Supplemental Material
for
Incorporating Covariates in the Measurement of Welfare and Inequality: Methods and Applications

Stephen G. Donald
Department of Economics
University of Texas at Austin

Yu-Chin Hsu
Department of Economics
University of Missouri at Columbia

Garry F. Barrett
Discipline of Economics
University of Sydney

March 15, 2012

This supplement contains two appendices. Summplemental Appendix A addresses some computational issues. Supplemental Appendix B gives proofs of the results in the main paper.

Appendix A: Computational Issues

A.1 Distributions and Partial Integrals

Here we describe the computational properties of the conditional CDF based on the semiparametric method as well as the single index and nonparametric methods. In the case of the semiparametric method, where for sake of generality we allow for the weighted version, we can write the estimator as

$$\begin{aligned}\hat{F}(y|x) &= \frac{1}{N}\sum_{i=1}^{N} 1\Big(\hat{\varepsilon}_i \leq \frac{\log y - x'\hat{\theta}}{h(x'\hat{\theta})}\Big) \\ &= \frac{1}{N}\sum_{i=1}^{N} 1(h(x'\hat{\theta})\hat{\varepsilon}_i + x'\hat{\theta} \leq \log y) \\ &= \frac{1}{N}\sum_{i=1}^{N} 1\left(\exp\left(h(x'\hat{\theta})\hat{\varepsilon}_i + x'\hat{\theta}\right) \leq y\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} 1\left(y_i(x,\hat{\theta}) \leq y\right).\end{aligned}$$

Using the same arguments as Davidson and Duclos (2000), we can show that

$$\mathcal{I}_J(y;\hat{F}(y|x)) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{(J-1)!}(y - y_i(x,\hat{\theta}))^{J-1}1(y_i(x,\hat{\theta}) \leq y).$$

This follows because the conditional distribution estimate, like the unconditional empirical distribution is a step function, with jumps occurring at the values $y_i(x,\hat{\theta})$ where the latter depend on the value x as well as the parameter estimate $\hat{\theta}$.

For the single index model and the non-parametric model we may represent the estimator in each case as

$$\hat{F}(y|x) = \sum_{i=1}^{N}\hat{w}_{ih}1(Y_i \leq y),$$

where in the case of the single index model $m = 1$ and the weights are

$$\hat{w}_{ih} = \frac{K\left(\frac{\hat{v}_1 - \widehat{V}_i}{h}\right)}{\sum_{i=1}^{N} K\left(\frac{\hat{v}_1 - \widehat{V}_i}{h}\right)},$$

while for the nonparametric method $m = p$ and

$$\hat{w}_{ih} = \frac{K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^{N} K\left(\frac{x-X_i}{h}\right)}.$$

Noting the similarity in the dependence of this object on y as for the empirical distribution and the semi-parametric method we have that

$$\mathcal{I}_J(y; \hat{F}(y|x)) = \sum_{i=1}^{N} \hat{w}_{ih} \frac{1}{(J-1)!}(y - Y_i)^{J-1} 1(Y_i \leq y).$$

A.2 Computing Lorenz Curves

The computation of conditional Lorenz curves for the semiparametric, single index and non-parametric conditional distributions is straightforward because in each case the conditional distributions are step functions. This implies that the conditional quantiles will also be step functions and then Lorenz curves will be piecewise linear convex functions. For the semiparametric method described in Section 2.2 denote the distinct values of $y_i(x, \hat{\theta}) = \exp(\hat{\varepsilon}_i h(x'\hat{\theta}) + x'\hat{\theta})$ by the values

$$0 < y_1 < y_2 ... < y_{N^*} < y_u,$$

where $N^* \leq N$ (and will equal N when the residuals are all distinct). Also, let

$$\hat{\pi}_j(x) = \frac{1}{N} \sum_{i=1}^{N} 1(y_i(x, \hat{\theta}) = y_j)$$

be the proportion of observations whose $y_i(x, \hat{\theta})$ value equals y_j. Then the conditional CDF can be written as

$$\hat{F}(y|x) = \sum_{j: y_j \leq y} \hat{\pi}_j(x) = \begin{cases} 0 \text{ for } y \in [0, y_1), \\ \hat{\pi}_1(x) \text{ for } y \in [y_1, y_2), \\ \hat{\pi}_1(x) + \hat{\pi}_2(x) \text{ for } y \in [y_2, y_3), \\ \vdots \\ 1 \text{ for } y \geq y_{N^*}. \end{cases}$$

Then the conditional quantile function can be simply obtained as

$$\hat{q}(p|x) = \begin{cases} y_1 \text{ for } p \in [0, \hat{F}(y_1|x)], \\ y_2 \text{ for } p \in (\hat{F}(y_1|x), \hat{F}(y_2|x)], \\ y_3 \text{ for } p \in (\hat{F}(y_2|x), \hat{F}(y_3|x)], \\ \vdots \\ y_{N^*} \text{ for } p \in (\hat{F}(y_{N^*-1}|x), 1]. \end{cases}$$

This can be written compactly as

$$\hat{q}(p|x) = \sum_{j=1}^{N^*} 1(p \in (\hat{F}(y_{j-1}|x), \hat{F}(y_j|x)])y_j,$$

where we define $\hat{F}(y_0|x) = 0$ and $\hat{F}(y_{N^*}|x) = 1$. Note that given the definition of $y_i(x, \hat{\theta})$ we can write $\hat{q}(p|x) = \exp(\hat{q}_\varepsilon(p)h(x'\hat{\theta}) + x'\hat{\theta})$ where $\hat{q}_\varepsilon(p)$ is the pth quantile of the residuals. The Generalized Lorenz curve can then be calculated by integrating the step function defined by $\hat{q}(p|x)$ to get for any $p \in [\hat{F}(y_{j-1}|x), \hat{F}(y_j|x))$,

$$\hat{G}^j(s|x) = (p - \hat{F}(y_{j-1}|x))y_j + \sum_{l=1}^{j-1} \hat{\pi}_l(x)y_l$$

so that for arbitrary $p \in [0, 1]$

$$\hat{G}(p|x) = \sum_{j=1}^{\dot{N}} 1(\hat{F}(y_{j-1}|x) \leq p < \hat{F}(y_j|x))\hat{G}^j(p|x)$$

and

$$\hat{G}(1|x) = \sum_{l=1}^{N^*} \hat{\pi}_l(x)y_l.$$

Note that this is simply

$$\frac{1}{N}\sum_{i=1}^{N} y_i(x, \hat{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \exp(\hat{\varepsilon}_i h(x'\hat{\theta}) + x'\hat{\theta}) = \exp(x'\hat{\theta})\frac{1}{N}\sum_{i=1}^{N} \exp(\hat{\varepsilon}_i h(x'\hat{\theta})),$$

which is the natural estimator of the conditional mean of Y given $X = x$ based on a log-linear model relating Y to X. Also, note that the GLC is a piecewise linear and continuous function. The LC can be calculated as the integral of the quantile process and is given by

$$\hat{L}(p|x) = \frac{\hat{G}(p|x)}{\hat{G}(1|x)}$$

and is also a piecewise linear function. As noted earlier, in the unweighted case where $h(.) = 1$ then the Lorenz curve does not depend on x.

The GLC and LC for the single index model and the nonparametric model can be computed in an analogous fashion. For each method denote the (ordered) distinct sample values for the Y_i by y_j (which will be different from the values used in the computation of the GLC and LC for the semiparametric method) so that

$$y_l \leq y_1 < y_2 < \ldots < y_{\dot{N}} \leq y_u,$$

where $\dot{N} \leq N$ is the number of distinct values in the sample. Then given this notation, the function $\hat{F}(y|x)$ is a step function with increments,

$$\hat{\pi}_j(x) = \sum_{i=1}^{N} \hat{w}_{i,h} 1(Y_i = y_j),$$

occurring at each of the y_j values where $\hat{w}_{i,h}$ was defined for each estimator in the previous section. Therefore, as was the case for the semiparametric method the estimator $\hat{F}(y|x)$ takes on the value $\sum_{l=1}^{j} \hat{\pi}_j(x)$ on the interval $[y_j, y_{j+1})$. The quantile function, the GLC and the LC can then be defined in the same form as the semiparametric method.

