TEST manuscript No.
(will be inserted by the editor)

Supplemental Appendix to "Robust estimation and inference for general varying coefficients models with missing observations"

Francesco Bravo

Received: date / Accepted: date

Abstract This supplemental appendix contains two additional examples (with related simulation studies), an additional empirical application and all the proofs of the results presented in the main paper.

Mathematics Subject Classification (2010) 62E20 · 62G10

Keywords Local linear estimation · MAR · M and Z estimators · Wald statistic

1 Additional details on the GEE example of Section 4.1

Let

$$\begin{aligned}
\widehat{\Gamma}_{\mu}(u) &= \frac{1}{n}\sum_{i=1}^{n}\eta'\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)^T V_i\left(\widehat{\eta},\widehat{\beta}\right)^{-1}\Omega_i\widehat{\Pi}\left(\delta_i,X_{2i},U_i\right)s_{\widehat{\alpha}}\left(X_i,U_i\right)\times \\
&\quad \frac{\partial}{\partial\alpha}\psi\left(\widehat{A}_i^{-1/2}\left(Y_i-\eta\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)\right)\right)^T K_h\left(U_i-u\right), \\
\widehat{\Sigma}_{\widehat{\pi}\mu}(u) &= \frac{1}{n}\sum_{i=1}^{n}\left[\left(\eta'\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)^T V_i\left(\widehat{\eta},\widehat{\beta}\right)^{-1}\Omega_i\widehat{\Pi}\left(\delta_i,X_{2i},U_i\right)\right)\times\right. \\
&\quad \left.\psi\left(\widehat{A}_i^{-1/2}\left(Y_i-\eta\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)\right)\right)\right]^{\otimes 2}K_h\left(U_i-u\right)- \\
&\quad \left[\frac{1}{n}\sum_{i=1}^{n}\eta'\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)^T V_i\left(\widehat{\eta},\widehat{\beta}\right)^{-1}\Omega_i\widehat{\Pi}\left(X_i,U_i\right)\times\right. \\
&\quad \left.\psi\left(\widehat{A}_i^{-1/2}\left(Y_i-\eta\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)\right)\right)K_h\left(U_i-u\right)\right]^{\otimes 2},
\end{aligned}$$

Francesco Bravo
Department of Economics, University of York, UK
Tel.: +44 01904 323794
Fax:
E-mail: francesco.bravo@york.ac.uk

where $\widehat{a}_{1\widehat{\pi}} =: \widehat{a}^{Z}_{1\widehat{\pi}}$, $\widehat{A}_i = \widehat{\phi} diag\left(\eta\left(X_{1i}^T \widehat{a}_{1\widehat{\pi}}\right), ..., \eta\left(X_{mi}^T \widehat{a}_{1\widehat{\pi}}\right)\right)$, $\Omega_i = diag\left(\omega\left(X_{1i}\right), ..., \omega\left(X_{mi}\right)\right)$ and $\widehat{\Pi}\left(\delta_i, X_{2i}, U_i\right) = diag\left(\delta_{1i}/\widehat{\pi}\left(X_{12i}, U_{1i}\right), ..., \delta_{mi}/\widehat{\pi}\left(X_{m2i}, U_{mi}\right)\right)$.

The following proposition shows that $\widehat{\Gamma}_\mu(u)$ and $\widehat{\Sigma}_{\widehat{\pi}\mu}(u)$ are consistent estimators for $f(u)\,\Gamma_{\mu 0}(u)$ and $f(u)\,\Sigma_{\pi\mu 0}(u)$.

Proposition 3 *Under the assumptions of Theorems 3-4 and if* $\left|\widehat{\phi} - \phi\right| = o_p(1)$, $\left\|\widehat{\beta} - \beta_0\right\| = o_p(1)$, *then*

$$\left\|\widehat{\Gamma}_\mu(u) - f(u)\,\Gamma_{\mu 0}(u)\right\| = o_p(1),$$
$$\left\|\widehat{\Sigma}_{\widehat{\pi}\mu}(u) - f(u)\,\Sigma_{\pi\mu 0}(u)\right\| = o_p(1).$$

The Raphson-Newton algorithm is based on the local estimating equations

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\left(\frac{\exp\left(X_i^T \widehat{a} W_i\right)}{\left(1+\exp\left(X_i^T \widehat{a} W_i\right)\right)^2}\begin{bmatrix} X_i \\ X_i\left(U_i - u\right)/h\end{bmatrix}\right)^T \times \right. \tag{1}$$
$$\left. \widehat{A}^{-1/2}\Omega_i \widehat{\Pi}\left(\delta_i, X_i, U_i\right)\psi_c\left(\widehat{t}_i\right) - G'_\omega\left(Y_i, \frac{\exp\left(X_i^T \widehat{a} W_i\right)}{1+\exp\left(X_i^T \widehat{a} W_i\right)}\right)\right\} K_h\left(U_i - u\right) :=$$
$$\frac{1}{n}\sum_{i=1}^{n}\widehat{\Psi}_{\widehat{\pi}c}\left(\widehat{t}_i, u\right) = 0,$$

where

$$\widehat{t}_i = \widehat{A}^{-1/2}\left(Y_i - \frac{\exp\left(X_i^T \widehat{a} W_i\right)}{1+\exp\left(X_i^T \widehat{a} W_i\right)}\right),$$
$$\widehat{A} = \widehat{\phi} diag\left(\frac{\exp\left(X_{1i}^T \widehat{a}_1\right)}{\left(1+\exp\left(X_{1i}^T \widehat{a}_1\right)\right)^2}, ..., \frac{\exp\left(X_{mi}^T \widehat{a}_1\right)}{\left(1+\exp\left(X_{mi}^T \widehat{a}_1\right)\right)^2}\right),$$

$\widehat{\phi} = \left\{1.483 med\left(\left|\widehat{\varepsilon}_{ki} - med\left(\widehat{\varepsilon}_{ij}\right)\right|\right)\right\}^2$ (the same estimator used by He, Fung and Zhu (2005)),

$$G'_\omega\left(Y_i, \frac{\exp\left(X_i^T \widehat{a} W_i\right)}{1+\exp\left(X_i^T \widehat{a} W_i\right)}\right) = \left(\frac{\exp\left(X_i^T \widehat{a} W_i\right)}{\left(1+\exp\left(X_i^T \widehat{a} W_i\right)\right)^2}\begin{bmatrix} X_i \\ X_i\left(U_i - u\right)/h\end{bmatrix}\right)^T \times$$
$$\widehat{A}^{-1/2}\Omega_i \widehat{\Pi}\left(\delta_i, X_i, U_i\right)\left\{c\left[\Pr\left(Y_i \geq \widehat{j}_{2i} + 1\right) - \Pr\left(Y_i \geq \widehat{j}_{1i}\right)\right] - \widehat{B}^{1/2}\Pr\left(\widehat{j}_{1i} + 1 \leq Y_i \leq \widehat{j}_{2i}\right)\right\},$$

$$\widehat{B} = diag\left(\exp\left(X_{1i}^T \widehat{a}_1\right), ..., \exp\left(X_{mi}^T \widehat{a}_1\right)\right),$$

$$j_{1i}=\begin{bmatrix}\frac{\exp\left(X_{1i}^TaW_i\right)}{1+\exp\left(X_{1i}^TaW_i\right)}\\ \vdots\\ \frac{\exp\left(X_{mi}^TaW_i\right)}{1+\exp\left(X_{mi}^TaW_i\right)}\end{bmatrix}-c\begin{bmatrix}\left(\frac{\exp\left(X_{1i}^TaW_i\right)}{\left(1+\exp\left(X_{1i}^TaW_i\right)\right)^2}\right)^{1/2}\\ \vdots\\ \left(\frac{\exp\left(X_{mi}^TaW_i\right)}{\left(1+\exp\left(X_{mi}^TaW_i\right)\right)^2}\right)^{1/2}\end{bmatrix},$$

$$j_{2i}=\begin{bmatrix}\frac{\exp\left(X_{1i}^TaW_i\right)}{1+\exp\left(X_{1i}^TaW_i\right)}\\ \vdots\\ \frac{\exp\left(X_{mi}^TaW_i\right)}{1+\exp\left(X_{mi}^TaW_i\right)}\end{bmatrix}+c\begin{bmatrix}\left(\frac{\exp\left(X_{1i}^TaW_i\right)}{\left(1+\exp\left(X_{1i}^TaW_i\right)\right)^2}\right)^{1/2}\\ \vdots\\ \left(\frac{\exp\left(X_{mi}^TaW_i\right)}{\left(1+\exp\left(X_{mi}^TaW_i\right)\right)^2}\right)^{1/2}\end{bmatrix},$$

$$\begin{aligned}\widehat{\Gamma}_{\mu}(u)=&\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)}{\left(1+\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)\right)^2}X_i\right)^T\widehat{A}^{-1/2}\mathit{\Omega}_i\widehat{\Pi}\left(\delta_i,X_i,U_i\right)\left[\widehat{A}^{-1}\frac{\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)}{1+\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)}\times\right.\\&c\left(\Pr\left(Y_i\le j_{1i}\right)-\Pr\left(Y_i\ge j_{2i}+1\right)\right)^T+\widehat{A}\Pr\left(\widehat{j}_{2i}+1\le Y_i\le\widehat{j}_{2i}\right)^T\times\\&\left(\frac{\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)}{\left(1+\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)\right)^2}X_i\right)K_h\left(U_i-u\right),\end{aligned}$$

$$\begin{aligned}\widehat{\Sigma}_{\widehat{\pi}\mu}(u)=&\frac{1}{n}\sum_{i=1}^{n}\left\{\left(\frac{\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)}{\left(1+\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)\right)^2}X_i\right)^T\widehat{A}^{-1}\mathit{\Omega}_i\widehat{\Pi}\left(\delta_i,X_i,U_i\right)\right.\\&\left(c^2\left[\Pr\left(Y_i\le\widehat{j}_{1i}\right)-\Pr\left(Y_i\ge\widehat{j}_{2i}+1\right)\right]^{\otimes 2}+\widehat{B}\Pr\left(\widehat{j}_{2i}+1\le Y_i\le\widehat{j}_{2i}\right)^{\otimes 2}\right)\times\\&\left.\widehat{\Pi}\left(\delta_i,X_i,U_i\right)\mathit{\Omega}_i\left(\frac{\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)}{\left(1+\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)\right)^2}X_i\right)\right\}K_h\left(U_i-u\right)-\\&\left(\frac{1}{n}\sum_{i=1}^{n}G'_{\omega}\left(Y_i,\frac{\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)}{1+\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)}\right)K_h\left(U_i-u\right)\right)^{\otimes 2}\end{aligned}$$

and the inequalities should be interpreted componentwise.

2 Additional examples and Monte Carlo

2.1 Robust quasi-likelihood estimation

The first additional example considers quasi-likelihood estimation of a varying coefficients model, and extends the approach of Cantoni and Ronchetti (2001) by proposing a general class of IPW based local Mallows quasi-likelihood estimators. In this example the robust loss function $\zeta_\omega(\cdot)$ is

$$\zeta_\omega(Y,\eta(X,\alpha(U))) = -\int_{v_0}^{\eta\left(X^T\alpha(U)\right)} \nu(Y,t)\,\omega(X)\,dt - G_\omega\left(Y,\eta\left(X^T\alpha(U)\right)\right),$$

where

$$\nu(Y,t) = Var(t)^{-1/2}\,\psi(r(Y,t)), \quad \nu(Y,t_0)=0,$$

$\eta(\cdot)$ is the inverse link function, $r(Y,t)=Var(t)^{-1/2}(Y-t)$ is the Pearson residual, $\psi(\cdot)$ is a robust function and $G_\omega(\cdot)$ is a correction factor used to achieve Fisher consistency.

Let

$$G'_\omega\left(Y,\eta\left(X^T\alpha(U)\right)\right) = E\left[\nu\left(Y,\eta\left(X^T\alpha(U)\right)\right)\eta'\left(X^T\alpha(U)\right)\omega(X)\right],$$

where $\eta'(\cdot)=\partial\eta(\cdot)/\partial a$. Calculations similar to those given in Cantoni and Ronchetti (2001) show that

$$\begin{aligned}
&\zeta_{\omega 1}\left(Y,\eta\left(X^T aW\right)\right) = \nu\left(Y,\eta\left(X^T aW\right)\right)\eta'\left(X^T aW\right)\omega(X) - G'_\omega\left(Y,\eta\left(X^T aW\right)\right),\\
&\zeta_{\omega 2}\left(Y,\eta\left(X^T aW\right)\right) = \nu'\left(Y,\eta\left(X^T aW\right)\right)\left[\eta'\left(X^T aW\right)\right]^{\otimes 2}\omega(X) + \\
&\sum_{j=1}^{2k}\nu\left(Y,\eta\left(X^T aW\right)\right)\frac{\partial}{\partial a_j}\eta'(XaW)\,\omega(X) - G''_\omega\left(Y,\eta\left(X^T aW\right)\right),\\
&\Gamma_{\zeta 0}(u) = E\left[\frac{\psi\left(r\left(Y,\eta\left(X^T\alpha_0(U)\right)\right)\right)}{V\left(\eta\left(X^T\alpha_0(U)\right)\right)^{1/2}} s_\alpha(X,U)\left[\eta'\left(X^T\alpha_0(U)\right)X\right]^{\otimes 2}\omega(X)\,|U=u\right],\\
&\Sigma_{\pi\zeta 0}(u) = E\left\{\frac{1}{\pi(X_2,U_2)}\left[\frac{\psi\left(r\left(Y,\eta\left(X^T\alpha_0(U)\right)\right)\right)}{V\left(\eta\left(X^T\alpha_0(U)\right)\right)^{1/2}}\eta'\left(X^T\alpha_0(U)\right)X\omega(X)\right]^{\otimes 2}|U=u\right\} - \\
&\left\{f(u)\,E\left[\frac{\psi\left(r\left(Y,\eta\left(X^T\alpha_0(U)\right)\right)\right)}{V\left(\eta\left(X^T\alpha_0(U)\right)\right)^{1/2}}\eta'\left(X^T\alpha_0(U)\right)X\omega(X)\,|U=u\right]\right\}^{\otimes 2},
\end{aligned}$$

where $s_\alpha(X,U)=\partial\log f(Y|X,U)/\partial\alpha$ and $f(Y|X,U)$ is the conditional density of the response Y.

Let

$$\begin{aligned}
\widehat{\Gamma}_\zeta(u) &= \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\widehat{\pi}(X_{2i},U_i)}\frac{\psi\left(r\left(Y_i,\eta\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)\right)\right)}{V\left(\eta\left(X_i^T\widehat{a}_1\right)\right)^{1/2}} s_{\widehat{\alpha}}(X_i,U_i)\left[\eta'\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)X_i\right]^{\otimes 2}\times\\
&\quad\omega(X_i)\,K_h(U_i-u),\\
\widehat{\Sigma}_{\widehat{\pi}\zeta}(u) &= \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\widehat{\pi}(X_{2i},U_i)^2}\left[\frac{\psi\left(r\left(Y_i,\eta\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)\right)\right)}{V\left(\eta\left(X_i^T\widehat{a}_1\right)\right)^{1/2}}\eta'\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)X_i\omega(X_i)\right]^{\otimes 2}K_h(U_i-u) - \\
&\quad\left\{\frac{1}{n}\sum_{i=1}^n\left[\frac{\delta_i}{\widehat{\pi}(X_{2i},U_i)}\frac{\psi\left(r\left(Y_i,\eta\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)\right)\right)}{V\left(\eta\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)\right)^{1/2}}\eta'\left(X_i^T\widehat{a}_1\right)X_i\omega(X_i)\,K_h(U_i-u)\right]\right\}^{\otimes 2} \quad (2)
\end{aligned}$$

where $\widehat{a}_{1\widehat{\pi}} = \widehat{a}_{1\widehat{\pi}}^{M}$The following proposition shows that $\widehat{\Gamma}_{\zeta}(u)$ and $\widehat{\Sigma}_{\widehat{\pi}\zeta}(u)$ are consistent estimators for $f(u)\,\Gamma_{\zeta 0}(u)$ and $f(u)\,\Sigma_{\pi\zeta 0}(u)$.

