$ .

*Proof.* First note that

$$\begin{aligned} & \max_{1 \leq i \leq n} \hat{t}_i \left| \frac{d\hat{\lambda}(\hat{V}_i)}{dv} - \frac{d\lambda(V_i)}{dv} \right| \\ & \leq \max_{1 \leq i \leq n} \hat{t}_i \left| \frac{d\hat{\lambda}(\hat{V}_i)}{dv} - \frac{d\lambda(\hat{V}_i)}{dv} \right| + \max_{1 \leq i \leq n} \hat{t}_i \left| \frac{d\lambda(\hat{V}_i)}{dv} - \frac{d\lambda(V_i)}{dv} \right| \\ & = O_p[\zeta_1(\kappa)(k/n)^{1/2}] + O_p[\max_i \hat{t}_i |\hat{V}_i - V_i|] \\ & = O_p[\zeta_1(\kappa)(k/n)^{1/2} + n^{-1/2}]. \end{aligned}$$

Define

$$\tilde{\Gamma}_{n\kappa} = (nh_n)^{-1} \sum_{i=1}^n \hat{t}_i K \left( \frac{\hat{\varepsilon}_i}{h_n} \right) \frac{d\lambda(V_i)}{dv} \hat{P}_{\kappa i}(1, Z'_i).$$

Then

$$\begin{aligned} & \|\hat{\Gamma}_{n\kappa} - \tilde{\Gamma}_{n\kappa}\| \\ & \leq \max_{1 \leq i \leq n} \hat{t}_i \left| \frac{d\hat{\lambda}(\hat{V}_i)}{dv} - \frac{d\lambda(V_i)}{dv} \right| \zeta_0(\kappa) (nh_n)^{-1} \sum_{i=1}^n \hat{t}_i \left\| K \left( \frac{\hat{\varepsilon}_i}{h_n} \right) \right\| \|(1, Z'_i)\| \\ & = O_p[\zeta_0(\kappa) \zeta_1(\kappa) (k/n)^{1/2}] = o_p(1), \end{aligned}$$

where the last equality follows from Assumption 3.14. Now it remains to show that  $\|\tilde{\Gamma}_{n\kappa} - \Gamma_\kappa\| = o_p(1)$ . This can be proved by using identical arguments as in the proof of Lemma A.9.  $\square$

*Proof of Theorem 3.2.* Recall that it is shown in the proof of Theorem 3.1 that  $\|\Omega_\kappa - \Omega\| \rightarrow_p 0$ . Thus, it suffices to prove that  $\|\hat{\Omega}_{n\kappa} - \Omega_\kappa\| \rightarrow_p 0$ . To do so, define

$$\begin{aligned}\hat{\Omega}_{n\kappa 1} &= A\hat{\Phi}_{n\kappa}^{-1}\hat{\Sigma}_{n\kappa}\hat{\Phi}_{n\kappa}^{-1}A', \\ \hat{\Omega}_{n\kappa 2} &= A\hat{\Phi}_{n\kappa}^{-1}\hat{\Gamma}_{n\kappa}\hat{\Sigma}_{\mu,\pi}\hat{\Gamma}'_{n\kappa}\hat{\Phi}_{n\kappa}^{-1}A', \\ \Omega_{\kappa 1} &= A\Phi_\kappa^{-1}\Sigma_\kappa\Phi_\kappa^{-1}A', \quad \text{and} \\ \Omega_{\kappa 2} &= A\Phi_\kappa^{-1}\Gamma_\kappa\Sigma_{\mu,\pi}\Gamma'_\kappa\Phi_\kappa^{-1}A'.\end{aligned}$$