A.3 Computing Generalized Gini Coefficients

Since the LC for the three main methods is piecewise linear it is straightforward to used the calculations in Barrett and Donald (2009) for calculating generalized Gini inequality measures. Here, due to space limitations we focus on the S-Gini index of relative inequality which, conditional on x, is defined as

$$\begin{aligned} I_R^{\delta}(x) &= 1 - \delta(\delta - 1) \int_0^1 (1-p)^{\delta-2} L(p|x) d(p) \\ &= 1 - \frac{\delta \int_0^1 (1-p)^{\delta-1} Q(p|x) d(p)}{G(1|x)}, \end{aligned}$$

where δ is chosen by the researcher – $\delta = 2$ gives the standard Gini index of relative inequality. Using analogous calculations to those in Barrett and Donald (2009) one can show that

$$\hat{I}_R^{\delta}(x) = 1 - \delta \frac{\sum_{j=1}^{\dot{N}} y_j \left\{ (1 - \hat{p}_{j-1}))^{\delta} - (1 - \hat{p}_j)^{\delta} \right\}}{\hat{G}(1|x)},$$

where $\sum_{l=1}^{j} \hat{\pi}_j = \hat{p}_j$ and where the $\hat{\pi}_j$ depend on the particular method as defined in the previous subsection. Given the linearity of the index in the LC arguments similar to those in Barrett and Donald (2009) can be used to show that

$$\sqrt{N}(\hat{I}_R^{\delta}(x) - I_R^{\delta}(x)) \xrightarrow{d} -\delta(\delta - 1) \int_0^1 (1-p)^{\delta-2} \mathcal{L}(\cdot) d(p) \sim \mathcal{N}(0, V(x))$$

for some asymptotic variance $V(x)$ which can be computed using the influence functions described in the next section. Indeed, because the influence functions have a similar form to those in the unconditional case (the semi-parametric method being slightly different), very similar calculations can be used.

A.4 Computing Influence Functions for Quantile Process and Lorenz Curves

To compute the influence function for the quantile process for each case we use the appropriate influence function and the general form given in (4.1). For the semi-parametric method, because

$$\hat{f}(\hat{q}(p|x)|x) = \hat{f}_\varepsilon(\log \hat{q}(p|x) - x'\hat{\theta}) = \hat{f}_\varepsilon(\hat{q}_\varepsilon(p)),$$

this becomes

$$\phi_i(p|x;\hat{q}) = \frac{p - 1(Y_i \le \hat{q}(p|x))}{\hat{f}_\varepsilon(\log \hat{q}(p|x) - x'\hat{\theta})} - (\overline{X} - x)'\phi(Y_i, X_i, \hat{\theta}).$$

For the single index model the influence functions can be written as

$$\phi_i(p|x;\hat{q}) = \frac{1}{\sqrt{h}} \frac{K\left(\frac{\hat{v}_1 - \widehat{V}_i}{h}\right)}{\hat{g}(\hat{v}_1, \hat{\theta})} \left(\frac{p - 1(Y_i \le \hat{q}(p|x))}{\hat{f}(\hat{q}(p|x)|x)}\right),$$

and

$$\phi_i(p|x;\hat{q}) = \frac{1}{\sqrt{h^d}} \frac{K\left(\frac{x - X_i}{h}\right)}{\hat{g}(x)} \left(\frac{p - 1(Y_i \le \hat{q}(p|x))}{\hat{f}(\hat{q}(p|x)|x)}\right).$$

These can be used to estimate pointwise variances. Also we can compute influence functions for the GLC and LC's using the forms in (4.2) and (4.3). For the GLC based on the semi-parametric method we have

$$\phi_i(p|x;\hat{G}) = (p\hat{q}(p|x) - \hat{G}(p|x)) - 1(Y_i \le \hat{q}(p|x))(\hat{q}(p|x) - Y_i) - p(\overline{X} - x)'\phi(Y_i, X_i, \hat{\theta}).$$

The first two terms in this expression are the natural extension of the influence function given in Barrett and Donald (2009) to the case where one conditions on x but knows the population value of θ_0. The last term is the effect on the influence function of having to estimate θ_0. For the single index and non-parametric models we obtain, respectively,

$$\phi_i(p|x;\hat{G}) = \frac{1}{\sqrt{h}} \frac{K\left(\frac{\hat{v}_1 - \widehat{V}_i}{h}\right)}{\hat{g}(\hat{v}_1, \hat{\theta})} \Big((p\hat{q}(p|x) - \hat{G}(p|x)) - 1(Y_i \le \hat{q}(p|x))(\hat{q}(p|x) - Y_i)\Big),$$

and

$$\phi_i(p|x;\hat{G}) = \frac{1}{\sqrt{h^d}} \frac{K\left(\frac{x - X_i}{h}\right)}{\hat{g}(x)} \Big((p\hat{q}(p|x) - \hat{G}(p|x)) - 1(Y_i \le \hat{q}(p|x))(\hat{q}(p|x) - Y_i)\Big).$$

Given the definition of quantile processes and $\hat{G}$ these can easily be calculated. Also, the influence functions for LC's can easily be obtained from these using (4.3).

Appendix B: Proofs of Results

Proof of Theorem 3.1: We have

$$\begin{aligned}
&\sqrt{N}(F(\cdot|x,\hat{\theta})-F(\cdot|x,\theta_0))\\
=&\nabla_\theta F(\cdot|x,\theta_0)'\sqrt{N}(\hat{\theta}-\theta_0)+\sqrt{N}(\hat{\theta}-\theta_0)'\nabla_{\theta\theta'}F(\cdot|x,\theta^*)(\hat{\theta}-\theta_0)\\
=&\nabla_\theta F_1(\cdot|x,\theta_0)'\sqrt{N}(\hat{\theta}-\theta_0)+o_p(1)=\nabla_\theta F_1(\cdot|x,\theta_0)'\frac{1}{\sqrt{N}}\sum_{i=1}^N\phi_1(Y_i,X_i,\theta_0)+o_p(1)\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^N\nabla_\theta F(\cdot|x,\theta_0)'\phi(Y_i,X_i,\theta_0)+o_p(1).
\end{aligned}$$

The first equality follows from the second-order mean-value expansion of $F(\cdot|x,\hat{\theta})$ at θ_0 and θ^* is the mean-value. The second equality holds because $\nabla_{\theta\theta'}F$ is bounded over y and $(\hat{\theta}-\theta_0)=O_p(N^{-1/2})$. The third equality follows from Assumption 2.2 i. and the boundedness of $\nabla_\theta F$.