Proposition 4 *Under the assumptions of Theorems 1 and 2 in the main paper*

$$\left\| \widehat{\Gamma}_{\zeta}(u) - f(u)\,\Gamma_{\zeta 0}(u) \right\| = o_p(1),$$
$$\left\| \widehat{\Sigma}_{\widehat{\pi}\zeta}(u) - f(u)\,\Sigma_{\pi\zeta 0}(u) \right\| = o_p(1).$$

2.2 Robust nonlinear regression estimation

The second additional example considers estimation of a possibly nonlinear varying coefficients regression model

$$Y = \eta(X, \alpha_0(U)) + \varepsilon,$$

where the unobservable error ε is an i.i.d. random variable from a zero mean symmetric distribution independent of X and U. In this example the robust loss function $\zeta_{\omega}(\cdot)$ is

$$\zeta_{\omega}(Y, \eta(X, \alpha(U))) = \rho(Y - \eta(X, \alpha(U)))\,\omega(X),$$

where $\rho(\cdot)$ is a robust function. Some calculations show that

$$\begin{aligned}
&\zeta_{\omega 1}(Y, \eta(X, aW)) = -\psi(Y - \eta(X, aW))\,\frac{\partial \eta(X, aW)}{\partial a}\omega(X) \\
&\zeta_{\omega 2}(Y, \eta(X, aW)) = \psi'(Y - \eta(X, aW))\,\frac{\partial \eta(X, aW)}{\partial a}^{\otimes 2}\omega(X) + \\
&\psi(Y - \eta(X, aW))\,\frac{\partial^2 \eta(X, aW)}{(\partial a)^{\otimes 2}}\omega(X), \\
&\Gamma_{\zeta 0}(u) = E[\psi'(\varepsilon)]\,E\left[\frac{\partial \eta(X, \alpha_0(U))}{\partial \alpha}^{\otimes 2}\omega(X)\,|U = u\right], \\
&\Sigma_{\pi\zeta 0}(u) = E\left[\psi(\varepsilon)^2\right] E\left[\frac{1}{\pi(X_2, U)}\left(\frac{\partial \eta(X, \alpha_0(U))}{\partial \alpha}\omega(X)\right)^{\otimes 2}|U = u\right],
\end{aligned}$$

where $\psi(\cdot) = \rho'(\cdot)$. Let

$$\widehat{\Gamma}_{\zeta}(u) = \widehat{\psi}'(\widehat{\varepsilon}_i)\,\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}(X_{2i}, U_i)}\frac{\partial \eta(X_i, \widehat{a}_{1\widehat{\pi}})}{\partial \alpha}^{\otimes 2}\omega(X_i)\,K_h(U_i - u), \tag{3}$$

$$\widehat{\Sigma}_{\widehat{\pi}\zeta}(u) = \widehat{\psi}(\widehat{\varepsilon}_i)^2\,\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}(X_{2i}, U_i)^2}\left(\frac{\partial \eta(X_i, \widehat{a}_{1\widehat{\pi}})}{\partial \alpha}\omega(X_i)\right)^{\otimes 2}K_h(U_i - u),$$

$$\widehat{\psi}'(\widehat{\varepsilon}_i) = \frac{1}{n}\sum_{i=1}^{n}\psi'(\widehat{\varepsilon}_i),\ \widehat{\psi}(\widehat{\varepsilon}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}\psi(\widehat{\varepsilon}_i)^2,$$

where$\widehat{\varepsilon}_i = Y_i - \eta\left(X_i, \widehat{\alpha}\left(U_i\right)\right)$ is the regression residual and $\widehat{\alpha}\left(\cdot\right) = \widehat{a}_{1\widehat{\pi}}\left(\cdot\right)$. The following proposition shows that $\widehat{\varGamma}_{\zeta}\left(u\right)$ and $\widehat{\varSigma}_{\widehat{\pi}\zeta}\left(u\right)$ are consistent estimators for $f\left(u\right)\varGamma_{\zeta 0}\left(u\right)$ and $f\left(u\right)\varSigma_{\pi\zeta 0}\left(u\right)$.

Proposition 5 *Under the assumptions of Theorems 1 and 2 in the main paper*

$$\left\|\widehat{\varGamma}_{\zeta}\left(u\right) - f\left(u\right)\varGamma_{\zeta 0}\left(u\right)\right\| = o_p\left(1\right),$$
$$\left\|\widehat{\varSigma}_{\widehat{\pi}\zeta}\left(u\right) - f\left(u\right)\varSigma_{\pi\zeta 0}\left(u\right)\right\| = o_p\left(1\right).$$

2.3 Monte Carlo

In the varying coefficients Poisson regression example, the covariates X_1 and X_2 are jointly normally distributed with mean zero, unit variances and correlation coefficient set to 0.3, the covariate U is independent of X and uniformly distributed between 0 and 1, and $\alpha_0\left(U\right) = \left[\sin\left(\pi U/2\right), \cos\left(\pi U\right)\right]^T$. The selection probabilities are specified as

$$\pi\left(X_2, U\right) = \frac{\exp\left(\gamma_{10} + \gamma_{20}X_2 + \gamma_{30}U\right)}{1 + \exp\left(\gamma_{10} + \gamma_{20}X_2 + \gamma_{30}U\right)}, \tag{4}$$

with $\left[\gamma_{10}, \gamma_{20}, \gamma_{30}\right]^T = \left[1, 1.5, 0.3\right]^T$, which implies that the average percentage of missing is about 35%. We consider four cases: the first one (case 0) has no outliers and no missing values; the other three cases, respectively, cases C0, C1 and C2) have no outliers but missing values and two contaminated data with missing values, where

$$\begin{aligned} &(C1)\ X_{1j} \sim N\left(10, 1\right), X_{2j} \sim N\left(10, 1\right), \\ &(C2)\ Y_j = 30 \text{ and } X_{2j} \sim N\left(10, 1\right), \quad \left(j = 1, ..., 5\right) \end{aligned} \tag{5}$$

that is in (C1) the first five elements of the covariates X_i are outliers, whereas in (C2) the first five elements of both the responses and the covariates are outliers.

The computation of $\widehat{a}_{\widehat{\pi}}$ is carried out using the Newton-Raphson algorithm on the local estimating equations (first order conditions)

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}\left(X_i, U_i\right)}\left\{\psi\left(\widehat{t}_i\right)\exp\left(\frac{X_i^T\widehat{a}W_i}{2}\right)\begin{bmatrix} X_i \\ X_i\left(U_i - u\right)/h\end{bmatrix}\omega\left(X_i\right) - \right. \\ &\left. G'_{\omega}\left(Y_i, \exp\left(X_i^T\widehat{a}W_i\right)\right)\right\}K_h\left(U_i - u\right) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\varPsi}_{\widehat{\pi}c}\left(\widehat{t}_i, u\right) = 0, \end{aligned} \tag{6}$$

where $\widehat{t}_i = Y_i - \exp\left(X_i^T\widehat{a}W_i\right)/\exp\left(X_i^T\widehat{a}W_i/2\right)$,

$$\begin{aligned} &G'_{\omega}\left(Y_i, \exp\left(X_i^T aW_i\right)\right) = \left\{c\left[\Pr\left(Y_i \geq j_{2i} + 1\right) - \Pr\left(Y_i \leq j_{1i}\right)\right] + \right. \\ &\left.\left[\Pr\left(Y_i = j_{1i}\right) - \Pr\left(Y_i = j_{2i}\right)\right]\right\}\exp\left(\frac{X_i^T aW_i}{2}\right)\begin{bmatrix} X_i \\ X_i\left(U_i - u\right)\end{bmatrix}\omega\left(X_i\right) \end{aligned} \tag{7}$$

and $j_{1i} = \exp\left(X_i^T a W_i\right) - c\exp\left(X_i^T a W_i/2\right)$, $j_{2i} = \exp\left(X_i^T a W_i\right) + c\exp\left(X_i^T a W_i/2\right)$. Note also that

$$\widehat{\Gamma}_\zeta(u) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}(X_{2i},U_i)}\left\{c\left[\Pr\left(Y_i = \widehat{j}_{1i}\right) + \Pr\left(Y_i = \widehat{j}_{2i}\right)\right] + \exp\left(\frac{X_i^T\widehat{a}_{1\widehat{\pi}}}{2}\right)\times\right.$$
$$\left[\Pr\left(Y_i = \widehat{j}_{1i} - 1\right) - \Pr\left(Y_i = \widehat{j}_{1i}\right) - \Pr\left(Y_i = \widehat{j}_{2i} - 1\right) + \right.$$
$$\left.\left.\Pr\left(Y_i = \widehat{j}_{2i}\right)\right] + \exp\left(\frac{X_i^T\widehat{a}_{1\widehat{\pi}}}{2}\right)\Pr\left(\widehat{j}_{1i} \le Y_i \le \widehat{j}_{2i} - 1\right)\right\}\times$$
$$\exp\left(\frac{3X_i^T\widehat{a}_{1\widehat{\pi}}}{2}\right)X_i^{\otimes 2}\omega(X_i)K_h(U_i - u)$$

$$\widehat{\Sigma}_{\widehat{\pi}\zeta}(u) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}(X_{2i},U_i)^2}\left\{c^2\left[\Pr\left(Y_i \le \widehat{j}_{1i}\right) - \Pr\left(Y_i \ge \widehat{j}_{2i} + 1\right)\right] + \frac{1}{\exp\left(\frac{X_i^T\widehat{a}_{1\widehat{\pi}}}{2}\right)}\times\right.$$
$$\left[\exp\left(2X_i^T\widehat{a}_{1\widehat{\pi}}\right)\Pr\left(\widehat{j}_{1i} - 1 \le Y_i \le \widehat{j}_{2i} - 1\right) + \left(\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right) - 2\exp\left(2X_i^T\widehat{a}_{1\widehat{\pi}}\right)\right)\times\right.$$
$$\left.\left.\Pr\left(\widehat{j}_{1i} \le Y_i \le \widehat{j}_{2i} - 1\right) + \exp\left(2X_i^T\widehat{a}_{1\widehat{\pi}}\right)\Pr\left(\widehat{j}_{1i} + 1 \le Y_i \le \widehat{j}_{2i}\right)\right\}\times$$
$$\exp\left(\frac{X_i^T\widehat{a}_{1\widehat{\pi}}}{2}\right)X_i^{\otimes 2}\omega(X_i)^2K_h(U_i - u) -$$
$$\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}(X_{2i},U_i)}G'_\omega\left(Y_i, \exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)\right)K_h(U_i - u)\right]^{\otimes 2},$$

and $\widehat{j}_{1i} = \exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right) - c\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}/2\right)$, $\widehat{j}_{2i} = \exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right) + c\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}/2\right)$. As in the Monte Carlo section in the main paper, the robust function $\psi(\cdot)$ is specified as the Huber function $\psi_c(\cdot)$ with tuning constant $c = 1.2$, the weight function $\omega(\cdot)$ is

$$\omega(X) = \left(1 + \left\|S^{-1}(X - M)\right\|^2/2\right)^{-1/2}, \tag{8}$$

where M and S are robust location-scatter estimators such covarianceas the minimum determinant estimator (MCD) and the kernel function used is the Epanechnikov one. The bandwidth h is selected by the same robust cross-validation procedure used in the main paper, in which in the first step for a given h, $\widehat{t}_{-i} = \sum_{j\neq i}^{n}\Psi_{\widehat{\pi}c}\left(\widehat{t}_j, u\right)/n = 0$ and in the second step the robust bandwidth is chosen as $\widehat{h} = \arg\min_h \sum_{i=1}^{n}\delta_i\zeta_\omega\left(\widehat{t}_{-i}\right)/\widehat{\pi}(X_{2i},U_i)$, that is, in this example $\widehat{h}$ is the minimizer of the robust quasi deviance. The selection probabilities (4) are estimated using either a robust parametric logit or a nonparametric estimator using the same weight function $\omega(\cdot)$ as that used in (8). The bandwidth b for the nonparametric estimator is chosen using another robust cross-validation procedure where $\widehat{b} = \arg\min_b \sum_{i=1}^{n}\left|\delta_i - \widehat{\pi}_{-i}(X_{2i},U_i)\right|\omega(X_{2i})$ and $\widehat{\pi}_{-i}(X_{2i},U_i)$ is a robust leave one out kernel estimator.

Table 3 reports the mean absolute bias (B) - defined in the main paper, and standard deviation (SD) of six different estimators: the robust local complete case estimator $\widehat{a}_{1kc}(\cdot)$ which is based on (6) without the estimated selection probabilities $\widehat{\pi}(X_{2i}, U_i)$, the IPW local robust estimator $\widehat{a}_{1k\widehat{\pi}p}(\cdot)$ with logit estimation of $\pi(X_{2i}, U_i)$, the IPW robust local estimator $\widehat{a}_{1k\widehat{\pi}np}(\cdot)$ with nonparametric estimation of $\pi(X_{2i}, U_i)$ $(k = 1, 2, 3)$ and their nonrobust analogs computed with the Huber function set to $c = \infty$ and no Fisher correction term $G'_{\omega}(\cdot)$.

Table 3 approx here.

Inference is based on the statistic $\max_{1\leq j\leq s} W(u_j)$ (with $s = 5$) for the hypothesis

$$H_{\delta} : \begin{bmatrix} \alpha_1(u_j) \\ \alpha_2(u_j) \end{bmatrix} = \begin{bmatrix} \alpha_{10}(u_j) \\ \alpha_{20}(u_j) \end{bmatrix} + \gamma \begin{bmatrix} \widehat{\alpha}^M_{1\widehat{\pi}} - \alpha_{10}(u_j) \\ \widehat{\alpha}^M_{2\widehat{\pi}} - \alpha_{20}(u_j) \end{bmatrix}, \tag{9}$$

where $u_j = [0.1, 0.3, 0.5, 0.7, 0.9]$ (i.e. for $j = 1$ $u_1 = 0.1$ and likewise for $j = 2, 3, 4, 5$), $\alpha_{10}(u_j) = \sin(\pi u_j/2) = [0.156, 0.453, 0.707, 0.891, 0.987]$, $\alpha_{20}(u_j) = \cos(\pi u_j) = [0.951, 0.587, 0, 0.587, 0.951]$, $\widehat{\alpha}^M_{\widehat{\pi}}$ is the parametric estimator defined in (12) in the main paper and $\gamma \in \mathbb{R}$ index the departure from the null hypothesis (corresponding to $\gamma = 0$). The upper 10% and 5% critical values of the nonstandard distribution given in Proposition 2 are $[5.070, 6.631]$ for $n = 200$ and $[4.907, 6.416]$ for $n = 500$. Table 4 reports the finite sample size at the 0.10 and 0.05 nominal level of $\max_{1\leq j\leq 5} W(u_j)$ based on the six estimators used in Table 3 under no missing observations and no contamination (case 0), no contamination (case C0) and contamination (Case c2): $\max_{1\leq j\leq 5} W_{co}(u_j)$ is based on the robust local complete case estimator $\widehat{a}_{1c}(\cdot)$, $\max_{1\leq j\leq 5} W_{\widehat{\pi}p}(u_j)$ is based on the IPW robust local estimator with robust logit estimation of the selection probabilities and $\max_{1\leq j\leq 5} W_{\widehat{\pi}np}(u_j)$ is based on the IPW robust local estimator with robust nonparametric estimation of the selection probabilities.

Table 4 approx. here

Figure 3 shows the contour plots of the size adjusted finite sample power curves for the test of H_{γ} over the grid $\gamma \times \gamma = [-1, -0.75, ..., 0, ...0.75, 1] \times [-1, -0.75, ..., 0, ...0.75, 1]$ at the contour level of 0.45 for the C0 and C2 cases. Smaller contour plots indicate higher finite sample power.

Figure 3 approx. here

In the nonlinear varying coefficients regression, the relation between the response variable Y and the covariates $X = [X_1, X_2]^T$ and U is

$$Y = \exp(X_1\alpha_{10}(U) + X_2\alpha_{20}(U)) + \varepsilon,$$

where X_1 and X_2 are jointly normally distributed with mean zero, unit variances and correlation coefficients between X_1 and X_2 set to 0.3, U is independent of X and uniformly distributed between 0 and 0.6, ε is a standard normal independent of X and U and $\alpha_0(U) = [\sin(2\pi U), \exp(-U)]^T$. The selection probabilities $\pi(X_{2i}, U_i)$ are the same as those used in the previous example and so are the $(C1)$ and $(C2)$ contamination of the data. The estimators for $\alpha_0(\cdot)$ are computed using the same Huber function and weight function $\omega(\cdot)$ as those used in the previous example, and the Newton-Raphson algorithm on the local estimating equations (first order conditions)

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}(X_{2i}, U_i)}\psi_c(\widehat{\varepsilon}_i)\exp\left(X_i^T\widehat{a}W_i\right)\begin{bmatrix} X_i \\ X_i(U_i - u)/h \end{bmatrix}\times$$
$$\omega(X_i) K_h(U_i - u) = 0,$$

where $\widehat{\varepsilon}_i = Y_i - \exp\left(X_i^T\widehat{a}W\right)$. Note also that

$$\widehat{\Gamma}_\zeta(u) = \widehat{\psi}'_c(\widehat{\varepsilon}_i)\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}(X_{2i}, U_i)}\exp\left(2X_i^T\widehat{a}_{1\widehat{\pi}}\right)X_i^{\otimes 2}\omega(X_i) K_h(U_i - u),$$
$$\widehat{\Sigma}_{\widehat{\pi}\zeta}(u) = \widehat{\psi}_c(\widehat{\varepsilon}_i)^2\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}(X_{2i}, U_i)^2}\left[\exp\left(X_i^T\widehat{a}_{1\widehat{\pi}}\right)X_i\omega(X_i)\right]^{\otimes 2} K_h(U_i - u).$$

Table 5 reports the bias (B) and standard deviation (SD) of the same six different local estimators $\widehat{a}_{1k*}(\cdot)$ used in Table 3 with bandwidths h and b chosen using the same robust cross validation method as that used in the previous example.

