A Semiparametric Model for Binary Response and Continuous Outcomes under Index Heteroscedasticity

Roger Klein (Rutgers University)
Francis Vella (Georgetown University)

January 15, 2009

1 Appendix

The Appendix is organized into two subsections, with the first stating and proving all intermediate lemmas that we require to establish the asymptotic properties of the estimators. The second subsection employs these lemmas to prove the main results in the paper.

1.1 Intermediate Results

From (D1-D7) of the Assumptions and Definitions section, recall that $\hat{f}_1(\bullet)$ estimates $\Pr(Y_2 = 1)g_1(w)$, where $g_1(w)$ is the density for W conditioned on $Y_2 = 1$. Similarly, $\hat{f}_0(\bullet)$ estimates $\Pr(Y_2 = 0)g_0(w)$, where $g_0(w)$ is the density for W conditioned on $Y_2 = 0$. Throughout, all lemmas apply to both $\hat{f}_1(\bullet)$ and $\hat{f}_0(\bullet)$. Accordingly, for notational convenience, we will simply write $\hat{f}(\bullet)$ to refer to either of these estimators. In so doing, we will refer to the local smoothing parameters as λ without subscripting. Throughout, we will write $\nabla_\eta^k f$ to mean the k^{th} partial derivative of f with respect to η, with $\nabla_\eta^0 f \equiv f$. Finally, in terms of notation, denote X_c and X_d as the vectors of continuous and discrete variables respectively, with realizations $x_c \,\epsilon \mathcal{X}_c$ and $x_d \,\epsilon \mathcal{X}_d$. With $\mathcal{X}_{c1}$ as the subset of $\mathcal{X}_c$ on which $\tau_x = 1$ (see D6), define $\mathcal{X}_1 \equiv \{x : x_c \,\epsilon \mathcal{X}_{c1},\ x_d \,\epsilon \mathcal{X}_d\}$.

Recalling that $w = [x_1 + x_3\eta_{31},\ x_2 + x_3\eta_{32}] \equiv [w_1(\eta), w_2(\eta)]$, let $\mathcal{X}_2 = \{x : \underline{w}_k < w_k(\eta) < \bar{w}_k,\ \ k = 1, 2\}$. Finally, let $\mathcal{A} \equiv \{x : x \in \mathcal{X}_1 \cup \mathcal{X}_2\}$. All uniform results will be on $\mathcal{A}$, and, when appropriate, the compact parameter space. Though not stated explicitly, for all of the results below, we employ all assumptions in (A1-6) and (D1-7).

The estimated conditional densities above depend on the sample covariance matrix for W. As W depends on the index parameters, η, we denote this covariance matrix as $\hat{\Sigma}(\eta)$. With $\Sigma(\eta)$ as the uniform (in η) probability limit of $\hat{\Sigma}(\eta)$, Lemma 1 below will enable us to treat this estimated matrix as if it were known.

Lemma 1: Denote $\hat{f}\left(w; \hat{\Sigma}(\eta)\right)$ as the estimator defined in (D1-3) and denote $\hat{f}(w; \Sigma(\eta))$ as the corresponding estimator with $\Sigma(\eta)$ replacing $\hat{\Sigma}(\eta)$. Define $\widehat{f_0}(\bullet)$ analogously. Then:

$$\sup_{\eta,\bar{x}} \left| \nabla_\eta^k \widehat{f}\left(\bar{w}; \hat{\Sigma}(\eta)\right) - \nabla_\eta^k \hat{f}(\bar{w}; \Sigma(\eta)) \right| = o_p\left(N^{-1/2}\right), \quad k = 0, 1, 2,$$

where uniformity is over the sets described above.

Proof of Lemma 1: From a Taylor series expansion:

$$\left| \nabla_\eta^k \hat{f}\left(w; \hat{\Sigma}(\eta)\right) - \nabla_\eta^k \hat{f}_m(w; \Sigma(\eta)) \right| \leq \sup_{\eta,x} \left| \nabla_\Sigma \nabla_\eta^k \hat{f}\left(w; \hat{\Sigma}(\eta)\right) \right| \sup_\eta \left| \hat{\Sigma}(\eta) - \Sigma(\eta) \right|.$$

Since $\hat{f}$ converges to f even under an inconsistent estimator for Σ, it can be shown that the first term above is $o_p(1)$. As the second term is $O_p\left(N^{-1/2}\right)$, the result follows.

Employing Lemma 1, we will proceed with $\Sigma(\eta)$ replacing $\hat{\Sigma}(\eta)$ throughout. To simplify the argument further, it is also convenient to replace all estimated components in local smoothing parameters with their expectations. From (D2-3) estimated smoothing parameters are given as:

$$\hat{\lambda}_j = \left[\hat{d}_j \hat{\gamma}_j + \left(1 - \hat{d}_j\right) / Ln(N)\right]^{-1/2} \equiv \lambda\left(\hat{\gamma}_j\right),$$

where $\hat{\gamma}_j \equiv [\hat{\pi}_j / \hat{m}]$ and $\hat{d}$ is the smoothed indicator:

$$\hat{d}_j \equiv \left\{1+\exp\left(-N^{\varepsilon}\left[\hat{\gamma}_j - \frac{1}{Ln\,(N)}\right]\right)\right\}^{-1} \equiv d\left(\hat{\gamma}_j\right).$$