By Kosorok (2008), we have that the class $\{a'\phi(x)\}|a\in\mathbb{R}^p\}$ is a Vapnik-Cervonenkis class and this implies that it is $\mathcal{P}$-Donsker. Given $\nabla_\theta F(y|x,\theta_0)$ is a continuous function of y on $\mathbb{R}^p$, we have

$$\sqrt{N}(F(\cdot|x,\hat{\theta})-F(\cdot|x,\theta_0))\Rightarrow\mathcal{X}(\cdot),$$

which shows Theorem 3.1.□

Proof of Theorem 3.2: The first part follows from the fact that $\widehat{\Omega}\xrightarrow{p}\Omega$. To show the second part, given $\hat{\theta}\xrightarrow{p}\theta_0$ and for fixed x, $\nabla_\theta F$ is continuous on $B(\theta_0)\times\mathcal{Y}$, we have

$$\sup_{y\in\mathcal{Y}}|\nabla_\theta F(y|x,\hat{\theta})-\nabla_\theta F(y|x,\theta)|=o_p(1).$$

By applying Theorem 2.1 of Kosorok (2008) conditional on the sample path, we can show that $\mathcal{X}^u(\cdot)\Rightarrow\mathcal{X}(\cdot)$. □

Proof of Lemma 3.1: We have

$$\begin{aligned}
F_y(y|X=x)&=E[1(Y\le y)|X=x]=E[1(\exp(X'\theta_0+\varepsilon)\le y)|X=x]\\
&=E[1(\exp(x'\theta_0+\varepsilon)\le y)]=E[1(\varepsilon\le\log y-x'\theta_0)]=F_\varepsilon(\log y-x'\theta_0).
\end{aligned}$$

By the definition of the conditional CDF, we have the first equality. By rewriting Y in terms of X and ε, we have the second equality. By independence between X and ε, we get the third equality. The fourth equality follows from the fact that $\exp(x'\theta_0+\varepsilon)\le y$ if and only if $\varepsilon\le\log y-x'\theta_0$. The last equality follows from the definition of the F_ε. □

Proof of Theorem 3.3: By Linton et al. (2005), we have $\sqrt{N}(\widehat{F}_\varepsilon(\cdot)-F_\varepsilon(\cdot))\Rightarrow\mathcal{X}_\varepsilon(\cdot)$ where $\mathcal{X}_\varepsilon(\cdot)$ is a mean zero Gaussian process with covariance kernel generated by $1(\varepsilon\le\cdot)-F_\varepsilon(\cdot)+f_\varepsilon(\cdot)E[X]'\phi(Y,X,\theta_0)$.

For a given x, $\widehat{F}_y(y|x) = \widehat{F}_\varepsilon(\log y - x'\hat{\theta})$ and $F_y(y|x) = F_\varepsilon(\log y - x'\theta_0)$. Hence,

$$\sqrt{N}(\widehat{F}_y(y|x) - F_y(y|x)) = \sqrt{N}(\widehat{F}_\varepsilon(\log y - x'\hat{\theta}) - F_\varepsilon(\log y - x'\theta_0))$$

$$=\sqrt{N}\Big((\widehat{F}_\varepsilon(\log y - x'\hat{\theta}) - F_\varepsilon(\log y - x'\hat{\theta}))$$

$$- (\widehat{F}_\varepsilon(\log y - x'\theta_0) - F_\varepsilon(\log y - x'\theta_0))\Big) \quad (1)$$

$$+ \sqrt{N}(\widehat{F}_\varepsilon(\log y - x'\theta_0) - F_\varepsilon(\log y - x'\theta_0) \quad (2)$$

$$+ \sqrt{N}(F_\varepsilon(\log y - x'\hat{\theta}) - F_\varepsilon(\log y - x'\theta_0)). \quad (3)$$

For (1), it will converge to a zero process. For (2), it is the process of $\sqrt{N}(\widehat{F}_\varepsilon(\cdot) - F_\varepsilon(\cdot))$ with a transformation such that $\varepsilon = \log y - x'\theta_0$. Hence, it will converge to a mean zero Gaussian process with covariance kernel generated by $1(\varepsilon \le \log y - x'\theta_0) - F_\varepsilon(\log y - x'\theta_0) + f_\varepsilon(\log y - x'\theta_0)E[X]'\phi(Y, X, \theta_0)$. For (3), we have

$$\sqrt{N}(F_\varepsilon(\log y - x'\hat{\theta}) - F_\varepsilon(\log y - x'\theta_0))$$

$$= - f_\varepsilon(\log y - x'\theta_0))x'\sqrt{N}(\hat{\theta} - \theta_0) + \sqrt{N}f'(\log y - x'\theta_0^*)(\hat{\theta} - \theta_0)'xx'(\hat{\theta} - \theta_0)$$

$$= - \frac{1}{\sqrt{N}}\sum_{i=1}^{N} f_\varepsilon(\log y - x'\theta_0)x'\phi(Y_i, X_i, \theta_0) + o_p(1).$$

Given f_ε is continuous and bounded, we have $\sqrt{N}(F_\varepsilon(\log y - x'\hat{\theta}) - F_\varepsilon(\log y - x'\theta_0))$ converge to a mean zero Gaussian process with the covariance kernel generated by $-f_\varepsilon(\log y - x'\theta_0)x'\phi(Y, X, \theta_0)$. These complete the proof of Theorem 3. □

Proof of Theorem 3.4: We divide the simulated process into two parts:

$$\mathcal{X}_1(y) = \sum_{i=1}^{N} \frac{U_i}{\sqrt{N}} \left(1(\hat{\varepsilon}_i \le \log y_1 - x'\hat{\theta}) - \hat{F}_\varepsilon(\log y_1 - x'\hat{\theta})\right),$$

$$\mathcal{X}_2(y) = \sum_{i=1}^{N} \frac{U_i}{\sqrt{N}} \left(\hat{f}_\varepsilon(\log y_1 - x'\hat{\theta})(\overline{X} - x)'\phi_1(Y_i, X_i, \hat{\theta})\right).$$

By similar argument in Donald and Hsu (2011), $\mathcal{X}_1(y)$ will converge to a mean zero Gaussian process with covariance kernel generated by $1(\varepsilon \le \log y_1 - x'\theta_0)$ conditional on sample path with probability 1. By similar argument for Theorem 2, $\mathcal{X}_2(y)$ will converge to a mean zero Gaussian process with covariance kernel generated by $f_\varepsilon(\log y_1 - x'\theta_0)(E[X] - x)'\phi(Y, X, \theta_0)$ conditional on sample path with probability 1. These complete the proof of Theorem 3.4. □

Proof of Lemma 3.2: First, for $e \in [e_L, e_H]$, the estimator for the $f_\varepsilon(e)$ based on the true ε_i is

$$\bar{f}_\varepsilon(e) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{\varepsilon_i - e}{h}\right).$$

For some $M > 0$, we have

$$\begin{aligned}
|\tilde{f}_\varepsilon(e) - \bar{f}_\varepsilon(e)| =& \left| \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{\hat{\varepsilon}_i - e}{h}\right) - K\left(\frac{\varepsilon_i - e}{h}\right) \right| \\
\leq& \frac{1}{Nh} \sum_{i=1}^{N} \left| K\left(\frac{\hat{\varepsilon}_i - e}{h}\right) - K\left(\frac{\varepsilon_i - e}{h}\right) \right| \leq \frac{M}{Nh} \sum_{i=1}^{N} \frac{|X_i'(\hat{\theta} - \theta_0)|}{h} \\
\leq& \frac{M}{\sqrt{N}h^2} \frac{1}{N} \sum_{i=1}^{N} |\sqrt{N}(\hat{\theta} - \theta_0)' X_i| \leq \frac{M}{\sqrt{N}h^2} O_p(1) = o_p(1). \qquad (4)
\end{aligned}$$

The second inequality holds because $K(u)$ is Lipschitz continuous by assumptions. The fourth inequality holds because X_i is bounded and $\sqrt{N}(\hat{\theta}-\theta_0) = O_p(1)$. The $o_p(1)$ result holds because $\sqrt{N}h^2 \to 0$. Note that the bound in (4) does not depend on e, so $\sup_{e\in[e_L,e_H]} |\tilde{f}_\varepsilon(e) - \bar{f}_\varepsilon(e)| = o_p(1)$. Define $S_N = [e_L + h, e_H - h]$. First, we claim that

$$\sup_{e\in S_N} |\bar{f}_\varepsilon(e) - f_\varepsilon(e)| = O_p\Big(h^2 + \sqrt{\frac{\log N}{Nh}}\Big). \qquad (5)$$