Table 5 approx. here

For inference the same hypothesis as that given in (9) with $u_j = [0.1, 0.2, 0.3, 0.4, 0.5]$, $\alpha_{10}(u_j) = \sin(2\pi u_j) = [0.587, 0.951, 0.951, 0.587, 0]$, $\alpha_{20}(u_j) = \exp(-u_j) = [0.904, 0.818, 0.740, 0.670, 0.606]$ and $\widehat{\alpha}_{\widehat{\pi}}^{M}$ is the parametric estimator defined in (10) in the main paper. Table 4 reports the finite sample size at the 0.10 and 0.05 nominal level of $\max_{1\le j\le 5} W(u_j)$ based on the six local estimators used in Table 4 under no missing observations and no contamination (case 0), no contamination (case C0) and contamination (case C2).

Table 6 approx. here

Figure 4 shows the contour plots of the size adjusted finite sample power curves for the test of H_γ over the grid $\gamma\times\gamma = [-1, -0.75, ..., 0, ...0.75, 1]\times[-1, -0.75, ..., 0, ...0.75, 1]$ at the contour level of 0.45.

Figure 4 approx. here

3 Additional empirical application

In this additional empirical application, the same damaged carrots data set of Phelps (1982) is used (see also McCullagh and Nelder (1989) and Cantoni and Ronchetti (2001)). The data is available in the R package `Robustbase` and is characterized by an outlier (namely observation 14) but no missing values. The following binomial varying coefficients model with logit link

$$\log\left(\frac{E\left(Y|X,U\right)}{s-E\left(Y|X,U\right)}\right)=\alpha_{10}\left(U\right)+X_{1}\alpha_{20}\left(U\right)+X_{2}\alpha_{30}\left(U\right)$$

is fitted to the carrots data, where Y represent the number of damaged carrots, X_1 and X_2 are two indicator variables taking the value of 1 if corresponding to blocks 1 or 2 and U is the logarithm of the dose of insecticide. The unknown varying coefficients are estimated using the same Huber function with $c = 1.2$ used in the previous two examples and the Newton-Raphson algorithm on the local estimating equations

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\psi_{c}\left(\widehat{t}_{i}\right)}{V_{i}\left(\widehat{a}\right)^{1/2}}\left(\frac{s_{i}\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)}{\left(1+\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)\right)^{2}}\begin{bmatrix}X_{i}\\ X_{i}\left(U_{i}-u\right)/h\end{bmatrix}\right)\omega\left(X_{i}\right)-\right.$$
$$\left.G_{\omega}'\left(Y_{i},\frac{\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)}{1+\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)}\right)\right\}K_{h}\left(U_{i}-u\right)=0, \tag{10}$$

where $\omega\left(\cdot\right)$ is as that given in (8), $X_i = [1, X_{2i}, X_{3i}]'$,

$$\widehat{t}_{i}=\frac{1}{V_{i}\left(\widehat{a}\right)^{1/2}}\left(Y_{i}-s_{i}\frac{\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)}{1+\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)}\right),$$

$$V_{i}\left(\widehat{a}\right)=\frac{\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)}{1+\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)}\left(\frac{s_{i}-\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)/\left(1+\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)\right)}{s_{i}}\right),$$

$$G_{\omega}'\left(Y_{i},\frac{\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)}{1+\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)}\right)=\left\{c\left[\Pr\left(Y_{i}\geq j_{2i}+1\right)-\Pr\left(Y_{i}\leq j_{1i}\right)\right]+\right.$$
$$\frac{\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)}{\left(1+\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)\right)V_{i}\left(a\right)^{1/2}}\left[\Pr\left(j_{1i}\leq\widetilde{Y}_{i}\leq j_{2i}-1\right)-\Pr\left(j_{1i}\leq Y_{i}\leq j_{2i}-1\right)\right]\times$$
$$\frac{\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)}{\left(1+\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)\right)^{2}}\begin{bmatrix}X_{i}\\ X_{i}\left(U_{i}-u\right)\end{bmatrix}\omega\left(X_{i}\right),$$

$$j_{1i}=\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)/\left(1+X_{i}^{T}\widehat{a}W_{i}\right)-cV_{i}\left(\widehat{a}\right)^{1/2},$$
$$j_{2i}=\exp\left(X_{i}^{T}\widehat{a}W_{i}\right)/\left(1+X_{i}^{T}\widehat{a}W_{i}\right)+cV_{i}\left(\widehat{a}\right)^{1/2}$$

and $\widetilde{Y}_i \sim Bin\left(s_i-1,\exp\left(X_i^T\widehat{a}W_i\right)\right)$.

Figure 5 shows the three varying coefficients, denoted $\widehat{\alpha}_j$ $(j = 1, 2, 3)$, which are estimated using both the robust local estimating equations given in (10) and the nonrobust local estimating equations[1], together with the associated 95% confidence intervals.

Figure 5 approx. here

The first estimated coefficient (intercept) shows the direct effect of the log dose of the insecticide on the number of damaged carrots and is, as expected, a decreasing function of it. The other two estimated varying coefficients represent the differential effect of the insecticide on blocks 1 and 2 with respect to block 3. The effect of the insecticide on the second estimated varying coefficient is ambiguous as it is first an increasing and then a decreasing function of it, whereas the third estimated varying coefficient appears to be an non decreasing function of the insecticide. In terms of the mean effect (i.e. $\overline{\widehat{\alpha}_j} = \sum_{i=1}^{n} \widehat{\alpha}_j (U_i)/n$ for $j = 1, 2, 3$) the nonrobust estimation procedure has $\left[\overline{\widehat{\alpha}_1}, \overline{\widehat{\alpha}_2}, \overline{\widehat{\alpha}_3}\right]^T = [-2.18, 0.53, 0.97]^T$, whereas the robust one has $\left[\overline{\widehat{\alpha}_1}, \overline{\widehat{\alpha}_2}, \overline{\widehat{\alpha}_3}\right]^T = [-2.10, 0.54, 0.91]^T$. The overall quasi deviance and robust quasi deviance, which can be used as a measure of fit of the model, are respectively 78.34 and 45.55. These results are interesting because they show that the proposed varying coefficients specification of the model has a better fit compared to that obtained by the parametric specification used by Cantoni and Ronchetti (2001). For inference, the null hypotheses of joint constancy of the three varying coefficients $H_0 : [\alpha_1 (u_j) - \alpha_{10}, \alpha_2 (u_j) - \alpha_{20}, \alpha_3 (u_j) - \alpha_{30}]^T = 0^T$ and of individual constancy of each of the varying coefficients $H_0 : [\alpha_k (u_j) - \alpha_{k0}] = 0$ $(k = 1, 2, 3)$ are tested using the $\max_{1\le j\le 5} W_c (u_j)$ statistic evaluated at the five points u_2, u_5, u_7, u_9 and u_{11}. The sample values of $\max_{1\le j\le 5} W_c (u_j)$ for the null hypotheses of joint and individual constancy of the three varying coefficients are, respectively, 11.23, 5.98, 6.84 and 7.23 with corresponding p-values of 0.035, 0.044, 0.032 and 0.027. Thus the null hypotheses of joint and individual constancy of the varying coefficients are rejected at the 0.05 significance level.

4 Proofs

Assume that

either

[1] The nonrobust local estimating equations are:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{t}_i}{V_i (\widehat{a})} \left(\frac{s_i \exp\left(X_i^T \widehat{a} W_i\right)}{\left(1 + \exp\left(X_i^T \widehat{a}) W_i\right)\right)^2} \begin{bmatrix} X_i \\ X_i (U_i - u)/h \end{bmatrix} \right) \times$$
$$K_h (U_i - u) = 0,$$

where $\widehat{t}_i = Y_i - s_i \exp\left(X_i^T \widehat{a} W_i\right)$, see for example Cai, Fan and Li (2000).

A1 (i) the function $\zeta(\cdot)$ is continuous for all $u \in \mathcal{U}$ and has the first two derivatives $\zeta_j(\cdot)$ $(j = 1, 2)$ $a.s.$, (ii) the function $\eta(\cdot)$ is twice continuously differentiable $a.s.$, (iii) there exists a unique $\alpha_0(U)$ such that $E[\zeta_\omega(Y, \eta(X, \alpha_0(U))) - \zeta_\omega(Y, \eta(X, \alpha(U)))\,|U = u] < 0$ for all $u \in \mathcal{U}$, (iv) $\sup_{u\in\mathcal{U}} \left\| H\left(\widehat{a}^M_{\widehat{\pi}} - a_0\right)\right\| = o_p(1)$,

A2 (i) $E\left[\sup_{\|\xi\|\le\epsilon} \|\zeta_{\omega2}(Y, \eta(X, \alpha_0(U) + \xi)) - \zeta_{\omega2}(Y, \eta(X, \alpha(U)))\|\,|U = u\right] = o(1)$ for all $u \in \mathcal{U}$, (ii)

$$E\left[\sup_{\|\xi\|\le\epsilon} \|\zeta_{\omega1}(Y, \eta(X, \alpha_0(U) + \xi)) - \zeta_{\omega1}(Y, \eta(X, \alpha_0(U))) - \right.$$
$$\zeta_{\omega2}(Y, \eta(X, \alpha_0(U)))\,\xi\|\;|U = u] = o(\epsilon)$$

as $\epsilon \to 0$ for all $u \in \mathcal{U}$,

A3 (i) $E[\zeta_{\omega1}(Y, \eta(X, \alpha_0(U)))\,|U = u] = 0$, (ii) $E\left[\zeta_{\omega1}(Y, \eta(X, \alpha_0(U)))^{\otimes2}\,|U = u\right]$ is continuous and positive definite (iii) $E\|\zeta_{\omega1}(Y, \eta(X, \alpha_0(U)))\,|U = u\|^{2+\beta} < \infty$ in a neighborhood of u for some $\beta > 0$ (iv) $E[\zeta_{\omega2}(Y, \eta(X, \alpha(U)))\,|U = u]$ is continuous and nonsingular for all $u \in \mathcal{U}$,

or

A1' (i) the function $\mu(\cdot)$ is continuous for all $u \in \mathcal{U}$ and has the first derivative $\mu_1(\cdot)$ $a.s.$, (ii) the function $\eta(\cdot)$ is twice continuously differentiable $a.s.$, (iii) there exists a unique $\alpha_0(U)$ such that $E[\mu_\omega(Y, \eta(X, \alpha_0(U)))\,|U = u] = 0$ for all $u \in \mathcal{U}$, (iv) $\sup_{u\in\mathcal{U}} \left\| H\left(\widehat{a}^Z_{\widehat{\pi}} - a_0\right)\right\| = o_p(1)$,

A2' (i) $E\left[\sup_{\|\xi\|\le\epsilon} \|\mu_{\omega1}(Y, \eta(X, \alpha_0(U) + \xi)) - \mu_{\omega1}(Y, \eta(X, \alpha(U)))\|\,|U = u\right] = o(1)$ for all $u \in \mathcal{U}$, (ii)

$$E\left[\sup_{\|\xi\|\le\epsilon} \|\mu_\omega(Y, \eta(X, \alpha_0(U) + \xi))\,\eta'(X, (\alpha_0(U) + \xi)) - \right.$$
$$\mu_\omega(Y, \eta(X, \alpha_0(U)))\,\eta'(X, \alpha_0(U)) - \mu_{\omega1}(Y, \eta(X, \alpha_0(U)))\,\xi\|\;|U = u] = o(\epsilon)$$

as $\epsilon \to 0$ for all $u \in \mathcal{U}$,

A3' (i) $E\left[(\mu_\omega(Y, \eta(X, \alpha_0(U)))\,\eta'(X, \alpha_0(U)))^{\otimes2}\,|U = u\right]$ is continuous and positive definite, (ii) $E\|\mu_\omega(Y, \eta(X, \alpha_0(U)))\,\eta'(X, \alpha_0(U))\,|U = u\|^{2+\beta} < \infty$ in a neighborhood of u for some $\beta > 0$ (iii) $E[\mu_{\omega1}(Y, \eta(X, \alpha_0(U)))\,|U = u]$ is continuous and nonsingular for all $u \in \mathcal{U}$,

and

A4 (i) U has bounded support $\mathcal{U}$, its density $f(\cdot)$ is continuous at u and $f(u) > 0$ for all $u \in \mathcal{U}$, (ii) $\alpha_0(\cdot)$ has continuous second derivative at u,

A5 (i) the kernel function $K(\cdot)$ is a probability density function on a bounded support and

$$\int K(v)\,dv = 1, \quad \int |v|^3 K(v)\,dv < \infty, \quad \int v^2 K(v)^2\,dv < \infty,$$

(ii) the bandwidth satisfies $h \to 0$ and $nh \to \infty$ as $n \to \infty$,

A6 *either* (i) $\pi(X_2, U, \gamma)$ is a parametric model for $\pi(X_2, U)$ where $\gamma \in \Gamma \subseteq \mathbb{R}^l$ such that $\inf_{X_2,U} \pi(X_2, U, \gamma) > 0$ for all $\gamma \in \Gamma$ and there exists a $\gamma_0 \in \Gamma$ such that $\pi(X_2, U, \gamma_0) = \pi(X_2, U)$, (ii) $E \sup_{\gamma \in \Gamma} \|\partial \pi(X_2, U, \gamma) / \partial \gamma\|^{\delta} < \infty$ for some $\delta > 2$ and the robust estimator $n^{1/2} \|\widehat{\gamma} - \gamma_0\| = O_p(1)$,
or (iii) X_2 has bounded support $\mathcal{X}_2$, $\inf_{X_2,U} \pi(X_2, U) > 0$ and $\pi(X_2, U)$ is twice continuously differentiable with bounded derivatives, (iv) $L(\cdot)$ is a bounded vth order ($v \geq 2$) kernel function and the bandwidth b satisfies $nb^{2(k_2+1)} \to \infty$ and $nhb^{2v} \to 0$.

A7 The preliminary estimator $\widehat{a}^*_{\widehat{\pi}}(u)$ for $* = M$ or$Z$ satisfies

$$\|H(\widehat{a}^*_{\widehat{\pi}} - a_0)\| = O_p\left(h^2 + (nh)^{-1/2}\right),$$

A8 *either* (i) $\pi(X, U, \gamma)$ is a parametric model for $\pi(X, U)$ where $\gamma \in \Gamma \subseteq \mathbb{R}^l$ such that $\inf_{X_2,U} \pi(X, U, \gamma) > 0$ for all $\gamma \in \Gamma$ and there exists a $\gamma_0 \in \Gamma$ such that $\pi(X, U, \gamma_0) = \pi(X, U)$, (ii) $E \sup_{\gamma \in \Gamma} \|\partial \pi(X, U, \gamma) / \partial \gamma\|^{\delta} < \infty$ for some $\delta > 2$ and the robust estimator $\widehat{\gamma}$ has the following (bounded) influence function

$$n^{1/2}(\widehat{\gamma} - \gamma_0) = M^{-1}(\gamma_0) \frac{1}{n^{1/2}} \sum_{i=1}^{n} r_{i\omega}(\gamma_0) + o_p(1) \tag{11}$$

or (iii) X has bounded support $\mathcal{X}$, $\inf_{X,U} \pi(X, U) > 0$ and $\pi(X, U)$ is twice continuously differentiable with bounded derivatives, (iv) $L(\cdot)$ is a bounded vth order ($v \geq 2$) kernel function and the bandwidth b satisfies $nb^{2(k+1)} \to \infty$ and $nhb^{2v} \to 0$.