Define $\bar{\gamma}_j \equiv [E\left(\hat{\pi}_j\right)/m]\,, \bar{d}_j \equiv d\left(\gamma_j\right)$, and

$$\bar{\lambda}_j \equiv \left[\bar{\gamma}_j \bar{d}_j + \left(1-\bar{d}\right)/Ln(N)\right]^{-1/2} = \lambda\left(\bar{\gamma}_j\right).$$

Write $\hat{f}\left(\bar{w};\hat{\lambda}\right)$ as the estimator of f at $\overline{w} = \bar{x}\eta$ and let $\hat{f}\left(\overline{w};\bar{\lambda}\right)$ be the corresponding estimator with $\bar{\lambda}$ replacing $\hat{\lambda}$. In the next three lemmas, we examine convergence rates under multi-stage local smoothing. For estimated densities and first derivatives Lemmas $2A-B$ provide the required intermediate results needed to establish convergence rates in the third stage of local smoothing (Lemma 2C). Throughout, $\overline{w} \equiv \bar{x}\eta$.

Lemma 2A: Stage 1 (No local smoothing). Let $\hat{\boldsymbol{\lambda}}_1 = \mathbf{1}$ Then, for $\bar{x} \,\epsilon\, \mathcal{A}$ η in a compact set, and $k = 0, 1, 2$:

$$a) \quad : \quad \sup_{\bar{x},\eta}\left|\nabla^k_\eta \hat{f}\left(\overline{w};\mathbf{1},h_1\right) - E\nabla^k_\eta \hat{f}\left(\overline{w};\mathbf{1},h_1\right)\right| = O_p\left(1/\left[N^{1/2}h_1^{k+2}\right]\right)$$

$$b) \quad : \quad \sup_{\bar{x},\eta}\left|E\nabla^k_\eta \hat{f}\left(\overline{w};\mathbf{1},h_1\right) - \nabla^k_\eta f\left(\overline{w}\right)\right| = O_p\left(h_1^2\right).$$

Proof of Lemma 2A: Standard bias and uniform convergence results provide the proof (see Klein and Spady(1993)).

Employing the above results without local smoothing, Lemma 2B below examines convergence rates in which local smoothing is based on the density estimator in Lemma 2A.

Lemma 2B: Stage 2 (Local Smoothing). Let $\hat{\boldsymbol{\lambda}}_2 \equiv \lambda\left(\hat{f}\left(w;h_1,\mathbf{1}\right)\right)$, $\bar{\boldsymbol{\lambda}}_2 \equiv \lambda\left(E\left[\hat{f}\left(w;h_1,\mathbf{1}\right)\right]\right)$, and $h_i = O(N^{-r_i})$, $i = 1, 2$. Assuming $r_1 < r_2$, for $k = 0, 1, 2$:

$$a) \quad : \quad \sup_{\bar{x},\eta}\left|\nabla^k_\eta \hat{f}\left(\overline{w};\hat{\boldsymbol{\lambda}}_2,h_2\right) - \nabla^k_\eta \hat{f}\left(\overline{w};\bar{\boldsymbol{\lambda}}_2,h_2\right)\right| = O_p\left(1/\left[N^{1/2}h_2^{k+2}\right]\right)$$

$$b) \quad : \quad \sup_{\bar{x},\eta}\left|\nabla^k_\eta \hat{f}\left(\overline{w};\bar{\boldsymbol{\lambda}}_2,h_2\right) - E\nabla^k_\eta \hat{f}\left(\overline{w};\bar{\boldsymbol{\lambda}}_2,h_2\right)\right| = O_p\left(1/\left[N^{1/2}h_2^{k+2}\right]\right)$$

$$c) \quad : \quad \sup_{\bar{x},\eta}\left|E\nabla^k_\eta \hat{f}\left(\overline{w};\bar{\boldsymbol{\lambda}}_2,h_2\right) - \nabla^k_\eta f\left(\overline{w}\right)\right| = O_p\left(h_2^2 h_1^2\right).$$

Proof of Lemma 2B. Employing a Taylor series approximation, the proof for (a) follows from the uniform convergence rate of $\hat{\boldsymbol{\lambda}}_{2i}$ to $\bar{\boldsymbol{\lambda}}_{2i}$ (Klein and Spady, 1993, Lemma 1) and the window condition: $r_1 < r_2$. The proof for (b) is essentially the same as that for (a). To establish (c), write (employing a dominance condition to differentiate under an integral):

$$E\left(\nabla_\eta^k \hat{f}\left(\overline{w}_{;2}, h_2, \bar{\lambda}_2\right)\right) = \nabla_\eta^k E\left(\hat{f}\left(\overline{w}_{;2}, h_2, \bar{\lambda}_2\right)\right) \equiv \nabla_\eta^k \Delta_2,$$

where the second expectation is taken with respect to the density for w. Taylor expand Δ_2 in h_2 about $h_2 = 0$ and use the symmetry in K about 0 to obtain:

$$\Delta_2 = \nabla_\eta^k h_2^2 \left[\hat{C}_2 - C_2\right] + \nabla_\eta^k h_2^2 C_2 + h_2^4 \hat{C}_4.$$

Here, $\hat{C}_2 \xrightarrow{p} C_2$, where C_2 contains terms (densities and density derivatives) that would follow from a Taylor series expansion using local smoothing parameters based on true densities.[1] For the first term in Δ_2, it consists of estimated densities and density derivatives. From the rate at which the expectation of an estimator (density or higher order derivatives) converges to the truth:

$$h_2^2 \left[\hat{C}_2 - C_2\right] = O_p\left[h_2^2 h_1^2\right].$$

From Abramson and Silverman, the second term vanishes as $C_2 = 0$. The argument now follows because in the final term: $\mathrm{h}_2^2 \hat{C}_4 = o_p\left(h_1^2\right)$.[2] Referring to (D2-4), $C_4 = O\left(N^{4a}\right)$, where $a = .01$ is a local smoothing parameter.