Masry (1996) shows that (5) holds for any fixed compact set S that is a subset of the interior point of $[e_L, e_H]$. Here, we extend his result to cases where the S can vary with N. Note that for some $\ddot{e} \in [e - h, e + h]$,

$$E[\bar{f}_\varepsilon(e)] - f_\varepsilon(e) = h^2 f_\varepsilon''(\ddot{e}) \int_u u^2 K(u) du = O(h^2). \qquad (6)$$

Because f_ε'' is bounded, (6) holds uniformly for all $e \in S_N$. Hence,

$$\sup_{e\in S_N} \big|E[\bar{f}_\varepsilon(e)] - f_\varepsilon(e)\big| = O(h^2). \qquad (7)$$

Further, because S_N is compact, it can be covered by a finite number (L_N) of closed intervals $I_{k,N}$ with center $e_{k,N}$ and length ℓ_N such that $L_N = O(\ell_N^{-1})$. Note that

$$\begin{aligned}
&\sup_{e\in S_N} \big|\bar{f}_\varepsilon(e) - E[\bar{f}_\varepsilon(e)]\big| \\
\leq& \max_{1\leq k\leq L_N} \sup_{e\in S_N\cap I_{k,N}} \big|\bar{f}_\varepsilon(e) - \bar{f}_\varepsilon(e_{k,N})\big| + \max_{1\leq k\leq L_N} \big|\bar{f}_\varepsilon(e_{k,N}) - E[\bar{f}_\varepsilon(e_{k,N})\big| \\
&+ \max_{1\leq k\leq L_N} \sup_{e\in S_N\cap I_{k,N}} \big|E[\bar{f}_\varepsilon(e)] - E[\bar{f}_\varepsilon(e_{k,N})]\big| \equiv J_1 + J_2 + J_3.
\end{aligned}$$

Note that for some $M > 0$ and for any e_1 and e_2,

$$\begin{aligned}
&\big|\bar{f}_\varepsilon(e_1) - \bar{f}_\varepsilon(e_2)\big| = \sup_{e\in S_N\cap I_{k,N}} \Big| \sum_{i=1}^{N} K\Big(\frac{\varepsilon_i - e_1}{h}\Big) - K\Big(\frac{\varepsilon_i - e_2}{h}\Big) \Big| \\
\leq& \frac{1}{Nh} \sum_{i=1}^{N} M \Big|\frac{e_1 - e_2}{h}\Big| \leq \frac{1}{h^2} M |e_1 - e_2|,
\end{aligned}$$

where the first inequality follows from that $K(u)$ is Lipschitz continuous. Therefore,

$$\begin{aligned} J_1 &= \max_{1\le k\le L_N} \sup_{e\in S_N\cap I_{k,N}} \left|\bar{f}_\varepsilon(e) - \bar{f}_\varepsilon(e_{k,N})\right| \\ &\le \frac{1}{h^2} M \max_{1\le k\le L_N} \sup_{e\in S_N\cap I_{k,N}} |e - e_{k,N}| \le M h^{-2}\ell_N. \end{aligned} \tag{8}$$

Similarly, we have

$$J_3 = \max_{1\le k\le L_N} \sup_{e\in S_N\cap I_{k,N}} \left|E[\bar{f}_\varepsilon(e)] - E[\bar{f}_\varepsilon(e_{k,N})]\right| \le M h^{-2}\ell_N.$$

Now, we consider the J_2 term. Define

$$\begin{aligned} W_N(e) &= \bar{f}_\varepsilon(e) - E[\bar{f}_\varepsilon(e)] \equiv \sum_{i=1}^N Z_{N,i}(e), \\ Z_{N,i}(e) &= \frac{1}{Nh}\Big(K\Big(\frac{\varepsilon_i - e}{h}\Big) - E\Big[K\Big(\frac{\varepsilon_i - e}{h}\Big)\Big]\Big). \end{aligned}$$

For any $\eta > 0$. we have

$$\begin{aligned} P[J_2 > \eta] &= P\Big[\max_{1\le k\le L_N} |W_N(e_{k,N})| > \eta\Big] \\ &= P\big[|W_N(e_{k,N})| > \eta, \dots, \text{ or } |W_N(e_{L_N,N})| > \eta\big] \\ &\le \sum_{k=1}^{L_N} P\big[|W_n(e_{k,N})| > \eta\big] \le L_N \sup_{1\le k\le L_N} P\big[|W_n(e_{k,N})| > \eta\big]. \end{aligned}$$

Define $\lambda_N = \sqrt{Nh\log N}$, then it is true that for N sufficiently large, $\lambda_N|Z_{N,i}(e)| \le 1/2$ for all $i = 1,\dots,N$. Since $\exp(x) \le 1 + x + x^2$ for $|x| \le 1/2$, we have $\exp(\lambda_N Z_{N,i}(e)) \le 1 + \lambda_N Z_{N,i}(e) + \lambda_N^2 Z_{N,i}^2(e)$. Consequently, when N is large enough

$$\begin{aligned} E\big[\exp(\lambda_N Z_{N,i}(e))\big] &\le 1 + E\big[\lambda_N Z_{N,i}(e)\big] + E\big[\lambda_N^2 Z_{N,i}^2(e)\big] \\ &= 1 + E\big[\lambda_N^2 Z_{N,i}^2(e)\big] \le \exp\big(E\big[\lambda_n^2 Z_{N,i}^2(e)\big]\big), \end{aligned} \tag{9}$$

where the first equality holds because $E[Z_{N,i}(e)] = 0$ and the second equality holds because $\exp(x) \ge 1 + x$ for all x. Also, for some $M_1 > 0$

$$\begin{aligned} P\big[|W_N(e)| > \eta\big] &= P\Big[\big|\sum_{i=1}^N Z_{N,i}\big| > \eta\Big] = P\Big[\sum_{i=1}^N Z_{N,i}(e) > \eta\Big] + P\Big[-\sum_{i=1}^N Z_{N,i} > \eta\Big] \\ &\le \frac{E\big[\exp(\lambda_N \sum_{i=1}^N Z_{N,i}(e))\big]}{\exp(\lambda_N\eta)} + \frac{E\big[\exp(-\lambda_N \sum_{i=1}^N Z_{N,i}(e))\big]}{\exp(\lambda_N\eta)} \\ &\le 2\exp(-\lambda_N\eta)\exp\Big(\lambda_N^2 \sum_{i=1}^N E[Z_{N,i}^2(e)]\Big) \\ &\le 2\exp(-\lambda_N\eta)\exp\Big(\frac{M_1\lambda_N^2}{Nh}\Big), \end{aligned} \tag{10}$$

where the second line follows the Markov inequality such that $P(Z > c) \leq E[\exp(aZ)]/\exp(ac)$ for a random variable Z and any positive numbers a and c. The third line follows from (9) and the fact that $Z_{N,i}(e)$ and $Z_{N,j}(e)$ are independent for any $i \neq j$. The last line holds because for some $M_1 > 0$,

$$\begin{aligned} E[Z_{N,i}^2(e)] &= V\Big(\frac{1}{Nh}K\Big(\frac{\varepsilon_i - e}{h}\Big)\Big) \\ &\leq \frac{1}{N^2h^2}E\Big[K^2\Big(\frac{\varepsilon_i - e}{h}\Big)\Big] = \frac{1}{N^2h}\int_u K^2(u)f(e+hu)du \\ &\leq \frac{1}{N^2h}\Big(f_\varepsilon(e)\int_u K^2(u)du + f_\varepsilon''(\ddot{e})\int_u u^2K^2(u)du\Big) \leq \frac{M_1}{N^2h}, \end{aligned} \tag{11}$$

where the first inequality follows that $V(Z) \leq E[Z^2]$ for any random variable and the last inequality follows that f_ε and f_ε'' are uniformly bounded. Given that M_1 does not depend on e, we have

$$\sup_{e \in S_N} P\big[|W_N(e)| > \eta\big] \leq 2\exp(-\lambda_N\eta)\exp\Big(\frac{M\lambda_N^2}{Nh}\Big),$$

which implies that

$$P[J_2 > \eta] = P\Big[\max_{1\leq k\leq L_N}|W_N(e_{k,N})| > \eta\Big] \leq 2L_N\exp(-\lambda_N\eta)\exp\Big(\frac{M\lambda_N^2}{Nh}\Big).$$