Assumptions A1(i), A1'(i), A2 and A2' are relatively mild smoothness conditions that are satisfied by many robust estimators including Huber's ρ and ψ functions. Assumptions A1(iii) and A1(iii)' are standard identification conditions on the existence of a unique minimum and a unique solution for each $u \in calU$, respectively. A1(iii) implies that the solution to (3) in the main paper is unique. Assumptions A1(iv) and A1'(iv) are high level assumptions that can be replaced by more primitive conditions on the loss function, the estimating equations and/or the parameter space of $\alpha_0(\cdot)$, say A. For example if $\zeta_\omega(\cdot)$ is convex, the so-called convexity lemma (Pollard 1991) coupled with A1(iii) implies A1(iv). Alternatively, regularity conditions on A such as $N_{[]}(\epsilon, A, L_1(P)) < \infty$ - where $N_{[]}(\cdot)$ is the bracketing number of A with respect to the $L_1(P)$ norm - could be used. Assumptions A3(i) and A1'(iii) are local Fisher consistency conditions. They are satisfied for example when $\zeta_1(\cdot)$ is an odd real valued function and the distribution of the data is symmetric around zero, or when the unobservable errors in regression models are independent of the covariates and are symmetrically distributed around zero (Bianco, Gracia Ben and Yohai 2005). Alternatively a correction term has to be included, see for example Cantoni and Ronchetti (2001) and Boente, He and Zhou (2006). A3(ii)-(iv), A3'(ii)–(iii), A4 and A5 are standard in the nonparametric and semiparametric estimation literature. Assumption A6 imposes standard smoothness conditions on the two selection probability specifications and regularity conditions on their estimators. Assumption A7 requires that the preliminary M or Z estimators have the same convergence rate as that of non a standard nonparametric estimator. Finally Assumption A8 is used for Theorem 5. For an example of the influence function (11), see Bianco and Yohai (1996).

Proof (Proof of Theorem 1) Let $\widehat{\pi}_i$ denote either $\pi(X_{2i}, U_i, \widehat{\gamma})$ or $\widehat{\pi}(X_{2i}, U_i,)$; note for $\widehat{\pi}_i$ estimated parametrically, A6(ii), Jensen inequality and the law of large numbers imply that

$$\sup_i \left\| \frac{\partial \widehat{\pi}_i}{\partial \gamma} \right\| \leq n^{1/\delta} \left(\frac{1}{n} \sum_{i=1}^n \sup_{\gamma \in \Gamma} \left\| \frac{\partial \pi_i}{\partial \gamma} \right\|^{\delta} \right)^{1/\delta} = O_p\left(n^{1/\delta}\right),$$

hence by mean value expansion

$$\sup_i |\widehat{\pi}_i - \pi_i| \leq \sup_i \sup_{\gamma \in \Gamma} \left\| \frac{\partial \pi_i}{\partial \gamma} \right\| \|\widehat{\gamma} - \gamma_0\| = O_p\left(n^{1/\delta}\right) O_p\left(n^{-1/2}\right) = o_p(1).$$

For $\widehat{\pi}_i$ estimated nonparametrically, A6(iii)-(iv) and the results of Masry (1996) imply that

$$\sup_{u, x_2} |\widehat{\pi}_i - \pi_i| = o_p(1),$$

therefore in either cases

$$\sup |\widehat{\pi}_i - \pi_i| = o_p(1). \tag{12}$$

By A1 and A3(i), it follows that $0 = \sum_{i=1}^n \delta_i \zeta_{\omega 1i}(\widehat{a}) K_h(U_i - u) / \widehat{\pi}_i$ is satisfied with probability approaching 1. Then by (12)

$$0 = \left(\frac{h}{n}\right)^{1/2} \sum_{i=1}^n \frac{\delta_i}{\widehat{\pi}_i} \zeta_{1\omega i}(\widehat{a}) K_h(U_i - u) =$$
$$\left(\frac{h}{n}\right)^{1/2} \left[\sum_{i=1}^n \frac{\delta_i}{\pi_i} \zeta_{1\omega i}(\widehat{a}) K_h(U_i - u) + \sum_{i=1}^n \left(\frac{\pi_i - \widehat{\pi}_i}{\pi_i^2} \right) \delta_i \zeta_{1\omega i}(\widehat{a}) K_h(U_i - u) \right] +$$
$$o_p(1) = T_1 + T_2 + o_p(1).$$

Let $\widehat{\xi}_i = -\left(\partial^2 \alpha_0(u) / \partial u^2\right)(U_i - u)^2 / 2 + (\widehat{a}(u) - a_0(u)) W_i$ and note that for $|U_i - u| \leq h$ and A1(iv) $\sup_{|U_i - u| \leq h} \left\| \widehat{\xi}_i \right\| = o_p(1)$, thus by A2

$$T_1 = \left(\frac{h}{n}\right)^{1/2} \sum_{i=1}^n \frac{\delta_i}{\pi_i} \{ \zeta_{1\omega i}(\alpha_0(U_i)) +$$
$$\zeta_{2\omega i}(\alpha_0(U_i)) \widehat{\xi}_i + \left[\zeta_{1\omega i}\left(\alpha_0(U_i) + \widehat{\xi}_i\right) - \zeta_{1\omega i}(\alpha_0(U_i)) \right.$$
$$\left. -\zeta_{2\omega i}(\alpha_0(U_i)) \widehat{\xi}_i \right] \Big\} K_h(U_i - u) = T_{11} + T_{12} + T_{13}.$$

Since

$$T_{12} = \left(\frac{h}{n}\right)^{1/2} \sum_{i=1}^{n} \frac{\delta_i}{\pi_i} \left[-\frac{1}{2} \zeta_{2\omega i} \left(\alpha_0 \left(U_i\right)\right) \frac{\partial^2 \alpha_0 \left(u\right)}{\partial u^2} \left(U_i - u\right)^2 + \right.$$
$$\left. \zeta_{2\omega i} \left(\alpha_0 \left(U_i\right)\right)\right] \left(\widehat{a} - a_0\right) W_i \right] K_h \left(U_i - u\right),$$

iterated expectations and a standard kernel calculation show that

$$E\left[\frac{\delta}{\pi} \zeta_{2\omega i} \left(\alpha_0 \left(U\right)\right) \frac{\partial^2 \alpha_0 \left(u\right)}{\partial u^2} \left(U - u\right)^2 K_h \left(U - u\right)\right] = f\left(u\right) h^2 \begin{bmatrix} v_{12} \\ v_{13} \end{bmatrix} \times$$
$$\otimes \Gamma_{\zeta 0} \left(u\right) \frac{\partial^2 \alpha_0 \left(u\right)}{\partial u^2}, \qquad (13)$$
$$E\left[\frac{\delta}{\pi} \zeta_{2\omega i} \left(\alpha_0 \left(U\right)\right) K_h \left(U - u\right)\right] = f\left(u\right) \begin{bmatrix} 1 & v_{12} \\ v_{12} & v_{12} \end{bmatrix} \otimes \Gamma_{\zeta 0} \left(u\right).$$

By A2 for $E_\varepsilon = \{\|\xi_i\| \leq \varepsilon, i = 1, ..., n\}$

$$\sup_{E_\varepsilon} \|T_{13}\| \leq \sup_{E_\varepsilon} \|\zeta_{1\omega i} \left(\alpha_0 \left(U_i\right) + \xi_i\right) - \zeta_{1\omega i} \left(\alpha_0 \left(U_i\right)\right) - \zeta_{2\omega i} \left(\alpha_0 \left(U_i\right)\right) \xi_i\| \times$$
$$\left(\frac{h}{n}\right)^{1/2} \sum_{i=1}^{n} \left\|\frac{\delta_i}{\pi_i} K_h \left(U_i - u\right)\right\|, \qquad (14)$$

$$E\left[\sup_{E_\varepsilon} \|T_{13}\|\right] \leq \eta_\varepsilon \left(\frac{h}{n}\right)^{1/2} \sum_{i=1}^{n} \left\|\frac{\delta_i}{\pi_i} K_h \left(U_i - u\right)\right\| = o\left(1\right)$$

for $\xi_i \to 0$ as $\varepsilon \to 0$. As $\widehat{\xi} \in E_\varepsilon$ with probability approaching 1, it follows that $\|T_{13}\| = o_p\left(1\right)$. For T_2, if $\widehat{\pi}_i$ is estimated parametrically then $\|T_2\| = O_p\left(n^{-1/2}\right)$, whereas for π_i estimated nonparametrically

$$T_2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \zeta_{1\omega i} \left(\widehat{a}\right) K_h \left(U_i - u\right) \frac{\left(\pi_i - \delta_i\right) L_b \left(V_j - V_i\right)}{n b^{k_2+1} f\left(V_j\right)} + o_p\left(1\right)$$

and $E\left(\left(\pi - \delta\right) L_b \left(V - v\right)\right) = 0$, $var\left(T_2\right) = C/nb^{k_2+1} \to 0$, hence $\|T_2\| = o_p\left(1\right)$. By (13) and (14) the conclusion follows by the continuous mapping theorem noting that for

$$\sum_{i=1}^{n} \frac{d^T \overline{\zeta}_{1\omega i} \left(\alpha_0 \left(U_i\right)\right) K_h \left(U_i - u\right)}{nVar\left(d^T \zeta_{1\omega} \left(\alpha_0 \left(U\right)\right) K_h \left(U - u\right)\right)} := \sum_{i=1}^{n} Z_{in},$$

where $\overline{\zeta}_{1\omega i} \left(\alpha_0 \left(U_i\right)\right) K_h \left(U_i - u\right), = \zeta_{1\omega i} \left(\alpha_0 \left(U_i\right)\right) K_h \left(U_i - u\right) - E\left[\zeta_{1\omega i} \left(\alpha_0 \left(U_i\right)\right) K_h \left(U_i - u\right)\right]$, A3(iii) implies that

$$\sum_{i=1}^{n} |Z_{in}|^{2+\beta} = \frac{C}{\left(hVar\left(d^T \zeta_{1\omega} \left(\alpha_0 \left(U\right)\right) K_h \left(U - u\right)\right)\right)^{1+\beta/2} \left(nh\right)^{\beta/2}} E\left|K_h \left(U - u\right)\right|^{2+\beta} \to 0$$

as $nh \to \infty$, hence the Lyapunov central limit theorem is satisfied and

$$\left(\frac{h}{n}\right)^{1/2} \sum_{i=1}^{n} \frac{\delta_i}{\pi_i} \overline{\zeta}_{1\omega i} \left(\alpha_0 \left(U_i\right)\right) K_h \left(U_i - u\right) \xrightarrow{d} N\left(0, f\left(u\right) \begin{bmatrix} v_{20} & v_{21} \\ v_{21} & v_{22} \end{bmatrix} \otimes \Sigma_{\pi \zeta 0} \left(u\right)\right).$$

Proof (Proof of Theorem 2) Note that

$$\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}\zeta_{2\omega i}\left(\widehat{a}\right)K_h\left(U_i-u\right)=\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\zeta_{2\omega i}\left(\alpha_0\left(U\right)\right)K_h\left(U_i-u\right)+$$
$$\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\left[\zeta_{2\omega i}\left(\alpha_0\left(U_i\right)+\widehat{\xi}_i\right)-\zeta_{2\omega i}\alpha_0\left(U_i\right)\right]K_h\left(U_i-u\right)+o_p\left(1\right)=$$
$$T_3+T_4+o_p\left(1\right)$$

where $\widehat{\xi}_i=-\left(\partial^2\alpha_0\left(u\right)/\partial u^2\right)\left(U_i-u\right)^2/2+\left(\widehat{a}-a_0\right)W_i$ and

$$\sup_{\left|U_i-u\right|\leq h}\left\|\widehat{\xi}_i\right\|=O_p\left(h^2+\left(nh\right)^{-1/2}\right)\tag{15}$$

by A7. As in the proof of Theorem 1

$$\left\|T_3-nf\left(u\right)\begin{bmatrix}1 & v_{11}\\ v_{11} & v_{12}\end{bmatrix}\otimes\mathit{\Gamma}_{\zeta 0}\left(u\right)\right\|=o_p\left(1\right),$$

and $\left\|T_4\right\|=o_p\left(n\right)$. Note also that

$$\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}\zeta_{1\omega i}\left(\widehat{a}\right)K_h\left(U_i-u\right)=\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}\zeta_{1\omega i}\left(\alpha_0\left(U_i\right)\right)K_h\left(U_i-u\right)+$$
$$\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}\zeta_{2\omega i}\left(\alpha_0\left(U_i\right)\right)\widehat{\xi}_iK_h\left(U_i-u\right)+\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}\left[\zeta_{1\omega i}\left(\alpha_0\left(U_i\right)+\widehat{\xi}_i\right)-\zeta_{1\omega i}\left(\alpha_0\left(U_i\right)\right)-\right.$$
$$\left.\zeta_{2\omega i}\left(\alpha_0\left(U_i\right)\right)\widehat{\xi}_i\right]K_h\left(U_i-u\right)=T_5+T_6+T_7.$$

As in the proof of Theorem 1

$$T_5=\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\zeta_{1\omega i}\left(\alpha_0\left(U_i\right)\right)K_h\left(U_i-u\right)+o_p\left(\left(\left(\frac{h}{n}\right)^{1/2}\right)\right),$$
$$T_6=-nf\left(u\right)\frac{h^2}{2}\begin{bmatrix}v_{12}\\ v_{13}\end{bmatrix}\otimes\mathit{\Gamma}_{\xi 0}\left(u\right)\frac{\partial^2\alpha_0\left(u\right)}{\partial u^2}+nf\left(u\right)\begin{bmatrix}1 & v_{11}\\ v_{11} & v_{12}\end{bmatrix}\otimes\mathit{\Gamma}_{\zeta 0}\left(u\right)\times$$
$$\begin{bmatrix}\widehat{a}_1-a_1\\ h\left(\widehat{a}_2-a_2\right)\end{bmatrix}+o_p\left(n\right),$$
$$\left\|T_7\right\|\leq\sum_{i=1}^{n}\left\|\frac{\delta_i}{\pi_i}\left[\zeta_{1\omega i}\left(\alpha_0\left(U_i\right)+\widehat{\xi}_i\right)-\zeta_{1\omega i}\left(\alpha_0\left(U_i\right)\right)-\zeta_{2\omega i}\left(\alpha_0\left(U_i\right)\right)\widehat{\xi}_i\right]K_h\left(U_i-u\right)\right\|=$$
$$o_p\left(n\left\|\widehat{\xi}_i\right\|\right)$$

by A7 and (15). A simple rearrangement of the above expressions yields

$$
\begin{aligned}
&(nh)^{1/2}\begin{bmatrix}\widetilde{a}_1-a_1\\ h\left(\widetilde{a}_2-a_2\right)\end{bmatrix}=(nh)^{1/2}\begin{bmatrix}\widehat{a}_1-a_1\\ h\left(\widehat{a}_2-a_2\right)\end{bmatrix}-\\
&\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}\zeta_{2\omega i}\left(Y_i,\eta\left(X_i,\widehat{a}\right)\right)K_h\left(U_i-u\right)\right]^{-1}\times\\
&\left(\frac{h}{n}\right)^{1/2}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}\zeta_{1\omega}\left(Y_i,\eta\left(X_i,\widehat{a}\right)\right)K_h\left(U_i-u\right)=\\
&\begin{bmatrix}\widehat{a}_1-a_1\\ h\left(\widehat{a}_2-a_2\right)\end{bmatrix}-\left\{f\left(u\right)\begin{bmatrix}1 & v_{11}\\ v_{11} & v_{12}\end{bmatrix}\otimes\mathit{\Gamma}_{\zeta 0}\left(u\right)\right\}^{-1}\times\left(\frac{h}{n}\right)^{1/2}\\
&\left[-nf\left(u\right)\frac{h^2}{2}\begin{bmatrix}v_{12}\\ v_{13}\end{bmatrix}\otimes\mathit{\Gamma}_{\zeta 0}\left(u\right)\frac{\partial^2\alpha_0\left(u\right)}{\partial u^2}+f\left(u\right)\begin{bmatrix}1 & v_{11}\\ v_{11} & v_{12}\end{bmatrix}\otimes\mathit{\Gamma}_{\zeta 0}\left(u\right)\times\right.\\
&\left.\begin{bmatrix}\widehat{a}_1-a_1\\ h\left(\widehat{a}_2-a_2\right)\end{bmatrix}+\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\zeta_{1\omega i}\left(\alpha_0\left(U_i\right)\right)K_h\left(U_i-u\right)\right]+o_p\left(h^2+(nh)^{-1/2}\right)=
\end{aligned}
$$

$$
\begin{aligned}
&\left[I-\left\{f\left(u\right)\begin{bmatrix}1 & v_{11}\\ v_{11} & v_{12}\end{bmatrix}\otimes\mathit{\Gamma}_{\zeta 0}\left(u\right)\right\}^{-1}f\left(u\right)\begin{bmatrix}1 & v_{11}\\ v_{11} & v_{12}\end{bmatrix}\otimes\mathit{\Gamma}_{\zeta 0}\left(u\right)\right]\begin{matrix}\widehat{a}_1-a_1\\ h\left(\widehat{a}_2-a_2\right)\end{matrix}+\\
&\left\{f\left(u\right)\begin{bmatrix}1 & v_{11}\\ v_{11} & v_{12}\end{bmatrix}\otimes\mathit{\Gamma}_{\zeta 0}\left(u\right)\right\}^{-1}\left[-(nh)^{1/2}f\left(u\right)\frac{h^2}{2}\begin{bmatrix}v_{12}\\ v_{13}\end{bmatrix}\otimes\mathit{\Gamma}_{\zeta 0}\left(u\right)\frac{\partial^2\alpha_0\left(u\right)}{\partial u^2}+\right.\\
&\left.\left(\frac{h}{n}\right)^{1/2}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\zeta_{1\omega i}\left(\alpha_0\left(U_i\right)\right)K_h\left(U_i-u\right)\right]+o_p\left(h^2+(nh)^{-1/2}\right),
\end{aligned}
$$

that is

$$
\begin{aligned}
&(nh)^{1/2}\begin{bmatrix}\widetilde{a}_1-a_1\\ h\left(\widetilde{a}_2-a_2\right)\end{bmatrix}=\left\{f\left(u\right)\begin{bmatrix}1 & v_{11}\\ v_{11} & v_{12}\end{bmatrix}\otimes\mathit{\Gamma}_{\zeta 0}\left(u\right)\right\}^{-1}\times\\
&\left[-(nh)^{1/2}f\left(u\right)\frac{h^2}{2}\begin{bmatrix}v_{12}\\ v_{13}\end{bmatrix}\otimes\mathit{\Gamma}_{\zeta 0}\left(u\right)\frac{\partial^2\alpha_0\left(u\right)}{\partial u^2}+\left(\frac{h}{n}\right)^{1/2}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\zeta_{1\omega i}\left(\alpha_0\left(U_i\right)\right)K_h\left(U_i-u\right)\right]+\\
&o_p\left(h^2+(nh)^{-1/2}\right)
\end{aligned}
$$

and the conclusion follows as in Theorem 1.