[1]Local smoothing parameters employ separate trimming to keep local smoothing parameters from becoming smaller than $\mathrm{O}_p\left(1/Ln(N)\right)$. In taking a Taylor series expansion about $\mathrm{h}_2 = 0$, derivatives of Local-smoothing trimming will appear. However, with densities evaluated at a "target" point for which they are bounded from below by $c > 0$, then such derivatives will vanish exponentially (and can therefore be ignored). This derivative can not be ignored in the final term of such an expansion as it is evaluated at an intermediate point.

[2]A typical term of $\hat{\mathrm{C}}_4$ depends on the integral of the product of a term involving the inverse of a density estimator raised to a power below 4 (T1), the fourth derivative of a density estimator (T2), the fourth derivative of the smooth trimming function (T3), the kernel, and the true density. Based on the smooth trimming of local smoothing, uniformly:

$$|T1T3| = o_p\left(N^{.04} Ln(N)\right)$$

Given the uniform rate at which the fourth derivative of a density estimator converges to the truth and the fact that $\mathrm{h}_2^2 N^{.04} Ln(N) = o\left(h_1^2\right)$, the result follows.

Under assumptions on smoothing parameters, the final term is of smaller order than the first, which completes the argument.[3]

Lemma 2C: Stage 3 (Local Smoothing). Let $\hat{\boldsymbol{\lambda}}_3 \equiv \lambda\left(\hat{f}\left(w;\ \hat{\boldsymbol{\lambda}}_2, h_2\right)\right)$ and $h_i = O(N^{-r_i})$, $i = 1,2,3$. With $r_i > 0$, assume $r_1 < r_2$ and that $r_1 + r_2 < r_3$. With $\bar{\boldsymbol{\lambda}}_2$ given as above, define $\bar{\boldsymbol{\lambda}}_3 \equiv \lambda\left(\hat{f}\left(w; \bar{\boldsymbol{\lambda}}_2\ , h_2\right)\right)$. Then, for $k = 0, 1, 2$:

$$a) \quad : \quad \sup_{\bar{x},\eta}\left|\nabla_\eta^k \hat{f}\left(\overline{w}; \hat{\boldsymbol{\lambda}}_3, h_3\right) - \nabla_\eta^k \hat{f}\left(\overline{w}; \bar{\boldsymbol{\lambda}}_3, h_3\right)\right| = O_p\left(1/\left[N^{1/2} h_3^{k+2}\right]\right)$$

$$b) \quad : \quad \sup_{\bar{x},\eta}\left|\nabla_\eta^k \hat{f}\left(\overline{w}; \bar{\boldsymbol{\lambda}}_3, h_3\right) - E\nabla_\eta^k \hat{f}\left(\overline{w}; \bar{\boldsymbol{\lambda}}_3, h_3\right)\right| = O_p\left(1/\left[N^{1/2} h_3^{k+2}\right]\right)$$

$$c) \quad : \quad \sup_{\bar{x},\eta}\left|E\nabla_\eta^k \hat{f}\left(\overline{w}; \bar{\boldsymbol{\lambda}}_3, h_3\right) - \nabla_\eta^k f\left(\overline{w}\right)\right| = O_p\left(h_3^2 h_2^2 h_1^2\right).$$

Proof of Lemma 2C. The proof of (a-b) is the same as that in the previous lemma. For (c), define Δ_3 as in the previous lemma with $\bar{\boldsymbol{\lambda}}_3$ replacing $\bar{\boldsymbol{\lambda}}_2$. Then from the same type of Taylor expansion as in the previous theorem:

$$\Delta_3 = \nabla_\eta^k h_3^2\left[\hat{C}_2^* - C_2\right] + h_3^4 \hat{C}_3^*.$$

From Lemma 2C, the first term above has order $h_3^2 h_2^2 h_1^2$. With $\hat{C}_3^* = O\left(N^{4a}\right)$, similar to the previous lemma, this last term is of smaller order than the first under the assumptions on smoothing parameters, which completes the proof.

Employing the above results, it is now possible to establish uniform rates of convergence (on compact sets) for estimated probability functions and derivatives.

Lemma 3 (Estimated Probability Functions).

$$\sup_{\bar{x},\eta}\left|\nabla_\eta^k \hat{P}\left(\overline{w};\eta\right) - \nabla_\eta^k P\left(\overline{w}\right)\right| = O_p\left(\max\left\{1/\left[N^{1/2} h_3^{k+2}\right],\ h_3^2 h_2^2 h_1^2\right\}\right).$$

---

[3] With $\varepsilon_a > 0$ and arbitrarily small, set $a = (r_3 - \varepsilon_a)/8$. Here, a = .01 and $r_3 = 1/11$. For $\delta < r_3/2$, set:

$$r_1 = (r_3 - \delta)/4 \text{ and } r_2 = (r_3 - \delta/2)/2.$$

For these settings, $r_1 < r_2 - 2a$.