Let $\ell_N = N^{-2}$ and $\eta_N = M_2\sqrt{N^{-1}h^{-1}\log N}$ where $M_2 - M_1 = 4$. Hence, for some $M > 0$, we have $L_N \leq MN^2$ and

$$P[J_2 > \eta_N] \leq 2L_N\exp\Big(M_2\log N - M_1\log N\Big) \leq 2MN^2N^{-4} = MN^{-2}.$$

Define $A_N \equiv \{\omega : J_2 > \eta_N\}$, then we have

$$\lim_{N\to\infty}\sum_{i=1}^N P[J_2 > \eta_N] = \lim_{N\to\infty}\sum_{i=1}^N P[A_N] \leq \lim_{N\to\infty}\sum_{i=1}^N 2MN^{-2} < \infty.$$

By Borel-Cantelli Lemma, it is true that $P(A_N \text{ infinitely often}) = 0$ or equivalently $P(A_N^c \text{ eventually}) = 1$. This implies that $J_2 = O_{a.s.}(\eta_N)$ and this is sufficient for $J_2 = O_p(\sqrt{\log N/Nh})$. Also, by the choice of ℓ_N, it is true that $J_1 \leq M(Nh)^{-2} = o(\eta_N)$ and similarly, $J_3 = o(\eta_N)$. These imply that

$$\sup_{e\in S_N}\big|\bar{f}_\varepsilon(e) - E[\bar{f}_\varepsilon(e)]\big| = O_p\Big(\sqrt{\frac{\log N}{Nh}}\Big). \tag{12}$$

Hence, (7) and (12) imply that

$$\sup_{e\in S_N}|\bar{f}_\varepsilon(e) - f_\varepsilon(e)| = O_p\Big(h^2 + \sqrt{\frac{\log N}{Nh}}\Big),$$

which is sufficient for $\sup_{e \in S_N} |\bar{f}_\varepsilon(e) - f_\varepsilon(e)| = o_p(1)$. In addition, for all $e \in [e_L, e_L + h]$ and for some $M > 0$,

$$\begin{aligned} |\hat{f}_\varepsilon(e_L + h) - f_\varepsilon(e)| &\leq |\hat{f}_\varepsilon(e_L + h) - f_\varepsilon(e + h)| + |f_\varepsilon(e_L + h) - f_\varepsilon(e)| \\ &\leq |\hat{f}_\varepsilon(e_L + h) - f_\varepsilon(e + h)| + Mh, \end{aligned}$$

which implies that $\sup_{e \in [e_L, e_L + h]} |\bar{f}_\varepsilon(e_L + h) - f_\varepsilon(e)| = o_p(1)$ and $\sup_{e \in [e_H - h, e_H]} |\bar{f}_\varepsilon(e_H - h) - f_\varepsilon(e)| = o_p(1)$. Given the fact that $\sup_{e \in [e_L, e_H]} |\tilde{f}_\varepsilon(e) - \bar{f}_\varepsilon(e)| = o_p(1)$ and by applying several triangular inequalities, we have $\sup_{e \in [e_L, e_H]} |\hat{f}_\varepsilon(e) - f_\varepsilon(e)| = o_p(1)$. □

Proof of Lemma 3.3: We have

$$\begin{aligned} F(y|X = x) &= E[1(Y \leq y)|X = x] = E[1(w(x'\theta_0, \varepsilon) \leq y)] \\ &= E[1(w(X'\theta_0, \varepsilon) \leq y)|X'\theta_0 = x'\theta_0] \\ &= E[1(Y \leq y)|X'\theta_0 = x'\theta_0] = F(y|X'\theta_0 = x'\theta_0). \end{aligned}$$

By the definition of the conditional CDF, the first equality follows. By rewriting Y as $w(x'\theta_0, \varepsilon)$ and by the independence between X and ε, the second equality holds. The third equality holds by the definition of conditional expectation. The last line follows from rewriting $w(x'\theta_0, \varepsilon)$ as Y and from the definition of CDF. □

Proof of Theorem 3.5: Define

$$\bar{r}(y, v) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{v - V_i}{h}\right) 1(Y_i \leq y), \hat{r}(y, v) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{v - \hat{V}_i}{h}\right) 1(Y_i \leq y).$$

Hence,

$$\sqrt{Nh}(\hat{r}(y, \hat{v}) - r(y, v)) = \sqrt{Nh}(\hat{r}(y, \hat{v}) - \bar{r}(y, v)) + \sqrt{Nh}(\bar{r}(y, v) - r(y, v)).$$

Using similar argument in Hall and Yao (2005), we can show that $\sup_y |\sqrt{Nh}(\hat{r}(y, \hat{v}) - \bar{r}(y, v))| = o_p(1)$. This implies that the estimation error from $\hat{\theta}$ will disappear in the limit, so using $\hat{\theta}$ is as good as using θ_0. Second, we claim that $\sqrt{Nh}(\bar{r}(y, v) - r(y, v))$ converges to a mean zero Gaussian process $\mathcal{Z}(\cdot)$ with covariance kernel $Cov(\mathcal{Z}(y_1), \mathcal{Z}(y_2)) = \|K\|_2^2 r(\min\{y_1, y_2\}, v)$ where $\|K\|_2^2 = \int_u K^2(u)du$. Define

$$\begin{aligned} f_{Ni}(y) &= \frac{1}{\sqrt{Nh}} K\left(\frac{v - V_i}{h}\right) 1(Y_i \leq y), F_{Ni} = \frac{1}{\sqrt{Nh}} K\left(\frac{v - V_i}{h}\right), \\ \mathcal{Z}_N(y) &= \sum_{i=1}^{N} f_{Ni}(y) - E[f_{Ni}(y)]. \end{aligned}$$

We first show that $\mathcal{Z}_N(\cdot)$ weakly converges to a Gaussian process $\mathcal{Z}(\cdot)$ by checking that (i)-(v) of Theorem 10.6 of Pollard (1990) hold.

For all $\omega \in \Omega$, define $\overrightarrow{F}_N(\omega) = (F_{N1}(\omega), ..., F_{NN}(\omega))$, $\overrightarrow{f}_N(y,\omega) = ((f_{N1}(\omega), ..., f_{NN}(y,\omega))$, $\mathcal{F}_{N\omega} = \{\overrightarrow{f}_N(y,\omega)|\ y \in [y_l, y_u]\}$, and $A_N = (a_1, ..., a_N) \in \mathbb{R}^N$ be a vector of non-negative weights. Let $\odot$ denote the pointwise product where $A_N \odot \overrightarrow{F}_N(\omega) \equiv (a_1 F_{N1}(\omega), ..., a_N F_{NN}(\omega))$. The packing number $D(\epsilon, T_0)$ for a subset of T_0 of a metric space with metric d is defined as the largest k for which there exist points $t_1, ..., t_k$ in T_0 with $d(t_i, t_j) > \epsilon$ for $i \neq j$. We use the ℓ_1 norm on $\mathbb{R}^N$ which is defined as $|(u_1, ..., u_N)|_1 = \sum_{i=1}^N |u_i|$. Since $D(\epsilon|A_N \odot \overrightarrow{F}_N|_1, A_N \odot \mathcal{F}_{N\omega_u}) = D(\epsilon|\alpha A_N \odot \overrightarrow{F}_N|_1, \alpha A_N \odot \mathcal{F}_{N\omega_u})$ for all $\alpha > 0$, without loss of generality (WLOG) we can re-scale A_N such that $|A_N \odot \overrightarrow{F}_N|_1 = 1$. Let

$$d(y_1, y_2) = |\overrightarrow{f}_N(y_1, \omega) - \overrightarrow{f}_N(y_2, \omega)|_1 = \sum_{i=1}^N |f_{Ni}(y_1, \omega) - f_{Ni}(y_2, \omega)|.$$