Proof (Proof of Theorem 3) By A1' and A3'(i) it follows that

$$
\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}\mu_{\omega i}\left(\widehat{a}\right)\eta_i'\left(\widehat{a}\right)K_h\left(U_i-u\right)=0
$$

is satisfied with probability approaching 1. Then by (12)

$$
\begin{aligned}
0 = & \left(\frac{h}{n}\right)^{1/2} \sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}_i} \mu_{\omega i}\left(\widehat{a}\right) \eta_i'\left(\widehat{a}\right) K_h\left(U_i-u\right) = \\
& \left(\frac{h}{n}\right)^{1/2} \left[\sum_{i=1}^{n} \frac{\delta_i}{\pi_i} \mu_{\omega i}\left(\widehat{a}\right) \eta_i'\left(\widehat{a}\right) K_h\left(U_i-u\right) + \right. \\
& \left. \sum_{i=1}^{n} \left(\frac{\pi_i-\widehat{\pi}_i}{\pi_i^2}\right) \delta_i \mu_{\omega i}\left(\widehat{a}\right) \eta_i'\left(\widehat{a}\right) K_h\left(U_i-u\right)\right] \\
= & T_8 + T_9 + o_p\left(1\right).
\end{aligned}
$$

Let $\widehat{\xi}_i = -\left(\partial^2 \alpha_0\left(u\right)/\partial u^2\right)\left(U_i-u\right)^2/2 + \left(\widehat{a}-a_0\right) W_i$; as in the proof of Theorem 1 $\sup_{|U_i-u|\leq h}\left\|\widehat{\xi}_i\right\| = o_p\left(1\right)$, hence by A1'(iv) and A2'

$$
\begin{aligned}
T_8 = & \left(\frac{h}{n}\right)^{1/2} \sum_{i=1}^{n} \frac{\delta_i}{\pi_i} \left\{\mu_{\omega i}\left(\alpha_0\left(U_i\right)\right) \eta_i'\left(\alpha_0\left(U_i\right)\right) + \right. \\
& \mu_{1\omega i}\left(\alpha_0\left(U_i\right)\right)\widehat{\xi}_i + \left[\mu_{\omega i}\left(\alpha_0\left(U_i\right)+\widehat{\xi}_i\right)\eta_i'\left(\alpha_0\left(U_i\right)+\widehat{\xi}_i\right) - \mu_{\omega i}\left(\alpha_0\left(U_i\right)\right)\right. \\
& \left.\left. -\mu_{1\omega i}\left(\alpha_0\left(U_i\right)\right)\widehat{\xi}_i\right]\right\} K_h\left(U_i-u\right) = T_{81}+T_{82}+T_{83}.
\end{aligned}
$$

The same arguments as those used in the proof of Theorem 1 show that $\left\|T_{8j}\right\| = o_p\left(1\right)$ $(j=2,3)$ and by the Lyapunov central limit theorem

$$
T_{81} \xrightarrow{d} N\left(0, f\left(u\right)\begin{bmatrix} v_{20} & v_{21} \\ v_{21} & v_{22}\end{bmatrix} \otimes \Sigma_{\pi\mu 0}\left(u\right)\right).
$$

The conclusion follows by the continuous mapping theorem and simple algebra.

Proof (Proof of Theorem 4) The proof is similar to that of Theorem 2, hence omitted.

Proof (Proof of Theorem 5) We first consider the case where π_i is estimated nonparametrically. Recall that $\widehat{\pi}_i - \pi_i = \sum_{j=1}^{n} w_{ji}\left(\delta_j^Y - \pi_j\right)$, where

$$
w_{ji} = \frac{\sum_{j=1}^{n} \delta_j^Y L_b\left(V_j-V_i\right)\omega\left(X_j\right)}{\sum_{j=1}^{n} L_b\left(V_j-V_i\right)\omega\left(X_j\right)}
$$

are robust kernel weights, $V_i = \left[X_i^T, U_i\right]^T$ and $L_b\left(\cdot\right) = L\left(\cdot/b\right)/b^{k+1}$. Then

$$
\begin{aligned}
n^{1/2}\left(\widehat{\mu}_{IPW}^{NP} - \mu_0\right) = & \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(\frac{\delta_i^Y Y_i}{\pi_i} - \mu_0\right) + \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(\frac{\delta_i^Y Y_i - E\left(\delta_i^Y Y_i | X_i, U_i\right)}{\pi_i \widehat{\pi}_i}\right)\left(\pi_i - \widehat{\pi}_i\right) + \\
& \frac{1}{n^{1/2}} \sum_{i=1}^{n} \frac{E\left(\delta_i^Y Y_i | X_{i_i}, U_i\right)}{\pi_i \widehat{\pi}_i}\left(\pi_i - \widehat{\pi}_i\right) = \sum_{j=1}^{3} T_{9j},
\end{aligned}
$$

and

$$T_{92} = -\frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(\frac{\delta_i^Y Y_i - E\left(\delta_i^Y Y_i | X_i, U_i\right)}{\pi_i^2} \right) \sum_{j=1}^{n} w_{ji} \left(\delta_j^Y - \pi_j\right) + o_p(1),$$

$$T_{93} = -\frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(\delta_i^Y - \pi_i\right) \frac{E\left(Y_i | X_{i_i}, U_i\right)}{\pi_i} -$$

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(\delta_i^Y - \pi_i\right) \sum_{j=1}^{n} \omega_{ji} \left(\frac{E\left(Y_j | X_{j_i}, U_j\right)}{\pi_j} - \frac{E\left(Y_i | X_{2i_i}, U_i\right)}{\pi_i} \right) + o_p(1) = T_{931} + T_{932}.$$

Note that T_{92} is a degenerate second order U statistic with zero mean and variance of order $O\left(1/nb^{k+1}\right)$, and $E\left(T_{932}\right) = 0$ and $Var\left(T_{932}\right) = c/nb^{k+1} \to 0$ hence $\left|T_{932}\right| = o_p(1)$. Thus

$$n^{1/2}\left(\widehat{\mu}_{IPW}^{NP} - \mu_0\right) = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(\frac{\delta_i^Y Y_i}{\pi_i} - \mu_0 \right) - \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(\delta_i^Y - \pi_i\right) \frac{E\left(Y_i | X_{i_i}, U_i\right)}{\pi_i} + o_p(1)$$

and the first conclusion follows by the standard central limit theorem, noting that $E\left(T_{91}\right)^2 = E\left(E\left(Y|XU\right)^2/\pi\right) + E\left(\sigma^2\left(X,U\right)/\pi\right) - \mu_0^2$, $E\left({T_{931}}^2\right) = E\left(E\left(Y|X,U\right)^2/\pi\right) - E\left(E\left(Y|X,U\right)^2\right)$ and $Cov\left(T_{91}, T_{931}\right) = -E\left(E\left(Y|X,U\right)^2/\pi\right) + E\left(E\left(Y|X,U\right)^2\right)$. For π_i estimated parametrically

$$\begin{aligned}
n^{1/2}\left(\widehat{\mu}_{IPW}^{P} - \mu_0\right) &= \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(\frac{\delta_i^Y Y_i}{\pi_i} - E\left(Y_i | X_{_i}, U_i\right) \right) + \frac{1}{n^{1/2}} \sum_{i=1}^{n} \frac{\delta_i^Y Y}{\pi_i^2} \left(\pi_i - \widehat{\pi}_i\right) + \\
&\quad \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(E\left(Y_i | X_{i_i}, U_i\right) - \mu_0\right) + o_p(1) \\
&= \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(\frac{\delta_i^Y Y_i}{\pi_i} - E\left(Y_i | X_{i_i}, U_i\right) \right) - \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i^Y Y}{\pi_i^2} \frac{\partial \overline{\pi}_i}{\partial \gamma^T} n^{1/2}\left(\widehat{\gamma} - \gamma_0\right) + \\
&\quad \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(E\left(Y_i | X_{2i_i}, U_i\right) - \mu_0\right) + o_p(1) = \sum_{j=1}^{3} T_{10j},
\end{aligned}$$

where $\overline{\pi}_i = \pi\left(X_{i_i}, U_i, \overline{\gamma}\right)$ and $\overline{\gamma}$ is the mean value. Thus the conclusion follows by the law of large numbers and the standard central limit theorem, noting that by iterated expectations

$$\begin{aligned}
Cov\left(T_{101}, T_{102}\right) &= E\left(\frac{\delta_i^Y Y_i}{\pi_i} E\left(\frac{E\left(Y | X_{_i}, U\right)}{\pi} \frac{\partial \pi}{\partial \gamma^T} \right) M^{-1}\left(\gamma_0\right) r_{i\omega}\left(\gamma_0\right) \right) \\
&= E\left(E\left(Y|X,U\right) E\left(\frac{E\left(Y|X,U\right)}{\pi} \frac{\partial \pi}{\partial \gamma^T} \right) M^{-1}\left(\gamma_0\right) r_{\omega}\left(\gamma_0\right) \right)
\end{aligned}$$

and $Cov\left(T_{101},T_{103}\right)=0$ and $Cov\left(T_{102},T_{103}\right)=0$ by a direct calculation. For $n^{1/2}\left(\widehat{\mu}_{DR}^{NP}-\mu_0\right)$ with π_i estimated nonparametrically, note that

$$\begin{aligned}n^{1/2}\left(\widehat{\mu}_{DR}^{NP}-\mu_0\right)=&\frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{\delta_i^Y Y_i}{\widehat{\pi}_i}-\mu_0+\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(1-\frac{\delta_i^Y}{\pi_i}\right)\left(\eta\left(X_i,\widehat{\alpha}\left(U_i\right)\right)-\eta\left(X_i,\alpha_0\left(U_i\right)\right)\right)+\\&\frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{\delta_i^Y}{\widehat{\pi}_i\pi_i}\left(\pi_i-\widehat{\pi}_i\right)\left(\eta\left(X_i,\widehat{\alpha}\left(U_i\right)\right)-\eta\left(X_i,\alpha_0\left(U_i\right)\right)\right)+\\&\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(1-\frac{\delta_i^Y}{\pi_i}\right)\eta\left(X_i,\alpha_0\left(U_i\right)\right)+\frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{\delta_i^Y}{\widehat{\pi}_i\pi_i}\left(\pi_i-\widehat{\pi}_i\right)\eta\left(X_i,\alpha_0\left(U_i\right)\right)=\sum_{j=1}^{5}T_{11j}.\end{aligned}$$

Then for T_{112}

$$\left|T_{112}\right|\leq\sup_{X_i,E_\epsilon}\left|\eta\left(X_i,\alpha_0\left(U_i\right)+\widehat{\xi}_i\right)-\eta\left(X_i,\alpha_0\left(U_i\right)\right)\right|\left|\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(1-\frac{\delta_i^Y}{\pi_i}\right)\right|=o_p\left(1\right),$$

by continuous mapping theorem, where $E_\varepsilon=\left\{\left\|\xi_i\right\|\leq\varepsilon,i=1,...,n\right\}$, $\widehat{\xi}_i=-\left(\partial^2\alpha_0\left(U_j\right)/\partial U_j^2\right)\left(U_j-U_i\right)^2/2+\left(\widehat{a}\left(U_i\right)-a_0\left(U_i\right)\right)W_i$ and $\widehat{\xi}_i\in E_\epsilon$ for $\epsilon\to 0$. Similarly $\left|T_{113}\right|=o_p\left(1\right)$, whereas by iterated expectations $E\left[\left(1-\delta_i^Y/\pi_i\right)\eta'\left(X_i,\alpha_0\left(U_i\right)\right)\right]=0$ and $Var\left(T_{114}\right)=C/nb^{k+1}\to 0$, hence $\left|T_{114}\right|=o_p\left(1\right)$; for T_{115}

$$T_{115}=\frac{1}{n^{1/2}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\delta_j-\pi_j\right)\frac{\delta_i^Y\eta\left(X_i,\alpha_0\left(U_i\right)\right)L_b\left(V_j-V_i\right)}{\pi_i nb^{k+1}f\left(V_j\right)}+o_p\left(1\right)$$

with $E\left(T_{115}\right)=0$ and $Var\left(T_{115}\right)=C/nb^{k_2+1}\to 0$, hence $\left|T_{115}\right|=o_p\left(1\right)$. Thus $n^{1/2}\left(\widehat{\mu}_{DR}^{NP}-\mu_0\right)=n^{1/2}\left(\widehat{\mu}_{IPW}^{NP}-\mu_0\right)+o_p\left(1\right)$ and the conclusion follows. Finally

$$\begin{aligned}n^{1/2}\left(\widehat{\mu}_{DR}^{P}-\mu_0\right)=&\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(\frac{\delta_i^Y Y_i}{\pi_i}-\eta\left(X_i,\alpha_0\left(U_i\right)\right)\right)+\frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{\delta_i^Y Y_i}{\widehat{\pi}_i\pi_i}\left(\pi_i-\widehat{\pi}_i\right)+\\&\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(\eta\left(X_i,\alpha_0\left(U_i\right)\right)-\mu_0\right)+\\&\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(1-\frac{\delta_i^Y}{\pi_i}\right)\left(\eta\left(X_i,\widehat{\alpha}\left(U_i\right)\right)-\eta\left(X_i,\alpha_0\left(U_i\right)\right)\right)+\\&\frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{\delta_i^Y}{\widehat{\pi}_i\pi_i}\left(\pi_i-\widehat{\pi}_i\right)\left(\eta\left(X_i,\widehat{\alpha}\left(U_i\right)\right)-\eta\left(X_i,\alpha_0\left(U_i\right)\right)\right)+o_p\left(1\right):=\sum_{j=1}^{5}T_{12j}.\end{aligned}$$

Then it is easy to see that $\left|T_{124}\right|=o_p\left(1\right)$ and

$$\left|T_{125}\right|\leq\sup_{X_i,E_\epsilon}\left|\eta\left(X_i,\alpha_0\left(U_i\right)+\widehat{\xi}_i\right)-\eta\left(X_i,\alpha_0\left(U_i\right)\right)\right|\frac{1}{n}\sum_{i=1}^{n}\sup_{\gamma\in\Gamma}\left\|\frac{\delta_i^Y}{\pi_i^2}\frac{\partial\pi_i}{\partial\gamma^T}\right\|\left\|n^{1/2}\left(\widehat{\gamma}-\gamma_0\right)\right\|=o_p\left(1\right),$$

hence $n^{1/2}\left(\widehat{\mu}_{DR}^{P}-\mu_0\right)=n^{1/2}\left(\widehat{\mu}_{IPW}^{P}-\mu_0\right)+o_p\left(1\right)$ and the conclusion follows immediately.

Proof (Proof of Proposition 1) By the results of Theorems 1-4 and the continuous mapping theorem

$$(nh)^{1/2}\left(R\widetilde{\widehat{\alpha}}^{*}_{\widehat{\pi}}(u)-r_0(u)\right)\xrightarrow{d}N\left(\gamma(u),\left(RS\Gamma^{v_1}_{\times 0}(u)^{-1}\frac{\Sigma^{v_2}_{\pi\times 0}(u)}{f(u)}\Gamma^{v_1}_{y0}(u)^{-1}S^TR^T\right)\right),$$

hence the conclusion follows by standard results on quadratic forms in nonzero mean asymptotically normal random vectors. The second conclusion follows immediately under the assumption that $(nh)^{1/2}\gamma_n(u)\to\infty$.