Proof of Lemma 3. The proof immediately follows from the lemmas above.

Below we will establish asymptotic normality by exploiting a "residual" property of semiparametric probability derivatives. The following lemma provides this property.

Lemma 4. Let $P(\eta)$ be the semiparametric probability function, where $P(\eta_0) = \Pr(Y_2 = 1 |\ X)$. Then, with $\nabla_\eta = \nabla_\eta^1$ as the first partial operator:

$$E\left[\nabla_\eta P(\eta)\ \mid W_1(\eta_1),\ W_2(\eta_2)\right]_{\eta\,=\,\eta_0} = 0.$$

Proof of Lemma 4. The proof of this result for the single index case is due to Whitney Newey and is contained in Klein and Spady (1993) and Klein and Sherman (2002). The extension to the double index case immediately follows from the same type of argument employed for the single index case.

As a final set of intermediate lemmas, we require results to deal with trimming. As discussed earlier, one trimming strategy below is based on a trimming sequence defined on the $X\prime s$. In particular, recall from (D6) that $\hat{\tau}_{xi}$ is a trimming indicator that is 1 on the set where each of the continuous variables is in a region defined by sample quantiles (e.g. the lower 1% and upper 99% sample quantiles). We refer to this trimming indicator as being estimated. Denote τ_{xi} as the corresponding trimming indicator with all sample quantiles replaced by their population counterparts. Lemma 5 provides a useful result relating estimated to known trimming. As such trimming occurs in normalized sums, the result below is written in this form to facilitate its subsequent use below.

Lemma 5: X-Trimming. Let r_i be random variables with $E\left|r_i\right|$ bounded. Then, under X-trimming, for any $\varepsilon > 0$:

$$\left|\frac{1}{N}\sum_{i=1}^{N}\left[\hat{\tau}_{xi} - \tau_{xi}\right] r_i\right| \leq \sum_{m=1}^{M} R_m \sum_{i=1}^{N} b_{im}\left|r_i\right|/N + o_p\left(N^{-1/2}\right) = O_p\left(N^{-(1/2)+\varepsilon}\right),$$

where M is finite, $R_m = O_p\left(N^{-(1/2)+\varepsilon}\right)$, and b_{im} is i.i.d., non-negative, and bounded.

Proof of Lemma 5. The proof for this lemma is based on an inequality due to Jim Powell for bounding $|(\hat{\tau}_{xi} - \tau_{xi})|$ from above by a "smoothed" indicator.[4] Once the indicator is smoothed, standard Taylor series arguments yield the above result. Here, ε is the "penalty" for approximating an indicator with a smooth function.

We will also be employing a trimming strategy based on the indices. Denote $\hat{\eta}_{kp}, k = 1, 2,$ as a matrix pilot estimates of nuisance parameters (obtained below under X-trimming) and define estimated indices as:

$$\hat{W}_1 \equiv X_1 + X_3 \hat{\eta}_{1p}; \; \hat{W}_2 \equiv X_2 + X_3 \hat{\eta}_{2p}$$

Recall that the smoothed trimming function in (D7) depends on $\hat{\eta}_{kp}$ and estimated sample quantiles. From (D2), we defined an underlying smooth trimming function as:

$$\tau(z; a) \equiv [\, 1 + \exp(N^a [z])]^{-1}.$$

The estimated trimming function then applies this smooth trimming function to each of the $k = 1, 2$ indices to insure that each indices stays (asymptotically) between lower and upper sample quantiles . Namely, from (D7),

$$\begin{aligned} \hat{\tau}_{wi} &\equiv \hat{\tau}_{1i}\, \hat{\tau}_{2i}, \;\; \hat{\tau}_{ki} \equiv \hat{L}_{ki} \hat{U}_{ki}, \\ \hat{L}_{ki} &\equiv \tau\left(\underline{w}_k(\hat{\eta}_p) - w_{ki}(\hat{\eta}_p); 1/12\right) \\ \hat{U}_{ki} &\equiv \tau\left(w_{ki}(\hat{\eta}_p) - \overline{w}_k(\hat{\eta}_p); 1/12\right) \quad for\ k = 1, 2. \end{aligned}$$

Here, $\underline{w}_k(\hat{\eta}_p)$ is a lower sample quantile of the $w_{ki}(\hat{\eta}_p)'s$ while $\overline{w}_k(\hat{\eta}_p)$ is the corresponding upper sample quantile. Letting η_0, λ_{kL}, and λ_{kU} be the probability limits for $\hat{\eta}_p, \underline{w}_k(\hat{\eta}_p), and\ \overline{w}_k(\hat{\eta}_p)$, τ_i is obtained from $\hat{\tau}_i$ by replacing all estimates with their probability limits. Analogously, L and U are defined by replacing all estimators by their population counterparts. To examine estimated trimming, we require a rate at which estimated quantiles ($\underline{w}_k(\hat{\eta}_p)$, $\overline{w}_k(\hat{\eta}_p)$) converge to the corresponding true quantiles. With virtually any rate sufficing, Lemma 6 below provides a rate that is subsequently employed in Lemma 7 in arguing that estimated trimming can be treated as known.