Since $f_{Ni}(y, \omega)$ is either monotonically increasing or monotonically increasing in y depending on the sign of $K((v - V_i)/h)$ for $i = 1, ..., N_1$, we have that for any $y_1 \leq y_2 \leq y_3$,

$$\begin{aligned} d(y_3, y_1) &= \sum_{i=1}^N |f_{Ni}(y_3, \omega) - f_{Ni}(y_1, \omega)| \\ &= \sum_{i=1}^N f_{Ni}|(y_3, \omega) - f_{Ni}(y_2, \omega)| + |f_{Ni}(y_2, \omega) - f_{Ni}(y_1, \omega)| \\ &= d(y_3, y_2) + d(y_2, y_1). \end{aligned}$$

We claim that $D(\epsilon, A_N \odot \mathcal{F}_{N\omega}) \leq 1/\epsilon + 1$. Suppose not, then there exists an integer $k \geq 1/\epsilon + 1$ and $y_l \leq y_1 < y_2 < ... < y_k \leq y_u$ such that $d(y_i, y_j) > \epsilon$ for all $i \neq j$ which implies that

$$\begin{aligned} d(y_l, y_u) &= d(y_l, y_1) + d(y_1, y_2) + \cdots + d(y_k, y_u) \\ &\geq d(y_1, y_2) + d(y_2, y_3) \cdots + d(y_{k-1}, y_k) \geq k\epsilon > (k-1)\epsilon \geq \epsilon/\epsilon = 1. \end{aligned}$$

But this contradicts to the fact that $d(y_l, y_u) \leq |A_N \odot \overrightarrow{F}_N|_1 = 1$. Hence, we have $D(\epsilon, A_N \odot \mathcal{F}_{N,\omega}) \leq 1/\epsilon + 1 \equiv \lambda(\epsilon)$ for all $\omega \in \Omega$ for any A_N and $\int_0^1 \sqrt{\lambda(\epsilon)} d\epsilon < \infty$ and (i) follows. To check

(ii), we first assume that $y_1 \leq y_2$ WLOG. Note that $f_{Ni}(y_1)f_{Ni}(y_2) = f_{Ni}^2(y_1)$ and

$$
\begin{aligned}
&\sum_{i=1}^{N} E\left[f_{Ni}(y_1)f_{Ni}(y_2)\right] = \sum_{i=1}^{N} E\left[f_{Ni}^2(y_1)\right] \\
=&N\left(E\left[\frac{1}{Nh}K^2\left(\frac{v-V}{h}\right)1(Y \leq y_1)\right]\right) \\
=&\frac{1}{h}\int_V\int_Y\left[K^2\left(\frac{v-V}{h}\right)1(Y \leq y_1)\right]f(Y,V)dYdV \\
=&\frac{1}{h}\int_V K^2\left(\frac{v-V}{h}\right)\left(\int_Y 1(Y \leq y_1)f(Y|V)dY\right)g(V)dV \\
=&\frac{1}{h}\int_V K^2\left(\frac{v-V}{h}\right)\left(\int_Y 1(Y \leq y_1)f(Y|V)dY\right)g(V)dV \\
=&\frac{1}{h}\int_V K^2\left(\frac{v-V}{h}\right)F(y_1|V)g(V)dV = \frac{1}{h}\int_V K^2\left(\frac{v-V}{h}\right)r(y_1,V)dV \\
=&\int_u K^2(u)r(y_1, v-uh)du = \int_u K^2(u)\left(r(y_1,v) - r_v(y_1, v^*(u))uh\right)du \\
=&\|K\|_2^2 r(y_1,v) - h\int_u uK^2(u)r_{vv}(y_1, v^*(u))du,
\end{aligned}
$$

where $v^*(u)$ denotes some point between v and $v-uh$ such that $r(y, v-uh) = r(y,v) - r_v(y, v^*(u))uh$ and r_v denotes the first order partial derivative of r with respect to v. Given $r_v(y,v)$ is uniformly bounded over y and v, we have $\sum_{i=1}^{N} E\left[f_{Ni}(y_1)f_{Ni}(y_2)\right] = \|K\|_2^2 r(y_1,v) + O(h)$. Similarly,

$$
\begin{aligned}
E[f_{Ni}(y)] &= \sqrt{\frac{1}{Nh}}\int_V\int_Y K\left(\frac{v-V}{h}\right)1(Y \leq y)f(Y,V)dYdV \\
&= \sqrt{\frac{h}{N}}\int_u K(u)r(y, v-uh)du \\
&= \sqrt{\frac{h}{N}}\int_u K(u)\big(r(y,v) - r_v(y,v)uh + r_{vv}(y, v^*(u))(uh)^2\big)du \\
&= \sqrt{\frac{h}{N}}r(y,v) + \sqrt{\frac{h^5}{N}}\int_u u^2K(u)r_{vv}(y, v^*(u))du,
\end{aligned}
$$

where $v^*(u)$ is some point between v and $v-uh$ such that $r(y,v-uh) = r(y,v) - r_v(y,v)uh + r_{vv}(y, v^*(u))(uh)^2$. Given $r(y,v)$ and $r_{vv}(y,v)$ are uniformly bounded over y and v, we have

$$\sum_{i=1}^{N} E[f_{Ni}(y_1)]E[f_{Ni}(y_2)] = NE[f_{N1}(y_1)]E[f_{N1}(y_2)] = O(h).$$

For all $y_l \leq y_1 \leq y_2 \leq y_u$, when N tends to infinity,

$$\begin{aligned}
E[\mathcal{Z}_N(y_1)\mathcal{Z}_N(y_2)] &= E\left[\left(\sum_{i=1}^N f_{Ni}(y_1) - E[f_{Ni}(y_1)]\right)\left(\sum_{i=1}^N f_{Ni}(y_2) - E[f_{Ni}(y_2)]\right)\right] \\
&= \sum_{i=1}^N E\big[(f_{Ni}(y_1) - E[f_{Ni}(y_1)])(f_{Ni}(y_2) - E[f_{Ni}(y_2)])\big] \\
&= N\left(E\big[(f_{Ni}(y_1) - E[f_{Ni}(y_1)])(f_{Ni}(y_2) - E[f_{Ni}(y_2)])\big]\right) \\
&= N\left(E\left[\frac{1}{Nh}K^2\left(\frac{v-V}{h}\right)1(Y \leq y_1)\right] - E[f_{N1}(y_1)]E[f_{N1}(y_2)]\right) \\
&= \|K\|_2^2 r(y_1, v) + O(h) \to \|K\|_2^2 r(y_1, v).
\end{aligned}$$