Note that by Lemmas A.9 and A.10,  $\|\Phi_\kappa - \hat{\Phi}_{n\kappa}\| = o_p(1)$ , and  $\|\hat{\Sigma}_{n\kappa} - \Sigma_\kappa\| = o_p(1)$ . Also, it can be shown that  $\|A\Phi_\kappa^{-1}\| = O(1)$ , and  $\|A\hat{\Phi}_{n\kappa}^{-1}\| = O_p(1)$ . Then by  $\lambda_{\max}(\Phi_\kappa^{-1})$ ,  $\lambda_{\max}(\hat{\Phi}_{n\kappa}^{-1})$ , and  $\lambda_{\max}(\Sigma_\kappa)$  bounded,

$$\begin{aligned}& \|\hat{\Omega}_{n\kappa 1} - \Omega_{\kappa 1}\| \\ & \leq \|A\hat{\Phi}_{n\kappa}^{-1}(\hat{\Sigma}_{n\kappa} - \Sigma_\kappa)\hat{\Phi}_{n\kappa}^{-1}A'\| + \|A(\hat{\Phi}_{n\kappa}^{-1}\Sigma_\kappa\hat{\Phi}_{n\kappa}^{-1} - \Phi_\kappa^{-1}\Sigma_\kappa\Phi_\kappa^{-1})A'\| \\ & \leq \|A\hat{\Phi}_{n\kappa}^{-1}(\hat{\Sigma}_{n\kappa} - \Sigma_\kappa)\hat{\Phi}_{n\kappa}^{-1}A'\| + \|A(\hat{\Phi}_{n\kappa}^{-1} - \Phi_\kappa^{-1})\Sigma_\kappa\hat{\Phi}_{n\kappa}^{-1}A'\| + \|A\Phi_\kappa^{-1}\Sigma_\kappa(\hat{\Phi}_{n\kappa}^{-1} - \Phi_\kappa^{-1})A'\| \\ & \leq \|A\hat{\Phi}_{n\kappa}^{-1}(\hat{\Sigma}_{n\kappa} - \Sigma_\kappa)\hat{\Phi}_{n\kappa}^{-1}A'\| \\ & \quad + \|A\hat{\Phi}_{n\kappa}^{-1}(\Phi_\kappa - \hat{\Phi}_{n\kappa})\Phi_\kappa^{-1}\Sigma_\kappa\hat{\Phi}_{n\kappa}^{-1}A'\| + \|A\Phi_\kappa^{-1}\Sigma_\kappa\hat{\Phi}_{n\kappa}^{-1}(\Phi_\kappa - \hat{\Phi}_{n\kappa})\Phi_\kappa^{-1}A'\| \\ & \leq \|A\hat{\Phi}_{n\kappa}^{-1}\|^2\|\hat{\Sigma}_{n\kappa} - \Sigma_\kappa\| + \|A\hat{\Phi}_{n\kappa}^{-1}\|^2\|(\Phi_\kappa - \hat{\Phi}_{n\kappa})\Phi_\kappa^{-1}\Sigma_\kappa\| + \|A\Phi_\kappa^{-1}\|^2\|\Sigma_\kappa\hat{\Phi}_{n\kappa}^{-1}(\Phi_\kappa - \hat{\Phi}_{n\kappa})\| \\ & \leq \|A\hat{\Phi}_{n\kappa}^{-1}\|^2\|\hat{\Sigma}_{n\kappa} - \Sigma_\kappa\| + \|A\hat{\Phi}_{n\kappa}^{-1}\|^2\|\Phi_\kappa - \hat{\Phi}_{n\kappa}\|\lambda_{\max}(\Phi_\kappa^{-1})\lambda_{\max}(\Sigma_\kappa) \\ & \quad + \|A\Phi_\kappa^{-1}\|^2\lambda_{\max}(\Sigma_\kappa)\lambda_{\max}(\hat{\Phi}_{n\kappa}^{-1})\|\Phi_\kappa - \hat{\Phi}_{n\kappa}\| \\ & = o_p(1).\end{aligned}$$

Similarly, by Assumption 3.15, Lemma A.11,  $\|A\hat{\Phi}_{n\kappa}^{-1}\hat{\Gamma}_{n\kappa}\| = O_p(1)$ , and  $\lambda_{\max}(\Gamma_\kappa\Sigma_{\mu,\pi}\Gamma'_\kappa)$  bounded,