[4] This inequality was provided in a private communication to one of the authors and is contained in Klein (1993, Lemmas 1-2, and the proof for Lemma 2).

Lemma 6: Estimated Quantiles. Assuming $\left(\hat{\eta}_p - \eta_0\right) = o_p(N^{-r})$, $r < 1/2$:

$$\begin{aligned} \underline{w}_k\left(\hat{\eta}_p\right) - \lambda_{kL} &= o_p\left(N^{-r+\varepsilon}\right) \\ \overline{w}_k\left(\hat{\eta}_p\right) - \lambda_{kU} &= o_p\left(N^{-r+\varepsilon}\right). \end{aligned}$$

Proof of Lemma 6. It suffices to consider the lower α^{th} quantile with $k = 1$. With $\{\bullet\}$ is an indicator on the indicated event, for this case:

$$\sum \left\{w_{1i}\left(\hat{\eta}_p\right) < \underline{w}_1\left(\hat{\eta}_p\right)\right\} / N = \alpha.$$

Employing the same type of smooth approximation argument used in the proof of Lemma 5 and with $\varepsilon > 0$:

$$\sum \left[\left\{w_{1i}\left(\hat{\eta}_p\right) < \underline{w}_1\left(\hat{\eta}_p\right)\right\} - \left\{w_{1i}\left(\eta_0\right) < \underline{w}_1\left(\hat{\eta}_p\right)\right\}\right] / N = O_p\left(N^{-(r-\varepsilon)}\right).$$

Define $\underline{w}_{10}$ such that:

$$\sum \left\{w_{1i}\left(\eta_0\right) < \underline{w}_{10}\right\} / N = \alpha.$$

Then it follows from above that:

$$\sum \left[\left\{w_{1i}\left(\eta_0\right) < \underline{w}_1\left(\hat{\eta}_p\right)\right\} - \left\{w_{1i}\left(\eta_0\right) < \underline{w}_{10}\right\}\right] / N = O_p\left(N^{-(r-\varepsilon)}\right).$$

Letting F_N be the empirical distribution for the $w_{1i}\left(\eta_0\right)$'$s$:

$$F_N\left(\underline{w}_1\left(\hat{\eta}_p\right)\right) - F_N\left(\underline{w}_{10}\right) = O_p\left(N^{-(r-\varepsilon)}\right).$$

From the uniform convergence of the empirical distribution function to the the true distribution function, F:

$$\begin{aligned} F\left(\underline{w}_1\left(\hat{\eta}_p\right)\right) - F\left(\underline{w}_{10}\right) &= O_p\left(N^{-(r-\varepsilon)}\right) \Rightarrow \\ \underline{w}_1\left(\hat{\eta}_p\right) - \underline{w}_{10} &= O_p\left(N^{-(r-\varepsilon)}\right). \end{aligned}$$

The lemma now follows since:

$$\left|\underline{w}_1\left(\hat{\eta}_p\right) - \lambda_{1L}\right| \le \left|\underline{w}_1\left(\hat{\eta}_p\right) - \underline{w}_{10}\right| + \left|\underline{w}_{10} - \lambda_{1L}\right|.$$

For the case of index-trimming, recall that the trimming function is a smooth exponential function. In employing Taylor series arguments to analyze this function, it is important that trimming function derivatives behave as trimming functions themselves in that they severely downweight the same observations as the initial trimming function. This follows, because derivatives have the structure of being a bounded function multiplied by the initial trimming function. For example:

$$\frac{\partial}{\partial z}\tau(z) = [\tau - 1]\,\tau;\ \frac{\partial^2}{\partial z \partial z}\tau(z) = [(2\tau - 1)(\tau - 1)]\,\tau.$$

The proof of Lemma 7 below, which is essentially the same as that in Klein and Spady[1993], exploits this replicative property.

Lemma 7: Index-Trimming. Let $\left(\hat{\eta}_p - \eta_o\right) = O\left(N^{-r_p}\right)$, and assume $r_p > r_3$, with $h_3 = O_p\left(N^{-r_3}\right)$ as specified in (D4). Then, for $R_m = o_p(1)$, M is finite, and b_{im} is i.i.d. and bounded over i for each m.

$$a) \ : \ N^{-1/2}\sum\left[\hat{\tau}_{wi} - \tau_{wi}\right]\left[Y_i - P_i\right]\hat{\rho} = \sum_{m=1}^{M} R_m \sqrt{N}\sum b_{im}\tau_{wi}\left[Y_i - P_i\right]\hat{\rho}/N + o_p(1)$$

$$b) \ : \ N^{-1/2}\sum\left[\hat{\tau}_{wi} - \tau_{wi}\right]\left[\hat{P}_i - P_i\right]\hat{\rho}_i = o_p(1.)$$