Proof (Proof of Proposition 2) Only the case where $\widetilde{\widehat{a}}^{*}_{1\widehat{\pi}}=\widehat{\alpha}^{M}_{\widehat{\pi}}(u)$ is considered in the proof, as the other cases can be dealt with in the same way. Note that for any $1\le j\ne k\le s$ by iterated expectations and a standard kernel calculation

$$\begin{aligned}&Cov\left((nh)^{1/2}\left(\widehat{\alpha}^{M}_{\widehat{\pi}}(u_j)-\alpha_0(u_j)\right),(nh)^{1/2}\left(\widehat{\alpha}^{M}_{\widehat{\pi}}(u_k)-\alpha_0(u_k)\right)\right)=\\
&E\left[\frac{1}{h}S\left(f(u_j)\Gamma^{v_1}_{\zeta 0}(u_j)\right)^{-1}\frac{\delta_i}{\pi_i^2}\zeta_{1\omega i}(\alpha_0(u_i))^{\otimes 2}K_h(u_i-u_j)K_h(u_i-u_k)\times\right.\\
&\left.S\left(f(u_k)\Gamma^{v_1}_{\zeta 0}(u_k)\right)^{-1T}\right]=\\
&S\left(f(u_j)\Gamma^{v_1}_{\zeta 0}(u_j)\right)^{-1}\Sigma^{v_2}_{\pi\zeta 0}(U)\int K_h(U-u_j)K_h(U-u_k)f(U)\,dU\times\\
&\left(S\left(f(u_k)\Gamma^{v_1}_{\zeta 0}(u_k)\right)^{-1}\right)^T,\end{aligned}$$

and

$$\begin{aligned}&\int K_h(U-u_j)K_h(U-u_k)f(U)\,dU=\int K(v)K\left(v+\frac{u_j-u_k}{h}\right)f(u_j+hv)\,dv=\\
&\int_{|v|<\frac{u_j-u_k}{h}}K(v)K\left(v+\frac{u_j-u_k}{h}\right)f(u_j+hv)\,dv+\\
&\int_{|v|\ge\frac{u_j-u_k}{2h}}K(v)K\left(v+\frac{u_j-u_k}{h}\right)f(u_j+hv)\,dv\le\\
&\sup_{|w|>\frac{u_j-u_k}{2h}}|K(w)|\sup_w|f(w)|\int|K(v)|\,dv+\sup_{|v|\ge\frac{u_j-u_k}{2h}}|K(v)|\sup_v|f(v)|\int|K(v)|\,dv\\
&=T_{20}+T_{21},\end{aligned}$$

where $w:=v+(u_j-u_k)/h>(u_j-u_k)/2h$ and $\sup_{|\cdot|>(u_j-u_k)/2h}|K(\cdot)|$, for $\cdot=v$ or w, and $T_{20},T_{21}\to 0$ as $h\to 0$ by A5(i), hence

$$Cov\left((nh)^{1/2}\left(\widehat{\alpha}^{M}_{\widehat{\pi}}(u_j)-\alpha_0(u_j)\right),(nh)^{1/2}\left(\widehat{\alpha}^{M}_{\widehat{\pi}}(u_k)-\alpha_0(u_k)\right)\right)=o(1).$$

Therefore by the Cramer-Wold device and the continuous mapping theorem

$$(nh)^{1/2}\begin{bmatrix}R\widehat{\alpha}^{M}_{\widehat{\pi}}(u_1)-r_0(u_1)\\ \vdots\\ R\widehat{\alpha}^{M}_{\widehat{\pi}}(u_s)-r_0(u_s)\end{bmatrix}\xrightarrow{d}N\left(\begin{bmatrix}\gamma(u_1)\\ \vdots\\ \gamma(u_s)\end{bmatrix},diag\left[RS\Omega_{\zeta 0}(u_1)S^TR^T,...,RS\Omega_{\zeta 0}(u_s)S^TR^T\right]\right),$$

where

$$\Omega_{\zeta 0}\left(u_j\right)=\Gamma_{\zeta 0}^{v_1}\left(u_j\right)^{-1} \Sigma_{\pi\zeta 0}^{\nu_2}\left(u_j\right) \Gamma_{\zeta 0}^{v_1}\left(u_j\right)^{-1},\ j=1,\ldots,s$$

and the conclusion follows again by the continuous mapping theorem.

Proof (Proof of Proposition 3) Note that

$$D=-2\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\left(\zeta_\omega\left(Y_i,\eta\left(X_i,\widehat{a}_{1\widehat{\pi}}^{M}\right)\right)-\zeta_\omega\left(Y_i,\eta\left(X_i,\alpha_0\left(u\right)\right)\right)\right)K_h\left(U_i-u\right)-$$
$$2\sum_{i=1}^{n}\left(\delta_i\frac{\left(\widehat{\pi}_i-\pi_i\right)}{\pi_i\widehat{\pi}_i}\right)\left(\zeta_\omega\left(Y_i,\eta\left(X_i,\widehat{a}_{1\widehat{\pi}}^{M}\right)\right)-\zeta_\omega\left(Y_i,\eta\left(X_i,\alpha_0\left(u\right)\right)\right)\right)K_h\left(U_i-u\right)$$
$$=T_{22}+T_{23}.$$

For T_{22}, a second order Taylor expansion about $\widehat{a}_{1\widehat{\pi}}^{M}$ with Lagrange remainder $\overline{\xi}_i$ shows that

$$\begin{aligned}T_{22}=&\,nh\left(\widehat{a}_{1\widehat{\pi}}^{M}-\alpha_0\left(u\right)\right)^T\frac{1}{nh}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\left[\zeta_{2\omega i}\left(\alpha_0\left(U\right)\right)+\right.\\ &\left.\left(\zeta_{2\omega i}\left(\alpha_0\left(U_i\right)+\overline{\xi}_i\right)-\zeta_{2\omega i}\alpha_0\left(U_i\right)\right)\right]K_h\left(U_i-u\right)\left(\widehat{a}_{1\widehat{\pi}}^{M}-\alpha_0\left(u\right)\right)+o_p\left(1\right)=\\ &nh\left(\widehat{a}_{1\widehat{\pi}}^{M}-\alpha_0\left(u\right)\right)^T\frac{1}{nh}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\zeta_{2\omega i}\left(\alpha_0\left(U\right)\right)K_h\left(U_i-u\right)\left(\widehat{a}_{1\widehat{\pi}}^{M}-\alpha_0\left(u\right)\right)+o_p\left(1\right)=\\ &nh\left(\widehat{a}_{1\widehat{\pi}}^{M}-\alpha_0\left(u\right)\right)^T f\left(u\right)\Gamma_{\zeta 0}\left(u\right)\left(\widehat{a}_{1\widehat{\pi}}^{M}-\alpha_0\left(u\right)\right)+o_p\left(1\right),\end{aligned}$$

where the first equality follows by $\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\zeta_{1\omega i}\left(\widehat{a}_{1\widehat{\pi}}^{M}\right)/\left(nh\right)^{1/2}=0$ and the second by the same arguments as those used in the proof of Theorem 1. The conclusion follows by a standard result in probability theory stating that if $v\sim N\left(0,\Omega\right)$, then, for any real symmetric matrix Q, $v'Qv\sim\sum_{j=1}^{\dim(v)}\lambda_j\chi_j^2\left(1\right)$, where the λ_j's are the eigenvalue of $Q\Omega$ (here $\Omega=\Gamma_{\zeta 0}\left(u\right)^{-1}\Sigma_{\pi\zeta 0}\left(u\right)\Gamma_{\zeta 0}\left(u\right)^{-1}/f\left(u\right)$ and $Q=f\left(u\right)\Gamma_{\zeta 0}\left(u\right)$). For T_{23}, note that $\left|T_{23}\right|\leq\sup\left|\widehat{\pi}_i-\pi_i\right|\left|T_{22}\right|=o_p\left(1\right)O_p\left(1\right)=o_p\left(1\right)$ by (12), hence $D=T_{22}+o_p\left(1\right)$ and the conclusion follows.

Proof (Proof of Proposition 4) Let $\eta'\left(X_i^T\alpha(U_i)\right) := \eta'(\alpha(U_i))$, $s_\alpha(X_i, U_i) := s(\alpha(U_i))$; by $\left\|\widehat{\beta}-\beta_0\right\| = o_p(1)$, $\left|\widehat{\phi}-\phi\right| = o_p(1)$

$$\begin{aligned}\widehat{\Gamma}_\mu(u) &= \frac{1}{n}\sum_{i=1}^n \eta'(\widehat{a}_{1\widehat{\pi}})^T V_i(\eta,\beta_0)^{-1}\,\Omega_i\widehat{\Pi}(\delta_i, X_i.U_i)\, s(\widehat{a}_{1\widehat{\pi}}) \times \\ &\quad \frac{\partial}{\partial\alpha}\psi\left(A_i^{-1/2}(Y_i-\eta(\widehat{a}_{1\widehat{\pi}}))\right)K_h(U_i-u) + o_p(1) \\ &= \frac{1}{n}\sum_{i=1}^n \eta'(\widehat{a}_{1\widehat{\pi}})^T V_i(\eta,\beta_0)^{-1}\,\Omega_i\Pi(\delta_i, X_i.U_i)\, s(\widehat{a}_{1\widehat{\pi}}) \times \\ &\quad \frac{\partial}{\partial\alpha}\psi\left(A_i^{-1/2}(Y_i-\eta(\widehat{a}_{1\widehat{\pi}}))\right)K_h(U_i-u) + \\ &\quad \frac{1}{n}\sum_{i=1}^n \eta'(\widehat{a}_{1\widehat{\pi}})^T V_i(\eta,\beta_0)^{-1}\,\Omega_i\left(\widehat{\Pi}(\delta_i, X_i.U_i) - \Pi(\delta_i, X_i.U_i)\right) s(\widehat{a}_{1\widehat{\pi}}) \times \\ &\quad \frac{\partial}{\partial\alpha}\psi\left(A_i^{-1/2}(Y_i-\eta(\widehat{a}_{1\widehat{\pi}}))\right)K_h(U_i-u) + o_p(1) = T_{24} + T_{25} + o_p(1)\,.\end{aligned}$$

For T_{24}

$$\begin{aligned}T_{24} &= \frac{1}{n}\sum_{i=1}^n \eta'(\alpha_0(U_i))^T V_i(\eta,\beta_0)^{-1}\,\Omega_i\Pi(\delta_i, X_i.U_i)\, s(\alpha_0(U_i)) \times \\ &\quad \frac{\partial}{\partial\alpha}\psi\left(A_i^{-1/2}(Y_i-\eta(\alpha_0(U_i)))\right)K_h(U_i-u) + \\ &\quad \frac{1}{n}\sum_{i=1}^n \left[\eta'\left(\alpha_0(U_i)+\widehat{\xi}_i\right)^T V_i(\eta,\beta_0)^{-1}\,\Omega_i\Pi(\delta_i, X_i.U_i)\, s\left(\alpha_0(U_i)+\widehat{\xi}_i\right) \times\right. \\ &\quad \frac{\partial}{\partial\alpha}\psi\left(A_i^{-1/2}\left(Y_i-\eta\left(\alpha_0(U_i)+\widehat{\xi}_i\right)\right)\right) - \eta'(\alpha_0(U_i))^T V_i(\eta,\beta_0)^{-1}\,\Omega_i \times \\ &\quad \left.\Pi(\delta_i, X_i.U_i)\, s(\alpha_0(U_i))\,\frac{\partial}{\partial\alpha}\psi\left(A_i^{-1/2}(Y_i-\eta(\alpha_0(U_i)))\right)\right]K_h(U_i-u) + o_p(1) = \\ &\quad T_{241} + T_{242} + o_p(1)\,,\end{aligned} \tag{16}$$

and

$$\begin{aligned}&E\left\{\sup_{\|\xi_i\|\le\varepsilon}\left\|\left\|\eta'(\alpha_0(U_i)+\xi_i)^T V_i(\eta,\beta_0)^{-1}\,\Omega_i\Pi(\delta_i, X_i.U_i)\, s(\alpha_0(U_i)+\xi_i) \times\right.\right.\right. \\ &\frac{\partial}{\partial\alpha}\psi\left(A_i^{-1/2}(Y_i-\eta(\alpha_0(U_i)+\xi_i))\right) - \eta'(\alpha_0(U_i))^T V_i(\eta,\beta_0)^{-1}\,\Omega_i \times \\ &\left.\left.\left.\Pi(\delta_i, X_i.U_i)\, s(\alpha_0(U_i))\,\frac{\partial}{\partial\alpha}\psi\left(A_i^{-1/2}(Y_i-\eta(\alpha_0(U_i)))\right)\right]K_h(U_i-u)\right\|^2\right\} = \\ &O\left(\frac{1}{nh}\right) + o(1)\end{aligned}$$

hence $Var\left(T_{242}\right)=o\left(1\right)$ and $\left\|T_{242}\right\|=o_{p}\left(1\right)$ since $\widehat{\xi}_{i}\in\left\|\xi_{i}\right\|\leq\varepsilon$ with probability approaching 1. For T_{25}, (12) and an expansion similar to that used in (16) show that

$$\begin{aligned}\left\|T_{25}\right\| \leq \sup_{i}\left\|\widehat{\Pi}\left(\delta_{i},X_{i}.U_{i}\right)-\Pi\left(\delta_{i},X_{i}.U_{i}\right)\right\| \Bigg\{ \Bigg\| \frac{1}{n}\sum_{i=1}^{n}\eta'\left(\alpha_{0}\left(U_{i}\right)\right)^{T}V_{i}\left(\eta,\beta_{0}\right)^{-1}\Omega_{i}s\left(\alpha_{0}\left(U_{i}\right)\right)\times \\ \frac{\partial}{\partial\alpha}\psi\left(A_{i}^{-1/2}\left(Y_{i}-\eta\left(\alpha_{0}\left(U_{i}\right)\right)\right)\right)K_{h}\left(U_{i}-u\right)\Bigg\|+ \\ \Bigg\|\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k}\frac{\partial}{\partial\widehat{\xi}_{ij}}\left(\eta'\left(\alpha_{0}\left(U_{i}\right)+\overline{\xi}_{i}\right)^{T}V_{i}\left(\eta,\beta_{0}\right)^{-1}\Omega_{i}s\left(\alpha_{0}\left(U_{i}\right)+\overline{\xi}_{i}\right)\times \\ \frac{\partial}{\partial\alpha}\psi\left(A_{i}^{-1/2}\left(Y_{i}-\eta\left(\alpha_{0}\left(U_{i}\right)+\overline{\xi}_{i}\right)\right)\right)\Bigg\|\left\|\widehat{\xi}_{ij}\right\|\Bigg\} \\ =o_{p}\left(1\right)\left(O_{p}\left(1\right)+o_{p}\left(1\right)\right)=o_{p}\left(1\right),\end{aligned}$$

hence by the triangle inequality $\left\|\widehat{\Gamma}_{\mu}\left(u\right)-f\left(u\right)\Gamma_{\mu0}\left(u\right)\right\|=o_{p}\left(1\right).$

Similar arguments can be used to show that $\widehat{\Sigma}_{\widehat{\pi}\mu}\left(u\right)=\widehat{\Sigma}_{\pi\mu}\left(u\right)+o_{p}\left(1\right)$, where

$$\begin{aligned}\widehat{\Sigma}_{\pi\mu}\left(u\right)=\frac{1}{n}\sum_{i=1}^{n}\Bigg[\left(\eta'\left(\widehat{a}_{1\widehat{\pi}}\right)^{T}V_{i}\left(\eta,\beta_{0}\right)^{-1}\Omega_{i}\Pi\left(\delta_{i},X_{i},U_{i}\right)\right)\times \\ \psi\left(A_{i}^{-1/2}\left(Y_{i}-\eta\left(X_{i}^{T}\widehat{a}_{1\widehat{\pi}}\right)\right)\right)\Bigg]^{\otimes2}K_{h}\left(U_{i}-u\right)- \\ \Bigg[\frac{1}{n}\sum_{i=1}^{n}\eta'\left(X_{i}^{T}\widehat{a}_{1\widehat{\pi}}\right)^{T}V_{i}\left(\eta,\beta_{0}\right)^{-1}\Omega_{i}\Pi\left(\delta_{i},X_{i},U_{i}\right)\times \\ \psi\left(A_{i}^{-1/2}\left(Y_{i}-\eta\left(X_{i}^{T}\widehat{a}_{1\widehat{\pi}}\right)\right)\right)K_{h}\left(U_{i}-u\right)\Bigg]^{\otimes2}+o_{p}\left(1\right)= \\ T_{26}+T_{27}+o_{p}\left(1\right),\end{aligned}$$

and that $\left\|T_{26}-f\left(u\right)\Sigma_{\pi\mu0}\left(u\right)\right\|=o_{p}\left(1\right)$, $\left\|T_{27}\right\|=o_{p}\left(1\right)$, hence by the triangle inequality $\left\|\widehat{\Sigma}_{\widehat{\pi}\mu}\left(u\right)-f\left(u\right)\Sigma_{\pi\mu0}\left(u\right)\right\|=o_{p}\left(1\right).$