$$\begin{aligned}\|\hat{\Omega}_{n\kappa 2} - \Omega_{\kappa 2}\| & \leq \|A\hat{\Phi}_{n\kappa}^{-1}\hat{\Gamma}_{n\kappa}(\hat{\Sigma}_{\mu,\pi} - \Sigma_{\mu,\pi})\hat{\Gamma}'_{n\kappa}\hat{\Phi}_{n\kappa}^{-1}A'\| \\ & \quad + \|A\hat{\Phi}_{n\kappa}^{-1}(\hat{\Gamma}_{n\kappa} - \Gamma_\kappa)\Sigma_{\mu,\pi}\hat{\Gamma}'_{n\kappa}\hat{\Phi}_{n\kappa}^{-1}A'\| + \|A\hat{\Phi}_{n\kappa}^{-1}\Gamma_\kappa\Sigma_{\mu,\pi}(\hat{\Gamma}'_{n\kappa} - \Gamma'_\kappa)\hat{\Phi}_{n\kappa}^{-1}A'\| \\ & \quad + \|A(\hat{\Phi}_{n\kappa}^{-1}\Gamma_\kappa\Sigma_{\mu,\pi}\Gamma'_\kappa\hat{\Phi}_{n\kappa}^{-1} - \Phi_\kappa^{-1}\Gamma_\kappa\Sigma_{\mu,\pi}\Gamma'_\kappa\Phi_\kappa^{-1})A'\|\end{aligned}$$

$$\begin{aligned}
&\leq \left\| A\hat{\Phi}_{n\kappa}^{-1}\hat{\Gamma}_{n\kappa} \right\|^2 \left\| \hat{\Sigma}_{\mu,\pi} - \Sigma_{\mu,\pi} \right\| + 2 \left\| A\hat{\Phi}_{n\kappa}^{-1} \right\| \left\| \hat{\Gamma}_{n\kappa} - \Gamma_{\kappa} \right\| \lambda_{\max}(\Sigma_{\mu,\pi}) \left\| A\hat{\Phi}_{n\kappa}^{-1}\hat{\Gamma}_{n\kappa} \right\| \\
&\quad + \left\| A\hat{\Phi}_{n\kappa}^{-1} \right\|^2 \left\| \Phi_{\kappa} - \hat{\Phi}_{n\kappa} \right\| \lambda_{\max}(\Phi_{\kappa}^{-1}) \lambda_{\max}(\Gamma_{\kappa} \Sigma_{\mu,\pi} \Gamma'_{\kappa}) \\
&\quad + \left\| A\Phi_{\kappa}^{-1} \right\|^2 \lambda_{\max}(\Gamma_{\kappa} \Sigma_{\mu,\pi} \Gamma'_{\kappa}) \lambda_{\max}(\hat{\Phi}_{n\kappa}^{-1}) \left\| \Phi_{\kappa} - \hat{\Phi}_{n\kappa} \right\| \\
&= o_p(1).
\end{aligned}$$

Now the theorem follows by the triangle inequality.  $\square$

## References

- Abadie, A., J. D. Angrist, and G. W. Imbens (2002): Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings, *Econometrica*, 70, 91-117.
- Amemiya, T. (1982): Two stage least absolute deviations estimators, *Econometrica*, 50, 689-711.
- Angrist, J. D., K. Graddy, and G. W. Imbens (2000): The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish, *Review of Economic Studies*, 67, 499-527.
- Angrist, J. D., G. W. Imbens, and A. B. Krueger (1999): Jackknife instrumental variables estimation, *Journal of Applied Econometrics*, 14, 57-67.
- Angrist, J. D. and A. B. Krueger (1991): Does compulsory schooling attendance affect schooling and earnings, *Quarterly Journal of Economics*, 106, 979-1014.
- Blundell, R. and J. L. Powell (2003a): Endogeneity in nonparametric and semiparametric regression models, in L. Hansen (ed.), *Advances in Econometrics, Eighth World Congress*, Cambridge University Press.
- Blundell, R. and J. L. Powell (2003b): Endogeneity in semiparametric binary regression models, *Review of Economic Studies*, forthcoming.
- Bound, J., D. Jaeger, and R. Baker (1995): Problems with instrumental variables estimation when the correlation between instruments and the endogenous explanatory variable is weak, *Journal of the American Statistical Association*, 61, 345-356.