Proof of Lemma 7. To establish (a), expand the components of $\hat{\tau}_{wi}$ in a Taylor series expansion, to obtain

$$\sqrt{N}\sum_i\left[\hat{\tau}_{wi} - \tau_{wi}\right]\left[Y_i - P_i\right]\hat{\rho}_i/N = \sqrt{N}\sum_{d=1}^{D} T_d/N\ ,$$

$$\begin{aligned} T_d &\equiv \sum_{s_d=1}^{S_d} R_{s_d}\sum_i b_{is_d}\tau_{wi}\left[Y_i - P_i\right]\hat{\rho}_i,\ d = 1, ..., D-1 \\ |T_D| &\leq \sum_{s_D=1}^{S_D} R_{s_D}\sum_i b_{is_D}\left|\hat{\rho}_i\right|, \end{aligned}$$

where S_d is finite, $d = 1, ..., D$ and b_{is_D} is i.i.d. over i and bounded. With D selected such that $D(r-a) > 1/2 + 2r_3$, d and D are both finite. The R-terms satisfy:

$$\begin{aligned} R_{s_d} &= O_p\left(N^{-d(r-a)}\right),\ d = 1, ..., D-1 \\ R_{sD} &= O_p\left(N^{-D(r-a)}\right),\ D(r-a) > 1/2 + 3r_3. \end{aligned}$$

The result now follows.

$$
\begin{aligned}
N^{1/2}N^{-d(r-a)}\sup \tau_i^{1/2}\left|\hat{P}_i - P_i\right| \sum_i \tau_i^{1/2} b_{isd}\left|\hat{\rho}_i\right| &= o_p(1) \\
N^{1/2}N^{-D(r-a)}\sup \left|\hat{P}_i - P_i\right| N^{2r3} &= o_p(1).
\end{aligned}
$$

The argument for (b) is similar.

To establish asymptotic normality in the next section, we will need to analyze several components that comprise the gradient. To simplify the exposition, we examine these components in Lemmas 8A-B below. In providing these results, recall that we use the notation τ_x and τ_w to refer respectively to X-trimming and Index-trimming. We employ the notation τ without x or w subscript for results that hold under either form of trimming. These gradient components have a standard form and depend on an estimated weight involving probability derivatives (see D5). With

$$
\hat{P} \equiv \left[\hat{f}_1 + \hat{\Delta}_1\right] / \left[\hat{g} + \hat{\Delta}\right] \equiv \hat{f}_1^*/\hat{g}^*
$$

denote the estimated weight as $\hat{\rho}_i^*$:

$$
\begin{aligned}
\hat{\rho}_i^* &= \nabla_\eta \hat{P}_i(\eta_0) / \left[\hat{P}_i(1-\hat{P}_i)\right] = \left[\nabla_\eta\left(\hat{f}_i^*(\eta_0)/\hat{g}_i^*(\eta_0)\right)\right]/\hat{P}_i(1-\hat{P}_i) \\
&= \frac{\hat{g}_i^*(\eta_0)\nabla_\eta \hat{f}_i^*(\eta_0) - \hat{f}_i^*(\eta_0)\nabla_\eta \hat{g}_i^*(\eta_0)}{\hat{g}_i^{*2}(\eta_0)\hat{P}_i(1-\hat{P}_i)} \equiv \frac{\hat{r}_i^*}{\hat{s}_i^*},
\end{aligned}
$$

With all $\hat{\Delta}$ adjustmen adjustment factorst ignored (they vanish exponentially under the trimming employed below), the unadjusted probability is then $\hat{P}_u \equiv \hat{f}_1/\hat{g}$ and the corresponding weight function becomes:

$$
\hat{\rho}_i = \frac{\hat{g}_i(\eta_0)\nabla_\eta \hat{f}_i(\eta_0) - \hat{f}_i(\eta_0)\nabla_\eta \hat{g}_i(\eta_0)}{\hat{g}_i^2(\eta_0)\hat{P}_{ui}(1-\hat{P}_{iu})} \equiv \frac{\hat{r}_i}{\hat{s}_i},
$$

Lemma 8A: Primary Gradient Components. Define:

$$
A_1 = \sum \tau_i\left[\,Y_i - P_i\right]\hat{\rho}_i^*/N; \qquad A_2 = \sum\left[\hat{\tau}_i - \tau_i\right]\left[\,Y_i - P_i\right]\hat{\rho}_i^*/N
$$

Then:

1) : $N^{1/2}A_1 = N^{-1/2}\sum \tau_i\left[\,Y_i - P_i\right]\rho_i + o_p(1)$
2) : $N^{1/2}A_2 = o_p(1)$, for $\hat{\tau}_i - \tau_i = \hat{\tau}_{wi} - \tau_{wi}$
3) : $N^{r_p}A_2 = o_p(1)$, for $\hat{\tau}_i - \tau_i = \hat{\tau}_{xi} - \tau_{xi}$ and $r_p > r_3$.