Therefore, (ii) holds. By similar argument, we have when N tends to infinity,

$$\sum_{i=1}^N E[F_{Ni}^2] = \sum_{i=1}^N E\big[f_{Ni}^2(y_u)\big] = \|K\|_2^2 r(y_u, v) + O(h) \to \|K\|_2^2 r(y_u, v),$$

which implies that $\limsup \sum_{i=1}^N E[F_{Ni}^2] < \infty$, and (iii) holds. Note that $1(F_{Ni} > \epsilon) = 1\left(K\left((v - V_i)/h\right) > \sqrt{Nh}\epsilon\right)$. Given $K(u)$ is bounded and $\sqrt{Nh} \to \infty$, then for any $\epsilon > 0$, $1(F_{Ni} > \epsilon) = 0$ for all N large enough. This implies that for any $\epsilon > 0$ and N large enough $\sum_{i=1}^N E[F_{Ni}^2 1(F_{Ni} > \epsilon)] = 0$, and (iv) follows. To show (v), first we define for any $y_l \leq y_1 < y_2 \leq y_u$,

$$\begin{aligned}
\rho_N(y_2, y_1) &= \left(\sum_{i=1}^N E\left[(f_{Ni}(y_2) - f_{Ni}(y_1))^2\right]\right)^{\frac{1}{2}}, \\
\rho(y_2, y_1) &= \left(\|K\|_2^2\left(r(y_2, v) - r(y_1, v)\right)\right)^{\frac{1}{2}}.
\end{aligned}$$

Note that for $y_l \leq y_1 < y_2 \leq y_u$,

$$f_{Ni}(y_2) - f_{Ni}(y_1) = \frac{1}{\sqrt{Nh}}K\left(\frac{v - V_i}{h}\right)1(y_1 < Y_i \leq y_2).$$

By the same argument, we have that

$$\begin{aligned}
|\rho_N^2(y_2, y_1) - \rho^2(y_2, y_1)| &= \sum_{i=1}^N E\left((f_{Ni}(y_2) - f_{Ni}(y_1))^2\right) - \|K\|_2^2\left(r(y_2, v) - r(y_1, v)\right) \\
=& N\left(E\left[\frac{1}{Nh}K^2\left(\frac{v-V}{h}\right)1(y_1 < Y \leq y_2)\right]\right) - \|K\|_2^2\left(r(y_2, v) - r(y_1, v)\right) \\
=& h\int_u uK^2(u)\left(r_v(y_2, v^*(u)) - r_v(y_1, v^*(u))\right)du \\
\leq& h\int_u K^2(u)\left(u(r_v(y_2, v^*(u)) - r_v(y_1, v^*(u))\right)du \\
\leq& h\int_u K^2(u)\,|u| \sup_{y_1, y_2, v}|r_v(y_2, v) - r_v(y_1, v)|\,du = hM_3
\end{aligned}$$

where $M_3 = \int_u |u|K^2(u)du \sup_{y_1,y_2,v} |r_v(y_2, v) - r_v(y_1, v)|$ which is a positive number not depending on y_1 and y_2. This implies that $\rho_N^2(y_2, y_1)$ converges to $\rho^2(y_2, y_1)$ uniformly over y_1 and y_2. It follows that $\rho_N(y_2, y_1)$ converges to $\rho(y_2, y_1)$ uniformly over y_1 and y_2 and this is sufficient for (v).

By Theorem 10.6 of Pollard (1990), $\mathcal{Z}_N(\cdot)$ converges to a Gaussian process $\mathcal{Z}(\cdot)$ with $Cov(\mathcal{Z}(y_1), \mathcal{Z}(y_2)) = \|K\|_2^2 r(\min\{y_1, y_2\}, v)$. We also have

$$\left|\sum_{i=1}^N E[f_{Ni}(y)] - \sqrt{Nh}r(y,v)\right| = \left|\sqrt{Nh^5}\int_u u^2 K(u) r_{vv}(y, v^*(u))du\right|$$
$$\leq \sup_{v,y} |r_{vv}(y,v)|\sqrt{Nh^5}\int_u u^2 K(u)du = M_4\sqrt{Nh^5},$$

where $M_4 = \sup_{v,y} |r_{vv}(y, v)| \int_u u^2 K(u)du$. It follows that when N tends to infinity,

$$\sup_y \left|\sum_{i=1}^N E[f_{Ni}(y)] - \sqrt{Nh}r(y,v)\right| \leq M_4\sqrt{Nh^5} \to 0.$$

We have

$$\sqrt{Nh}(\bar{r}(\cdot, v) - r(\cdot, v)) = \mathcal{Z}_N(\cdot) + \sum_{i=1}^N E[f_{Ni}(y)] - \sqrt{Nh}r(\cdot, v) \Rightarrow \mathcal{Z}(\cdot),$$

because $\sup_y |\sum_{i=1}^N E[f_{Ni}(y)] - \sqrt{Nh}r(y, v)| \to 0$. In addition, $\bar{r}(y_u, v) = \bar{g}(v)$. For any function b with $b(y_u) > 0$, let $\Psi(y; b(\cdot)) = b(y)/b(y_u)$ for $y_l \leq y \leq y_u$. The $\Psi(y; b(\cdot))$ is Hadamard-differentiable at $b(\cdot) = r(\cdot, v)$ with

$$\frac{\Psi(\cdot; r(\cdot, v) + t_N\psi_N) - \Psi(\cdot; r(\cdot, v))}{t_N} \to \frac{1}{g(v)}\psi(\cdot) - \frac{F(\cdot|x)}{g(v)}\psi(y_u)$$

for all $t_N \to 0$ and for all $\psi_N \in \ell^\infty[y_l, y_u]$ such that $\psi_N \to \psi \in \ell^\infty[y_l, y_u]$ in sup-norm. By delta method, we have

$$\begin{aligned}
&\sqrt{Nh}(\widehat{F}(\cdot|x) - F(\cdot|x)) \sim \sqrt{Nh}(\Psi(\cdot; \bar{r}(\cdot, v)) - \Psi(\cdot; r(\cdot, v))\\
\Rightarrow &\frac{1}{g(v)}\mathcal{Z}(\cdot) - \frac{F(\cdot|x)}{g(v)}\mathcal{Z}(y_u) \equiv \mathcal{X}(\cdot).
\end{aligned}$$

Hence, for $y_1 \leq y_2$

$$\begin{aligned}
&Cov(\mathcal{X}(y_1), \mathcal{X}(y_2))\\
=&E\left[\left(\frac{1}{g(v)}\mathcal{Z}(y_1) - \frac{F(y_1|x)}{g(v)}\mathcal{Z}(y_u)\right)\left(\frac{1}{g(v)}\mathcal{Z}(y_2) - \frac{F(y_2|x)}{g(v)}\mathcal{Z}(y_u)\right)\right]\\
=&\|K\|_2^2\left(\frac{r(y_1, v)}{g^2(v)} - \frac{F(y_1|x)r(y_2, v)}{g^2(v)} - \frac{F(y_2|x)r(y_1, v)}{g^2(v)} + \frac{F(y_1|x)F(y_2|x)g(v)}{g^2(v)}\right)\\
=&\frac{\|K\|_2^2}{g(v)}(F(y_1|x) - F(y_1|x)F(y_2|x)).
\end{aligned}$$

Hence, these complete the proof of Theorem 5. □

Proof of Theorem 3.6: First, we show that $\mathcal{Z}_u(\cdot) \Rightarrow \mathcal{Z}(\cdot)$ conditional on sample path with probability approaching 1. Let $\mathcal{W}$ denote the sample path of $\{(Y_1, X_1), (Y_2, X_2), ...\}$. Define

$$f_{Ni}^u(y|\mathcal{W}) = \frac{U_i}{\sqrt{Nh}}\left(K\left(\frac{\hat{v} - \widehat{V}_i}{h}\right)1(Y_i \leq y)\right), F_{Ni|\mathcal{W}}^u = \left|\frac{U_i}{\sqrt{Nh}}K\left(\frac{\hat{v} - \widehat{V}_i}{h}\right)\right|$$

We want to show that (i)-(v) of Theorem 10.6 of Pollard (1990) hold conditional on sample path with probability approaching 1. Note that $\{f_{Ni}^u(y|\mathcal{W})\}$ are manageable because f_{Ni}^u are monotonic in y for all i. Hence, (i) holds. For $y_1 \leq y_2$

$$E[\mathcal{Z}_u(y_1)\mathcal{Z}_u(y_2)] = \sum_{i=1}^{N} \frac{1}{Nh}K^2\left(\frac{\hat{v} - \widehat{V}_i}{h}\right)1(Y_i \leq y_1).$$