- Buchinsky, M. (1994): Changes in the U.S. wage structure 1963-1987: application of quantile regression, *Econometrica*, 62, 405-458.
- Buchinsky, M. (1998a): The dynamics of changes in the female wage distribution in the USA: a quantile regression approach, *Journal of Applied Econometrics*, 13, 1-30.
- Buchinsky, M. (1998b): Recent advances in quantile regression models: a practical guide for empirical research, *Journal of Human Resources*, 33, 88-126.
- Card, D. (2001): Estimating the return to schooling: progress on some persistent econometric problems, *Econometrica*, 69, 1127-1160.
- Chen, X., O. Linton, and I. Van Keilegom (2003): Estimation of semiparametric models when the criterion function is not smooth, *Econometrica*, 71, 1591-1608.
- Chaudhuri, P. (1991): Nonparametric estimation of regression quantiles and their local Bahadur representation, *Annals of Statistics*, 19, 760-777.
- Chernozhukov, V. and C. Hansen (2001): An IV model of quantile treatment effects, MIT Department of Economics Working Paper Series, No. 02-06.
- Chesher, A. D. (2003): Identification in nonseparable models, *Econometrica*, 71, 1405-1441.
- Graddy, K. (1995): Testing for imperfect competition at the Fulton Fish Market, *Rand Journal of Economics*, 26, 75-92.
- Heckman, J. and E. Vytlacil (1998): Instrumental variables methods for the correlated random coefficient model: estimating the average rate of return to schooling when the return is correlated with schooling, *Journal of Human Resources*, 33, 974-987.
- Hong, H. and E. Tamer (2003): Inference in censored models with endogenous regressors, *Econometrica*, 71, 905-932.
- Honoré, B.E. and L. Hu (2003): On the performance of some robust instrumental variables estimators, *Journal of Business and Economic Statistics*, forthcoming.
- Imbens, G. W. and W. K. Newey (2003): Identification and estimation of triangular simultaneous equations models without additivity, MIT working paper.

- Koenker, R. and G. Bassett (1978): Regression quantiles, *Econometrica*, 46, 33-50.
- Ma, L. and R. Koenker (2003): Quantile regression methods for recursive structural equation models, University of Illinois at Urbana-Champaign working paper.
- Newey, W. K. (1997): Convergence rates and asymptotic normality for series estimators, *Journal of Econometrics*, 79, 147-168.
- Newey, W. K., J. L. Powell, and F. Vella (1999): Nonparametric estimation of triangular simultaneous equations models, *Econometrica*, 67, 565-603.
- Powell, J. L. (1983): The asymptotic normality of two-stage least absolute deviations estimators, *Econometrica*, 51, 1569-1576.
- Powell, J. L. (1984): Least absolute deviation estimation for the censored regression model, *Journal of Econometrics*, 25, 303-25.
- Powell, J.L. (1986): Censored regression quantiles, *Journal of Econometrics*, 32, 143-55.
- Staiger, D. and J. H. Stock (1997): Instrumental variables regression with weak instruments, *Econometrica*, 65, 557-586.
- van der Vaart, A. W. and J. A. Wellner (1996), *Weak Convergence and Empirical Processes*, New-York: Springer-Verlag.
- Welsh, A. H. (1989): On M-processes and M-estimation, *Annals of Statistics*, 17, 337-361.
- Wooldridge, J. M. (1997): On two-stage least squares estimation of the average treatment effect in a random coefficients model, *Economics Letters*, 56, 129-133.

Table 1. Results of Monte Carlo Experiments  
(Results for  $\beta$ )

Estimator		Bias	Std Dev	RMSE		Bias	Std Dev	RMSE
QR	$n = 100$	0.873	0.232	0.903	$n = 900$	0.851	0.074	0.854
TS(1)		0.098	0.344	0.357		0.137	0.103	0.171
TS(2)		-0.018	0.407	0.408		0.006	0.122	0.123
TS(3)		0.008	0.372	0.372		0.027	0.114	0.118
TS(4)		-0.014	0.390	0.390		0.018	0.116	0.118
TS(5)		-0.017	0.382	0.383		0.028	0.107	0.111
TS(6)		-0.023	0.391	0.392		0.006	0.114	0.114
TS(7)		-0.027	0.388	0.389		0.013	0.107	0.107
TS(8)		-0.028	0.389	0.390		0.002	0.111	0.111
FV		0.561	0.348	0.661		0.501	0.112	0.513
QR	$n = 400$	0.854	0.110	0.861	$n = 1600$	0.850	0.054	0.852
TS(1)		0.129	0.158	0.204		0.143	0.073	0.161
TS(2)		0.004	0.186	0.186		0.016	0.090	0.091
TS(3)		0.026	0.175	0.177		0.036	0.084	0.091
TS(4)		0.016	0.179	0.179		0.026	0.086	0.090
TS(5)		0.020	0.167	0.168		0.037	0.078	0.087
TS(6)		0.005	0.174	0.174		0.015	0.085	0.086
TS(7)		0.006	0.168	0.168		0.020	0.079	0.081
TS(8)		-0.004	0.171	0.171		0.010	0.083	0.083
FV		0.513	0.172	0.541		0.503	0.084	0.510