Proof of Lemma 8A. Beginning with A_1, in (1), let:

$$\delta \equiv N^{-1/2} \sum_i \tau_i \left[Y_i - P_i\right] \left[\hat{\rho}_i^* - \rho_i\right] = o_p(1).$$

With adjustment functions vanishing exponentially, $\delta \equiv \delta_1 + \delta_2 + o_p(1)$, where:

$$\begin{aligned} \delta_1 &= N^{-1/2} \sum_i \tau_i \left[Y_i - P_i\right] \left[\hat{\rho}_i - \rho_i\right] \left(\hat{s}_i / s_i\right) \\ \delta_2 &= N^{-1/2} \sum_i \tau_i \left[Y_i - P_i\right] \left[\hat{\rho}_i - \rho_i\right] \left[\left(\hat{s}_i / s_i\right) - 1\right]. \end{aligned}$$

Here,

$$|\delta_2| \leq N^{-1/2} \sup \left| \tau_i^{1/2} \left[\hat{\rho}_i - \rho_i\right] \right| \sup \left| \tau_i^{1/2} \left[\left(\hat{s}_i / s_i\right) - 1\right] \right|,$$

which is $o_p(1)$ from Lemma 2C. Therefore, to show that $\delta = o_p(1)$, it suffices to show $\delta_1 = o_p(1)$. We have:

$$\delta_1 = N^{-1/2} \sum_i \tau_i \left[Y_i - P_i\right] \left[s_i \left(\hat{r}_i - r_i\right) - r_i \left(\hat{s} - s_i\right) / s_i^2\right].$$

In what follows, we analyze the first term above, with an analogous argument holding for the second. Let :

$$\Delta \equiv N^{-1/2} \sum_i \tau_i \left[Y_i - P_i\right] s_i \left(\hat{r}_i - r_i\right)$$

and write:

$$\begin{aligned} \Delta &= N^{-1/2} \sum_i \tau_i \left[Y_i - P_i\right] s_i \left[\left(\hat{g}_i - g_i\right) \nabla_\eta f_i + \left(\nabla_\eta \hat{f}_i - \nabla_\eta f_i\right) \hat{g}_i\right] \\ &\equiv .\Delta_{11} + \Delta_{12} \end{aligned}$$

As the analysis for both of the above terms is similar, here we analyze the first. Exploiting the fact that $[Y_i - P_i]$ has 0 conditional expectation, we show that $\Delta_{11} = o_p(1)$ by showing that its expected square converges to zero. With $\varepsilon_i \equiv s_i \left(\hat{g}_i - g_i\right) \nabla_\eta f_i$:

$$\begin{aligned} E\left(\Delta_{11}^2\right) &= E\left[\sum_i \tau_i^2 \left[Y_i - P_i\right]^2 \varepsilon_i^2 / N\right] + \\ & \sum_{i \neq j} \sum E\left[\left(\tau_i \left[Y_i - P_i\right] \varepsilon_i\right) \left(\tau_j \left[Y_j - P_j\right] \varepsilon_j\right)\right] / N. \end{aligned}$$

The first term is bounded from above by:

$$\sum_i E\left(\tau_i^2 \varepsilon_i^2\right)/N,$$

which converges to zero. Employing the fact that $E\left[Y_i - P_i \mid \mathrm{X}_i\right] = 0$, it can be shown that the second term also converges to zero. The result now follows.

Turning to (2), the argument for the smooth index-trimming function is based on a Taylor expansion of $\hat{\tau}_{wi}$, the observation that the derivative of a trimming function behaves as a trimming function, and the proof for A_1 above. Lemma 7 contains the details of this argument from which (2) follows. For (3), the argument is based on a characterization result for indicator X-trimming in Lemma 5.

Lemma 8B: Secondary Gradient Components. Define

$$B_1 = \sum \tau_i \left[\hat{P}_i - P_i\right] \hat{\rho}_i^* / N; \quad B_2 = \sum \left[\hat{\tau}_i - \tau_i\right] \left[\hat{P}_i - P_i\right] \hat{\rho}^* / N.$$

Then:

$$\begin{aligned} 1) &: \; N^{1/2} B_1 = o_p(1) \text{ and } N^{1/2} B_2 = o_p(1) \text{ for } \hat{\tau}_i - \tau_i = \hat{\tau}_{wi} - \tau_{wi} \\ 2) &: \; N^{r_p} B_1 = o_p(1), \, N^{r_p} B_2 = o_p(1) \text{ for: } \hat{\tau}_i - \tau_i = \hat{\tau}_{xi} - \tau_{xi}, \; r_p > r_3. \end{aligned}$$

Proof of Lemma 8B. As above with adjustment factors vanishing exponentially, we may replace $\hat{\rho}_i^*$ with $\hat{\rho}_i$ throughout. Proceeding with this subsittution, we first simplify B_1 by showing::

$$\Delta \equiv N^{-1/2} \sum_i \tau_{wi} \left[\hat{P}_i - P_i\right] \left[\hat{\rho}_i - \rho_i\right] = o_p(1).$$

Bounding this term:

$$\begin{aligned} |\Delta| &= N^{-1/2} \left| \sum_i \tau_{wi} \left[\hat{P}_i - P_i\right] \left[\hat{\rho}_i - \rho_i\right] \right| \\ &\leq N^{1/2} \sup \left| \tau_{wi}^{1/2} \left[\hat{P}_i - P_i\right] \right| \sup \left| \tau_{wi}^{1/2} \left[\hat{\rho}_i - \rho_i\right] \right|, \end{aligned}$$

which is $o_p(1)$ from Lemma 3. Therefore:

$$N^{1/2}\mathbf{B}_1 = N^{-1/2}\sum_i \tau_{wi}\left[\hat{P}_i - P_i\right]\rho_i + o_p(1).$$