Proof (Proof of proposition 5) Let $\psi\left(r\left(Y_{i},\eta\left(X_{i}^{T}\alpha\left(U_{i}\right)\right)\right)\right):=\psi\left(\alpha\left(U_{i}\right)\right)$, $V\left(\eta\left(X_{i}^{T}\alpha\left(U_{i}\right)\right)\right):=V\left(\alpha\left(U_{i}\right)\right)$, $s_{\alpha}\left(X_{i},U_{i}\right):=s\left(\alpha\left(U_{i}\right)\right)$, $\eta'\left(X_{i}^{T}\alpha\left(U_{i}\right)\right):=\eta'\left(\alpha\left(U_{i}\right)\right)$ and $\widehat{\xi}_{i}=$

$\widehat{a}_{1\widehat{\pi}} - \alpha_0(U_i)$. By (12)

$$\widehat{\Gamma}_\zeta(u) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\frac{\psi(\alpha_0(U_i))}{V(\alpha_0(U_i))^{1/2}} s(\alpha_0(U_i))\left[\eta'(\alpha_0(U_i))X_i\right]^{\otimes 2}\omega(X_i)K_h(U_i-u) +$$
$$\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\left\{\left[\frac{\psi\left(\alpha_0(U_i)+\widehat{\xi}_i\right)}{V(\widehat{\alpha}(U_i))^{1/2}} s\left(\alpha_0(U_i)+\widehat{\xi}_i\right)\left[\eta'\left(\alpha_0(U_i)+\widehat{\xi}_i\right)X_i\right]^{\otimes 2}\right] -\right.$$
$$\left.\frac{\psi(\alpha_0(U_i))}{V(\alpha_0(U_i))^{1/2}} s(\alpha_0(U_i))\left[\eta'(\alpha_0(U_i))X_i\right]^{\otimes 2}\right\}\omega(X_i)K_h(U_i-u) + o_p(1) =$$
$$T_{28} + T_{29}.$$

By a standard kernel calculation $\|T_{28} - f(u)\Gamma_{\zeta 0}(u)\| = o_p(1)$, whereas for $|U_i - u| \le h$ a Taylor expansion and the results of Masry (1996) show that $\sup_{|U_i-u|\le h}\left\|\widehat{\xi}_i\right\| = o_p(1)$. Note that by A2(i) and iterated expectations

$$\frac{1}{n}\sum_{i=1}^{n} E\left\{\sup_{\|\xi_i\|\le\varepsilon}\left\|\frac{\delta_i}{\pi_i}\left[\frac{\psi(\alpha_0(U_i)+\xi_i)}{V(\widehat{\alpha}(U_i))^{1/2}} s(\alpha_0(U_i)+\xi_i)\left[\eta'(\alpha_0(U_i)+\xi_i)X_i\right]^{\otimes 2} -\right.\right.\right.$$
$$\left.\left.\left.\frac{\psi(\alpha_0(U_i))}{V(\alpha_0(U_i))^{1/2}} s(\alpha_0(U_i))\left[\eta'(\alpha_0(U))X_i\right]^{\otimes 2}\right]\omega(X_i)K_h(U_i-u)\right\|\right\} \le$$
$$o(1)\frac{1}{n}\sum_{i=1}^{n} E\left[K_h(U_i-u)\right] = o(1)$$

and

$$E\left\{\sup_{\|\xi_i\|\le\varepsilon}\left\|\frac{\delta_i}{\pi_i}\left[\frac{\psi(\alpha_0(U_i)+\xi_i)}{V(\widehat{\alpha}(U_i))^{1/2}} s(\alpha_0(U_i)+\xi_i)\left[\eta'(\alpha_0(U_i)+\xi_i)X_i\right]^{\otimes 2} -\right.\right.\right.$$
$$\left.\left.\left.\frac{\psi(\alpha_0(U_i))}{V(\alpha_0(U_i))^{1/2}} s(\alpha_0(U_i))\left[\eta'(\alpha_0(U))X_i\right]^{\otimes 2}\right]\omega(X_i)K_h(U_i-u)\right\|^2\right\} =$$
$$O\left(\frac{1}{nh}\right) + o(1)$$

hence $Var(T_{29}) = o(1)$ and $\|T_{29}\| = o_p(1)$ since $\widehat{\xi}_i \in \|\xi_i\| \le \varepsilon$ with probability approaching 1, so by the triangle inequality $\left\|\widehat{\Gamma}_\zeta(u) - f(u)\Gamma_{\zeta 0}(u)\right\| = o_p(1)$. For $\widehat{\Sigma}_{\widehat{\pi}\zeta}(u)$ by (12)

$$\widehat{\Sigma}_{\widehat{\pi}\zeta}(u) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i^2}\left[\frac{\psi(\widehat{a}_{1\widehat{\pi}})}{V(\widehat{a}_{1\widehat{\pi}})^{1/2}}\eta'(\widehat{a}_{1\widehat{\pi}})X_i\omega(X_i)\right]^{\otimes 2} K_h(U_i-u) -$$
$$\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\delta_i}{\pi_i}\frac{\psi(\widehat{a}_{1\widehat{\pi}})}{V(\widehat{a}_{1\widehat{\pi}})^{1/2}}\eta'(\widehat{a}_{1\widehat{\pi}})X_i\omega(X_i)K_h(U_i-u)\right]\right\}^{\otimes 2} + o_p(1)$$
$$= T_{30} + T_{31} + o_p(1).$$

Note that by a mean value expansion

$$\frac{\psi\left(\widehat{a}_{1\widehat{\pi}}\right)}{V\left(\widehat{a}_{1\widehat{\pi}}\right)^{1/2}}\eta'\left(\widehat{a}_{1\widehat{\pi}}\right)X_i\omega\left(X_i\right)=\frac{\psi\left(\alpha_0\left(U_i\right)+\widehat{\xi}_i\right)}{V\left(\alpha_0\left(U_i\right)+\widehat{\xi}_i\right)^{1/2}}\eta'\left(\alpha_0\left(U_i\right)+\widehat{\xi}_i\right)X_i\omega\left(X_i\right)=$$

$$\frac{\psi\left(\alpha_0\left(U_i\right)\right)}{V\left(\alpha_0\left(U_i\right)_i\right)^{1/2}}\eta'\left(\alpha_0\left(U_i\right)\right)X_i\omega\left(X_i\right)+\sum_{j=1}^{k}\frac{\partial}{\partial\widehat{\xi}_{ij}}\left[\frac{\psi\left(\alpha_0\left(U_i\right)+\overline{\xi}_i\right)}{V\left(\alpha_0\left(U_i\right)+\overline{\xi}_i\right)^{1/2}}\eta'\left(\alpha_0\left(U_i\right)+\overline{\xi}_i\right)\right]\widehat{\xi}_{ij},$$

hence

$$\begin{aligned}T_{30}&=\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i^2}\left\{\left(\frac{\psi\left(\alpha_0\left(U_i\right)\right)}{V\left(\alpha_0\left(U_i\right)\right)^{1/2}}\eta'\left(\alpha_0\left(U_i\right)\right)X_i\omega\left(X_i\right)\right)^{\otimes 2}+\right.\\&\left(\sum_{j=1}^{k}\frac{\partial}{\partial\widehat{\xi}_{ij}}\left(\frac{\psi\left(\alpha_0\left(U_i\right)+\overline{\xi}_i\right)}{V\left(\alpha_0\left(U_i\right)+\overline{\xi}_i\right)^{1/2}}\eta'\left(\alpha_0\left(U_i\right)+\overline{\xi}_i\right)\right)X_i\omega\left(X_i\right)\widehat{\xi}_{ij}\right)^{\otimes 2}+\\&2\left(\frac{\psi\left(\alpha_0\left(U_i\right)\right)}{V\left(\alpha_0\left(U_i\right)\right)^{1/2}}\eta'\left(\alpha_0\left(U_i\right)\right)X_i\omega\left(X_i\right)\right)\times\\&\left.\left(\sum_{j=1}^{k}\frac{\partial}{\partial\widehat{\xi}_{ij}}\left(\frac{\psi\left(\alpha_0\left(U_i\right)+\overline{\xi}_i\right)}{V\left(\alpha_0\left(U_i\right)+\overline{\xi}_i\right)^{1/2}}\eta'\left(\alpha_0\left(U_i\right)\right)\right)X_i\omega\left(X_i\right)\widehat{\xi}_{ij}\right)^{T}\right\}K_h\left(U_i-u\right)+o_p\left(1\right)\\&=T_{301}+T_{302}+T_{303}+o_p\left(1\right).\end{aligned}$$

By a standard kernel calculation

$$\left\|T_{301}-f\left(u\right)E\left\{\left[\frac{\psi\left(r\left(\alpha_0\left(U\right)\right)\right)}{V\left(\alpha_0\left(U\right)\right)^{1/2}}\eta'\left(\alpha_0\left(U\right)\right)X\omega\left(X\right)\right]^{\otimes 2}|U=u\right\}\right\|=o_p\left(1\right)$$

and by the law of large numbers

$$\begin{aligned}\left\|T_{302}\right\|\leq&\sup_{\left|U_i-u\right|\leq h}\left\|\widehat{\xi}_i\right\|^2\frac{1}{n}\sum_{i=1}^{n}\left\|\left(\frac{\partial}{\partial\widehat{\xi}}\left(\frac{\psi\left(\alpha_0\left(U_i\right)+\overline{\xi}_i\right)}{V\left(\alpha_0\left(U_i\right)+\overline{\xi}_i\right)^{1/2}}\eta'\left(\alpha_0\left(U_i\right)+\overline{\xi}_i\right)\right)X_i\omega\left(X_i\right)\right)\right\|^{\otimes 2}=\\&o_p\left(1\right)O_p\left(1\right),\end{aligned}$$

whereas by the Cauchy-Schwarz inequality

$$\begin{aligned}\left\|T_{303}\right\|\leq 2&\sup_{\left|U_i-u\right|\leq h}\left\|\widehat{\xi}_i\right\|\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{\delta_i}{\pi_i^2}\frac{\psi\left(\alpha_0\left(U_i\right)\right)}{V\left(\alpha_0\left(U_i\right)\right)^{1/2}}\eta'\left(\alpha_0\left(U_i\right)\right)X_i\omega\left(X_i\right)\right\|^{\otimes 2}K_h\left(U_i-u\right)\right)^{1/2}\\&\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{\delta_i}{\pi_i^2}\frac{\partial}{\partial\widehat{\xi}_{ij}}\left(\frac{\psi\left(\alpha_0\left(U_i\right)+\overline{\xi}_i\right)}{V\left(\alpha_0\left(U_i\right)+\overline{\xi}_i\right)^{1/2}}\eta'\left(\alpha_0\left(U_i\right)+\overline{\xi}_i\right)\right)X_i\omega\left(X_i\right)\right\|^{2}K_h\left(U_i-u\right)\right)^{1/2}=\\&o_p\left(1\right)O_p\left(1\right),\end{aligned}$$

thus by the triangle inequality $\|T_{30} - f(u)\,\Gamma_{\zeta 0}(u)\| = o_p(1)$. The same argument and the continuous mapping theorem imply that

$$\left\| T_{31} - \left\{ f(u)\,E\left[\frac{\psi(\alpha_0(U))}{V(\eta(\alpha_0(U)))^{1/2}} \eta'(\alpha_0(U))\,X\omega(X)\,|U=u \right] \right\}^{\otimes 2} \right\| = o_p(1),$$

hence by the triangle inequality $\left\| \widehat{\Sigma}_{\widehat{\pi}}(u) - f(u)\,\Sigma_{\pi\zeta 0}(u) \right\| = o_p(1)$.

Proof (Proof of Proposition 6) Let $\eta(X_i, \alpha(U_i)) := \eta(\alpha(U_i))$; by (12)

$$\widehat{\Gamma}(u) = \widehat{\psi}'(\widehat{\varepsilon}_i)\,\frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{\pi_i} \frac{\partial \eta(\widehat{a}_{1\widehat{\pi}})}{\partial \alpha}^{\otimes 2} \omega(X_i)\,K_h(U_i - u) + o_p(1)$$

and by the same arguments used in Proposition 5

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{\pi_i} \frac{\partial \eta(\widehat{a}_{1\widehat{\pi}})}{\partial \alpha}^{\otimes 2} \omega(X_i)\,K_h(U_i - u) - f(u)\,E\left[\frac{\partial \eta(\alpha_0(U))}{\partial \alpha}^{\otimes 2} \omega(X)\,|U=u \right] \right\| = o_p(1).$$

Let $\widehat{\xi}_i = \eta(X_i, \widehat{\alpha}(U_i)) - \eta(X_i, \alpha_0(U_i))$ and note that by the continuity of $\eta(\cdot)$ and consistency of $\widehat{a}_{1\widehat{\pi}}$

$$\sup_{U_i} \left| \widehat{\xi}_i \right| \leq \sup_{U_i} \left| \eta(X_i, \widehat{\alpha}(U_i)) - \eta(X_i, \alpha_0(U_i)) \right| = o_p(1),$$

hence, noting that $\psi'(\widehat{\varepsilon}_i) = \psi'\left(\varepsilon_i + \widehat{\xi}_i\right)$, it follows by the triangle inequality and the law of large numbers that

$$\left| \widehat{\psi}'(\widehat{\varepsilon}_i) - E(\psi'(\varepsilon)) \right| \leq \left| \widehat{\psi}'(\varepsilon_i) - E(\psi'(\varepsilon)) \right| + \frac{1}{n}\sum_{i=1}^{n} \sup_{|\xi_i|<\varepsilon} \left| \psi'(\varepsilon_i + \xi_i) - \psi'(\varepsilon_i) \right| = o_p(1),$$

hence $\left\| \widehat{\Gamma}_\zeta(u) - f(u)\,\Gamma_{\zeta 0}(u) \right\|$. Similarly for $\widehat{\Sigma}_{\widehat{\pi}\zeta}(u)$, continuity of $\psi(\cdot)^2$ and $\eta(\cdot)$ in a neighborhood of ε_i implies that

$$\begin{aligned} \left| \widehat{\psi}(\widehat{\varepsilon}_i)^2 - E\left(\psi(\varepsilon)^2\right) \right| &\leq \left| \widehat{\psi}(\varepsilon_i)^2 - E\left(\psi(\varepsilon)^2\right) \right| + \frac{1}{n}\sum_{i=1}^{n} \sup_u \left| \psi\left(\varepsilon_i + \widehat{\xi}_i\right)^2 - \psi(\varepsilon_i)^2 \right| \\ &= o_p(1), \end{aligned}$$

and

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}_i^{\,2}} \left[\frac{\partial \eta(\widehat{a}_{1\widehat{\pi}})}{\partial \alpha} \omega(X_i)\,K_h(U_i - u) \right]^{\otimes 2} - f(u)\,E\left[\frac{1}{\pi} \frac{\partial \eta(X, \alpha_0(U))}{\partial \alpha} \omega(X)\,|U=u \right]^{\otimes 2} \right\| = o_p(1),$$

hence by the triangle inequality $\left\| \widehat{\Sigma}_\zeta(u)/\widehat{\pi} - f(u)\,\Sigma_{\pi\zeta 0}(u) \right\| = o_p(1)$.

References

Bianco, A., Gracia Ben, M. and Yohai, V.: 2005, Robust estimation for linear regression with asymmetric errors, *Canadian Journal of Statistics* **33**, 511–528.

Bianco, A. and Yohai, V.: 1996, Robust estimation in the logistic regression model, *Robust statistics, Data analysis and computer intensive methods, Lecture Notes in Statistics 109*, Springer and Verlag, New York.

Boente, G., He, X. and Zhou, J.: 2006, Robust estimates in generalized partially linear models, *Annals of Statistics* **34**, 2856–2878.

Cai, Z., Fan, J. and Li, R.: 2000, Efficient estimation and inference for varying-coefficient models, *Journal of the American Statistical Association* **95**, 888–902.

Cantoni, E. and Ronchetti, E.: 2001, Robust inference for generalized linear models, *Journal of the American Statistical Association* **96**, 1022–1030.

He, X., Fung, W. and Zhu, Z.: 2005, Robust estimation in generalized partial linear models for clustered data, *Journal of the American Statistical Association* **100**, 1176–1184.

Masry, E.: 1996, Multivariate local polynomial regression for time series: Uniform strong consistency and rates, *Journal of Time Series Analysis* **17**, 571–599.