Note: Entries are the mean bias, standard deviation, and root mean squared error for the linear quantile regression estimator (QR), the two-step estimator with a  $k$ -th order polynomial approximation (TS( $k$ )), and fitted value estimator (FV).

Table 2. Results of Monte Carlo Experiments  
(Results for  $\gamma$ )

Estimator		Bias	Std Dev	RMSE		Bias	Std Dev	RMSE
QR	$n = 100$	-2.618	0.751	2.724	$n = 900$	-2.547	0.235	2.557
TS(1)		-0.287	1.081	1.119		-0.409	0.322	0.520
TS(2)		0.064	1.273	1.274		-0.017	0.382	0.382
TS(3)		-0.024	1.165	1.165		-0.080	0.356	0.365
TS(4)		0.047	1.222	1.223		-0.053	0.362	0.366
TS(5)		0.056	1.199	1.200		-0.084	0.334	0.344
TS(6)		0.076	1.222	1.225		-0.017	0.355	0.355
TS(7)		0.085	1.216	1.219		-0.039	0.331	0.333
TS(8)		0.086	1.215	1.218		-0.005	0.345	0.345
FV		-1.698	1.071	2.008		-1.523	0.351	1.563
QR	$n = 400$	-2.558	0.354	2.583	$n = 1600$	-2.549	0.170	2.555
TS(1)		-0.385	0.498	0.630		-0.428	0.233	0.488
TS(2)		-0.009	0.582	0.582		-0.049	0.285	0.289
TS(3)		-0.078	0.544	0.550		-0.109	0.266	0.288
TS(4)		-0.045	0.560	0.562		-0.080	0.272	0.283
TS(5)		-0.061	0.517	0.521		-0.112	0.247	0.272
TS(6)		-0.017	0.542	0.542		-0.047	0.266	0.270
TS(7)		-0.018	0.522	0.522		-0.062	0.247	0.255
TS(8)		0.011	0.531	0.531		-0.031	0.259	0.260
FV		-1.562	0.532	1.650		-1.532	0.258	1.554

Note: Entries are the mean bias, standard deviation, and root mean squared error for the linear quantile regression estimator (QR), the two-step estimator with a  $k$ -th order polynomial approximation (TS(k)), and fitted value estimator (FV).