As in Theorem 3.5, the estimation error from $\hat{\theta}$ would disappear in the limit. Hence,

$$\begin{aligned} E[\mathcal{Z}_u(y_1)\mathcal{Z}_u(y_2)] &= \sum_{i=1}^{N} \frac{1}{Nh}K^2\left(\frac{\hat{v} - \widehat{V}_i}{h}\right)1(Y_i \leq y_1) \\ &\sim \sum_{i=1}^{N} \frac{1}{Nh}K^2\left(\frac{v - V_i}{h}\right)1(Y_i \leq y_1) \xrightarrow{p} \|K\|_2^2 r(y_1, v). \end{aligned}$$

We have

$$\sum_{i=1}^{N} E[(F_{Ni|\mathcal{W}}^u)^2] = \sum_{i=1}^{N} \frac{1}{Nh}K^2\left(\frac{\hat{v} - \widehat{V}_i}{h}\right) \xrightarrow{p} \|K\|_2^2 r(y_u, v),$$

and this implies (iii). Given U_i and $K(u)$ are all bounded, for fixed $\epsilon > 0$, $F_{Ni|\mathcal{W}}^u \leq \epsilon$ for all i and for N is large enough. Therefore, for N large enough,

$$\sum_{i=1}^{N} E[(F_{Ni|\mathcal{W}}^u)^2 1(F_{Ni|\mathcal{W}}^u > \epsilon)] = 0$$

which implies (iv). Finally, for $y_1 \leq y_2$, we have

$$\begin{aligned} \rho_N^2(y_1, y_2) &= \sum_{i=1}^{N} E[(f_{Ni}^u(y_1|\mathcal{W}) - f_{Ni}^u(y_2|\mathcal{W}))^2] \\ &= \sum_{i=1}^{N} \frac{1}{Nh}K^2\left(\frac{\hat{v} - \widehat{V}_i}{h}\right)1(y_1 < Y_i \leq y_2) \\ &\sim \sum_{i=1}^{N} \frac{1}{Nh}K^2\left(\frac{v - V_i}{h}\right)1(y_1 < Y_i \leq y_2) \\ &\xrightarrow{p} \|K\|_2^2 \left(r(y_2, v) - r(y_1, v)\right) \equiv \rho^2(y_1, y_2) \end{aligned}$$

uniformly in (y_1, y_2) which is sufficient for (v). To see this, by the same argument in Theorem 5,

$$\sqrt{Nh}\left(\sum_{i=1}^{N}\frac{1}{Nh}K^2\left(\frac{\hat{v}-\widehat{V}_i}{h}\right)1(Y_i \le y) - \|K\|_2^2 r(y,v)\right)$$

will converge to a mean zero Gaussian process and this implies that

$$\sup_y \left|\sum_{i=1}^{N}\frac{1}{Nh}K^2\left(\frac{\hat{v}-\widehat{V}_i}{h}\right)1(Y_i \le y_1) - \|K\|_2^2 r(y_1,v)\right| = o_p(1). \tag{13}$$

Second,

$$\begin{aligned}\rho_N^2(y_1,y_2) &= \sum_{i=1}^{N}\frac{1}{Nh}K^2\left(\frac{\hat{v}-\widehat{V}_i}{h}\right)1(< y_1 < Y_i \le y_2)\\ =&\sum_{i=1}^{N}\frac{1}{Nh}K^2\left(\frac{\hat{v}-\widehat{V}_i}{h}\right)1(Y_i \le y_2) - \sum_{i=1}^{N}\frac{1}{Nh}K^2\left(\frac{\hat{v}-\widehat{V}_i}{h}\right)1(Y_i \le y_1).\end{aligned} \tag{14}$$

(13) and (14) imply that $\rho_N^2(y_1, y_2)$ will converge in probability to $\rho^2(y_1, y_2)$ uniformly in (y_1, y_2). Therefore, we show that $\mathcal{Z}_u(\cdot) \Rightarrow \mathcal{Z}(\cdot)$ conditional on sample path with probability approaching 1.

Also, we have

$$\begin{aligned}&\sup_y \left|\frac{1}{\hat{g}(\hat{v})}\mathcal{Z}_u(y) - \frac{\widehat{F}(\cdot|x)}{\hat{g}(\hat{v})}\mathcal{Z}(y_u) - \frac{1}{g(v)}\mathcal{Z}_u(y) + \frac{F(\cdot|x)}{g(v)}\mathcal{Z}(y_u)\right|\\ \le&\left|\frac{1}{\hat{g}(\hat{v})} - \frac{1}{g(v)}\right|\sup_y|\mathcal{Z}_u(y)| + \sup_y\left|\frac{\widehat{F}(\cdot|x)}{\hat{g}(\hat{v})} - \frac{F(\cdot|x)}{g(v)}\right||\mathcal{Z}(y_u)| = o_p(1).\end{aligned}$$

The first term in the second line is $o_p(1)$, because $|1/\hat{g}(\hat{v}) - 1/g(v)| = o_p(1)$ given $\hat{g}(\hat{v}) \xrightarrow{p} g(v) > 0$ and $\sup_y |\mathcal{Z}_u(y)|$ is $O_p(1)$. Because $\sup_y \left|\widehat{F}(y|x)/\hat{g}(\hat{v}) - F(y|x)/g(v)\right| = o_p(1)$ by Theorem 3.5 and $|\mathcal{Z}(y_u)| = O_p(1)$, the second term is $o_p(1)$. These imply that

$$\mathcal{X}^u(\cdot) \sim \frac{1}{g(v)}\mathcal{Z}_u(y) - \frac{F(\cdot|x)}{g(v)}\mathcal{Z}(y_u).$$

By applying Theorem 2.1 of Kosorok (2008), we have $\mathcal{X}^u(\cdot) \Rightarrow \mathcal{X}(\cdot)$ conditional on sample path with probability approaching 1. □

Proof of Theorem 3.7: The proof is similar to the proof of Theorem 3.5 when θ_0 is known. We omit it. □

Proof of Theorem 3.8: The proof is similar to the proof of Theorem 3.6 when θ_0 is known. We omit it. □

Proof of Theorem 4.1: Theorem 4.1 follows when we apply the functional delta method on the quantile functions. □

Proof of Theorem 4.2: The proof is similar to the proof of Theorem 3.6. □

Proof of Theorem 4.3: The first part follows from the continuous mapping theorem and the second part follows from the functional delta method. □

Proof of Theorem 4.4: The first part follows from the continuous mapping theorem. The second part is similar to the proof of Theorem 4.2. □

References

[1] Barrett, G.F. and S.G. Donald (2009). Statistical Inference with Generalized Gini Indices of Inequality and Poverty. *Journal of Business and Economic Statistics 27*, 1-17.

[2] Davidson, R. and J-Y. Duclos (2000). Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality. *Econometrica 68*, 1435-1464.

[3] Donald, S.G. and Y.-C. Hsu (2011). Estimation and Inference for Distribution Functions and Quantile Functions in Treatment Effect Models, Working Paper.

[4] Hall, P. and Q. Yao (2005). Approximating Conditional Distribution Functions Using Dimension Reduction. *Annals of Statistics 33*, 1404-1421.

[5] Kosorok, M.R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* New York: Springer.

[6] Linton, O., Maasoumi, E. and Whang (2005). Consistent Testing for Stochastic Dominance: A Subsampling Approach. *Review of Economic Studies 72*, 735-765.

[7] Masry, E.(1996). Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates. *ournal of Time Series Analysis 17*, 571-599.

[8] Pollard, D. (1990). *Empirical Processes: Theory and Applications.* Hayward California: Institute of Mathematical Statistics.