McCullagh, P. and Nelder, J.: 1989, *Generalized Linear Models*, Chapman and Hall, London.

Phelps, K.: 1982, Use of the complementary log-log function to describe dose-response relationships in insecticide evaluation field trials, *Lecture Notes in Statistics, No. 14. GLIM.82: Proceedings of the International Conference on Generalized Linear Models.*

Pollard: 1991, Asymptotics for least absolute deviation regression estimators, *Econometric Theory* **7**, 186–199.

5 Tables and Figures

Table 3 Bias (B) and standard deviation (SD) for robust and nonrobust estimators in the Poisson regression example

n	200		500		200		500		200		500		200		500	
	B	SD	B	SD	B	SD	B	SD	B	SD	B	SD	B	SD	B	SD
	$0^{(NR)}$				$0^{(R)}$				$C0^{(NR)}$				$C0^{(R)}$			
$\widehat{a}_{11c}$	.120	.459	.081	.227	.126	.478	.092	.251	.204	.509	.128	.307	.206	.522	.124	.310
$\widehat{a}_{11\widehat{\pi}p}$	–		–		–		–		.182	.513	.111	.313	.185	.524	.114	.314
$\widehat{a}_{11\widehat{\pi}np}$	–		–		–		–		.185	.518	.114	.319	.185	.528	.116	.317
$\widehat{a}_{12c}$	.117	.440	.078	.218	.123	.467	.081	.235	.196	.504	.119	.313	.199	.517	.121	.314
$\widehat{a}_{12\widehat{\pi}p}$	–		–		–		–		.175	.510	.105	.317	.185	.519	.107	.320
$\widehat{a}_{12\widehat{\pi}np}$	–		–		–		–		.179	.515	.106	.321	.192	.521	.107	.322
	$C1^{(NR)}$				$C1^{(R)}$				$C2^{(NR)}$				$C2^{(R)}$			
$\widehat{a}_{11c}$	.259	.720	.226	.616	.208	.530	.129	.306	.310	.787	.274	.707	.210	.535	.131	.315
$\widehat{a}_{11\widehat{\pi}p}$	.268	.733	.239	.627	.188	.534	.120	.311	.333	.816	.287	.721	.193	.542	.117	.320
$\widehat{a}_{11\widehat{\pi}np}$	.271	.738	.246	.634	.185	.539	.123	.315	.356	.844	.296	.736	.190	.548	.121	.322
$\widehat{a}_{12c}$	.257	.723	.221	.610	.202	.527	.120	.310	.292	.764	.280	.701	.203	.537	.125	.310
$\widehat{a}_{12\widehat{\pi}p}$	.265	.730	.237	.618	.188	.533	.110	.313	.299	.773	.295	.712	.191	.519	.118	.311
$\widehat{a}_{12\widehat{\pi}np}$	.269	.741	.245	.629	.194	.539	.111	.316	.301	.776	.299	.724	.194	.522	.120	.315

(NR) nonrobust, estimation , (R) robust estimation

Table 4 Finite sample size for robust and nonrobust tests in the Poisson regression example

n	200		500		200		500	
	$0^{(NR)}$				$0^{(R)}$			
$\max_j W_{co}(u_j)$	.104	.058	.108	.055	.115	.060	.110	.061
n	$C0^{(NR)}$				$C0^{(R)}$			
$\max_j W_{co}(u_j)$	.114	.060	.112	.059	.116	.065	.110	.061
$\max_j W_{\widehat{\pi}p}(u_j)$	.112	.056	.108	.055	.114	.058	.108	.056
$\max_j W_{\widehat{\pi}np}(u_j)$	.115	.058	.113	.056	.116	.060	.112	.058
	$C2^{(NR)}$				$C2^{(R)}$			
$\max_j W_{co}(u_j)$	.166	.100	.165	.088	.118	.070	.114	.060
$\max_j W_{\widehat{\pi}p}(u_j)$	.179	.099	.175	.095	.115	.060	.109	.055
$\max_j W_{\widehat{\pi}np}(u_j)$	.181	.110	.171	.099	.117	.063	.110	.060

(NR) nonrobust, estimation , (R) robust estimation

Table 5 Bias (B) and standard deviation (SD) for robusr and nonrobust estimators in the nonlinear regression example

	200		500		200		500		200		500		200		500	
n	B	SD	B	SD	B	SD	B	SD	B	SD	B	SD	B	SD	B	SD
	$0^{(NR)}$				$0^{(R)}$				$C0^{(NR)}$				$C0^{(R)}$			
$\widehat{a}_{11c}$	.094	.409	.080	.201	.101	.418	.089	.221	.104	.436	.094	.234	.106	.450	.097	.241
$\widehat{a}_{11\widehat{\pi}p}$	–			–	–			–	.094	.445	.081	.242	.096	.458	.080	.250
$\widehat{a}_{11\widehat{\pi}np}$	–			–	–			–	.096	.449	.084	.250	.098	.459	.083	.254
$\widehat{a}_{12c}$	.088	.401	.072	.197	.095	.407	.082	.215	.094	.420	.072	.218	.068	.430	.068	.220
$\widehat{a}_{12\widehat{\pi}p}$	–			–	–			–	.087	.427	.070	.224	.088	.435	.071	.228
$\widehat{a}_{12\widehat{\pi}np}$	–			–	–			–	.088	.431	.072	.321	.090	.438	.073	.232
	$C1^{(NR)}$				$C1^{(R)}$				$C2^{(NR)}$				$C2^{(R)}$			
$\widehat{a}_{11c}$	.136	.522	.130	.441	.108	.434	.098	.264	.201	.634	.190	.627	.110	.440	.096	.266
$\widehat{a}_{11\widehat{\pi}p}$	.153	.554	.147	.448	.096	.438	.081	.270	.213	.637	.194	.628	.101	.446	.085	.281
$\widehat{a}_{11\widehat{\pi}np}$	.153	.553	.148	.454	.099	.446	.087	.275	.218	.638	.199	.630	.108	.451	.089	.286
$\widehat{a}_{12c}$	.125	.501	.123	.427	.098	.422	.078	.246	.170	.584	.165	.574	.105	.431	.081	.248
$\widehat{a}_{12\widehat{\pi}p}$	.131	.511	.126	.431	.089	.426	.073	.255	.181	.591	.169	.581	.090	.435	.071	.253
$\widehat{a}_{12\widehat{\pi}np}$	.133	.515	.129	.435	.090	.430	.075	.260	.185	.595	.170	.584	.092	.441	.074	.259

(NR) nonrobust, estimation , (R) robust estimation

Table 6 Finite sample size for robust and nonrobust tests in the nonlinear regression example

n	200		500		200		500	
	$0^{(NR)}$				$0^{(R)}$			
$\max_j W_{co}(u_j)$	.108	.057	.105	.053	.112	.060	.107	.057
	$C0^{(NR)}$				$C0^{(R)}$			
$\max_j W_{co}(u_j)$	.110	.063	.108	.058	.114	.066	.109	.060
$\max_j W_{\widehat{\pi}p}(u_j)$	.108	.058	.106	.056	.111	.060	.105	.057
$\max_j W_{\widehat{\pi}np}(u_j)$	.110	.060	.108	.059	.106	.063	.108	.059
	$C2^{(NR)}$				$C2^{(R)}$			
$\max_j W_{co}(u_j)$	.155	.095	.150	.091	.119	.068	.112	.072
$\max_j W_{\widehat{\pi}p}(u_j)$	.168	.109	.160	.103	.109	.061	.106	.058
$\max_j W_{\widehat{\pi}np}(u_j)$	.174	.110	.168	.102	.112	.065	.107	.060

(NR) nonrobust, estimation , (R) robust estimation

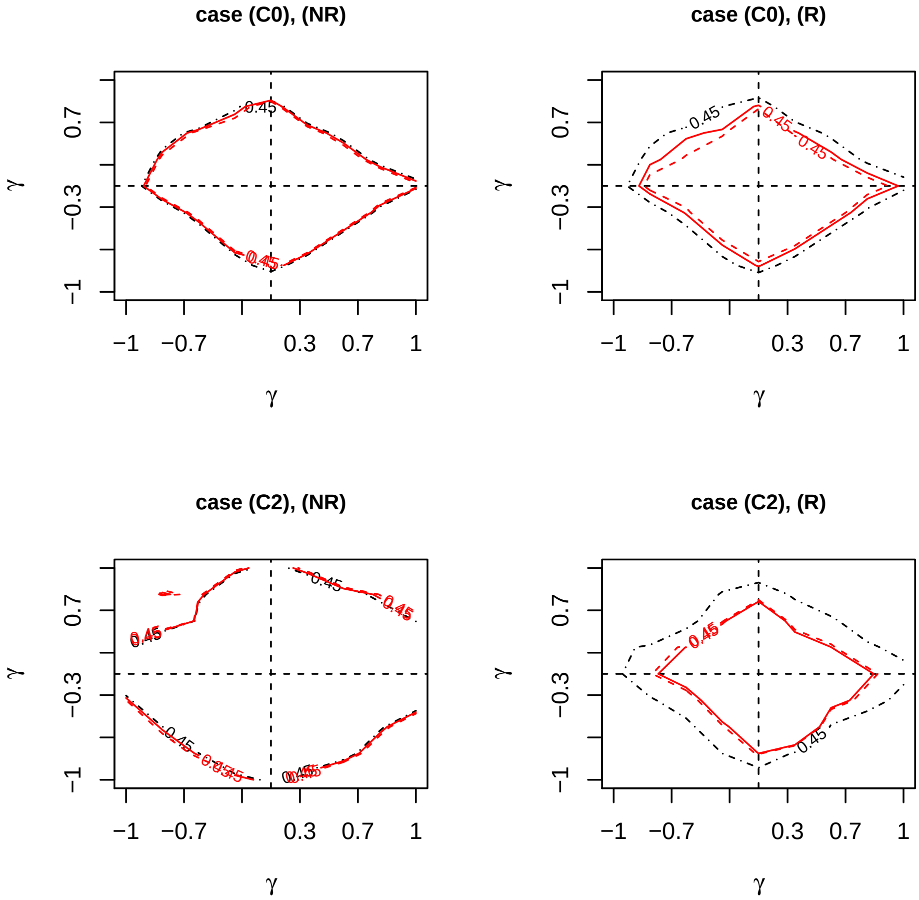

Fig. 3 Finite sample adjusted power for the test statistic $\max_{1\leq j\leq 5} W(u_j)$; solid lines indicate $\max_{1\leq j\leq 5} W_{\widehat{\pi}p}(u_j)$, dashed lines indicate $\max_{1\leq j\leq 5} W_{\widehat{\pi}np}(u_j)$ and dot dashed lines indicate $\max_{1\leq j\leq 5} W_{co}(u_j)$ for the Poisson regression example

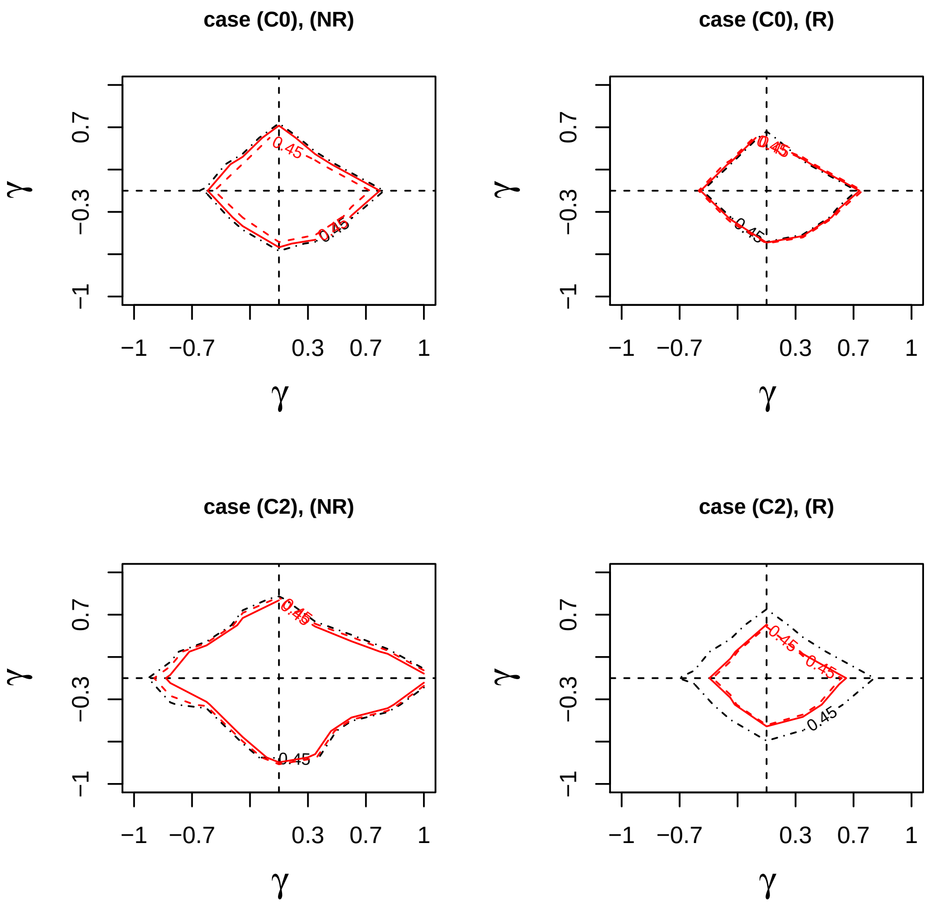

Fig. 4 Finite sample adjusted power for the test statistic $\max_{1\leq j\leq 5} W(u_j)$; solid lines indicate $\max_{1\leq j\leq 5} W_{\widehat{\pi}p}(u_j)$, dashed lines indicate $\max_{1\leq j\leq 5} W_{\widehat{\pi}np}(u_j)$ and dot dashed lines indicate $\max_{1\leq j\leq 5} W_{co}(u_j)$ for the nonlinear regression example

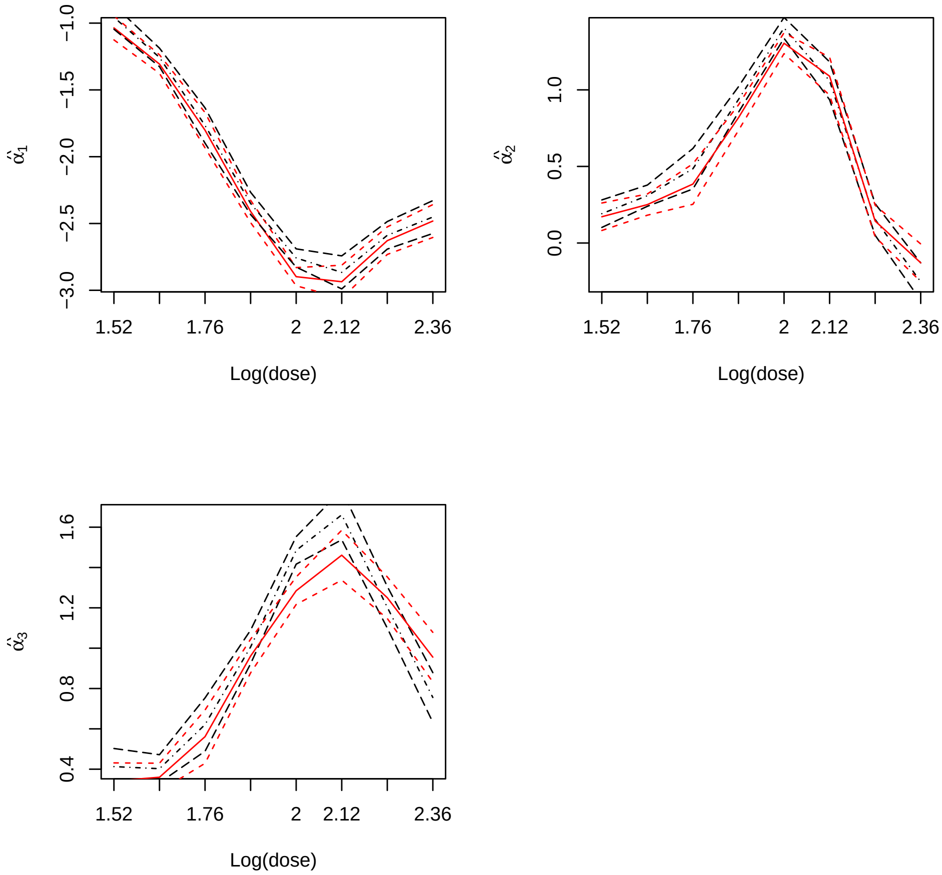

Fig. 5 Estimated varying coefficients for the damaged carrot data. The solid line represents the robust local estimator with 95% confidence interval (dashed lines). The dot dashed line represents the nonrobust local estimator with 95% confidence interval (long dashed lines).