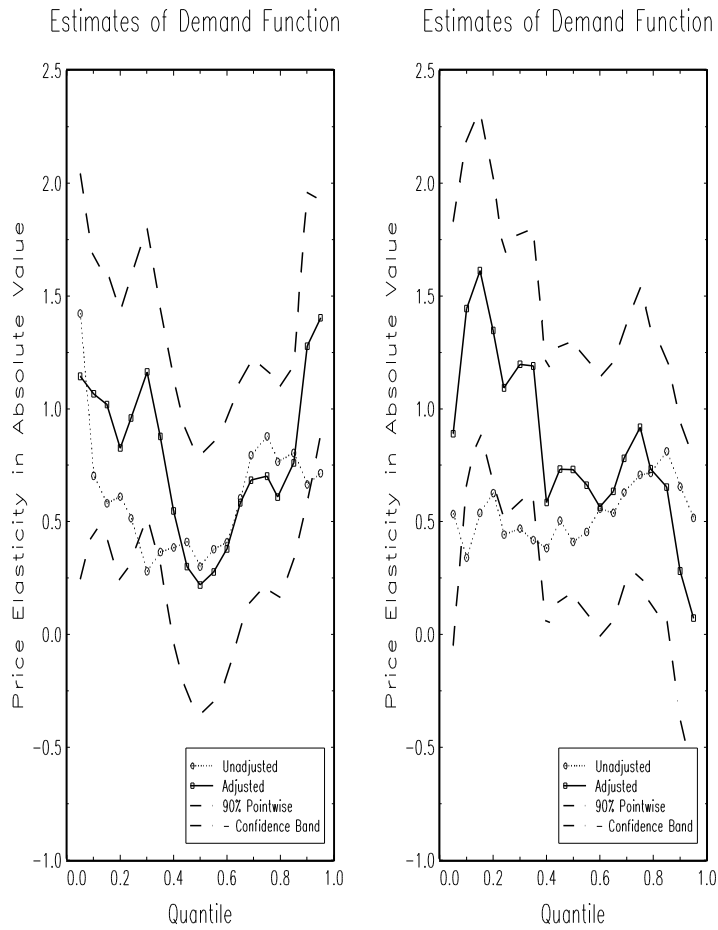


Figure 1. Estimation results for demand for fish.

Note: The left-hand panel shows estimation results for the model with covariates using stormy and mixed as instruments, and the right-hand panel shows those for the model without covariates using wind speed as the instrument.

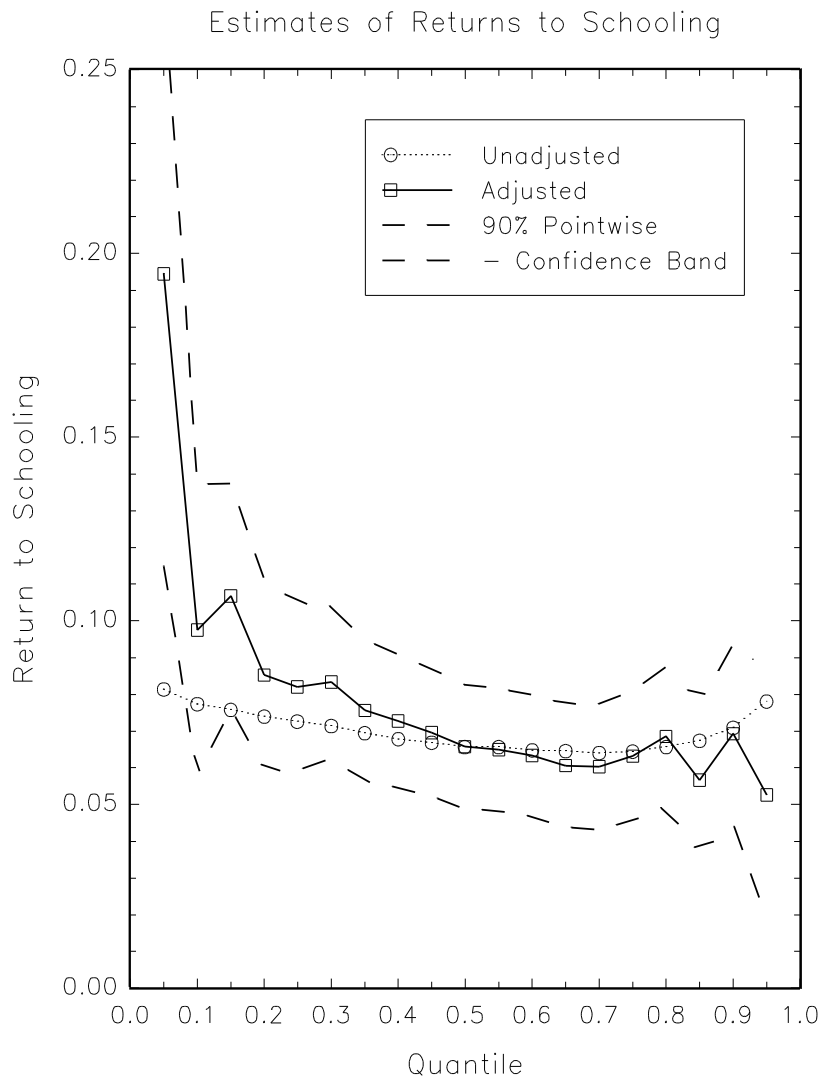


Figure 2. Estimation results for returns to schooling using quarter of birth as instrument.