To further simplify $\mathbf{B}_1$ and show that it is $o_p\left(N^{-1/2}\right)$, note that under an argument similar to that above:

$$N^{-1/2}\sum_i \tau_{wi}\left[\hat{P}_i - P_i\right]\rho_i\left[(\hat{g}_i/g_i) - 1\right] = o_p(1),$$

which implies:

$$N^{1/2}\mathbf{B}_1 = N^{-1/2}\sum_i \tau_{wi}\left[\hat{P}_i - P_i\right]\rho_i\left(\hat{g}_i/g_i\right) + o_p(1).$$

Next, recall that $\hat{P}_i = \left[\hat{f}_i + \hat{\Delta}_{1N}\right] / \left[\hat{g}_i + \hat{\Delta}_N\right]$. Under τ-trimming, the Δ-adjustment factors and their derivatives vanish exponentially when evaluated at the true densities. Accordingly, under a Taylor series argument, we may replace $\hat{P}_i$ with $\hat{f}_i/\hat{g}_i$ to obtain:

$$N^{1/2}\mathbf{B}_1 = N^{-1/2}\mathbf{B}_1^* + o_p(1), \;\; \mathbf{B}_1^* \equiv \sum_i \tau_{wi}\left[\hat{f}_i - P_i\hat{g}_i\right]e_i, \;\; e_i \equiv \rho_i/g_i.$$

Noting that e_i has expectation conditioned on the indices of 0 (Lemma 4), employ the same type of mean-square convergence argument used to analyze A. We have:

$$\begin{aligned} E\left[\left(N^{1/2}\mathbf{B}_1^*\right)^2\right] &= \frac{1}{N}E\left[\sum_i \tau_{wi}^2\left[\hat{f}_i - P_i\hat{g}_i\right]^2 e_i^2\right] + C, \\ C &= \sum_{i\neq j}\sum E\left[\tau_{wi}\left(\hat{f}_i - P_i\hat{g}\right)\tau_{wj}\left(\hat{f}_j - P_j\hat{g}_j\right)e_ie_j\right]/N. \end{aligned}$$

It can readily be shown directly that the first term above vanishes for large N. Turning to the cross-product terms in C, from iterated expectations:

$$\begin{aligned} C &= EE\left[\tau_{wi}\left(\hat{f}_i - P_i\hat{g}\right)\left(\tau_{wj}\left[\left(\hat{f}_j - P_j\hat{g}_j\right)\right]e_ie_j\right) \mid X\right] \\ &= E\left[E\left[\tau_{wi}\left(\hat{f}_i - P_i\hat{g}\right)\left(\tau_{wj}\left[\left(\hat{f}_j - P_j\hat{g}_j\right)\right]\right) \mid X\right]e_ie_j\right]. \end{aligned}$$

As the inner expectation only depends on the indices $W : Nx2$, denote this inner expectation as $H(W)$ and write:

$$C = E\left[H(W)\, e_i e_j\right] = E\left[H(W)\, E\left[e_i e_j\right] \mid W\right] = 0,$$

with the last result directly following from Lemma 4. The proof for B_1 in (1) now follows.

The proof for B_2 in (1), which is analogous to that for A_2 in Lemma 8A, part (2), readily follows from Lemma 7. To establish (2), we need to analyze B_1 and B_2 under X-trimming, The argument here, which is essentially the same as that for A_2 in Lemma 8A, part 3, follows from Lemma 5.

1.2 Main Results

As in the previous section, throughout this section, all results are provided under Assumptions (A1-6) and Definitions (D1-7).

Theorem 1. With $\hat{\tau}_i = \hat{\tau}_{xi}$ or $\hat{\tau}_{wi}$, define the quasi-likelihood as in Section 4.1:

$$\hat{Q}(\eta) \equiv \frac{1}{N}\sum_{i=1}^{N} \hat{\tau}_i \left(Y_{2i} Ln\left[\hat{P}_i(\eta)\right] + \left[1 - Y_{2i}\right] Ln\left[1 - \hat{P}_i(\eta)\right] \right)$$

and define $\hat{\eta} \equiv \arg\sup \hat{Q}(\eta)$. Then : $\hat{\eta} \xrightarrow{p} \eta_0$, the vector of true parameter values.

Proof of Theorem 1. Employing (D5) and deleting the i subscript for notational simplicity, define the probability functions:

$$\begin{aligned} \hat{P}(\eta) &\equiv \left[\hat{f}_1 + \hat{\Delta}_1\right] / \left[\hat{g} + \hat{\Delta}\right] \\ P_N(\eta) &\equiv \left[f_1 + \Delta_N\right] / \left[g + \Delta_N\right] \\ P(\eta) &\equiv f_1/g. \end{aligned}$$

With $P_N(\eta)$ replacing $\hat{P}(\eta)$ throughout, denote $Q_N(\eta)$ as the corresponding objective function. Finally, denote $Q(\eta)$ as the objective function obtained by replacing $\hat{P}(\eta)$ with $P(\eta)$ throughout. Then:

$$\left|\hat{Q}(\eta) - Q(\eta)\right| \leq \left|\hat{Q}(\eta) - Q_N(\eta)\right| + \left|Q_N(\eta) - Q(\eta)\right|.$$

Employing arguments similar to those in Klein and Spady (1993) and Lemma 4, it can be shown that each of these terms vanish in probability, uniformly in η.[5] Next, write:

$$\begin{aligned} Q(\eta) &\equiv \frac{1}{N}\sum_{i=1}^{N}\hat{\tau}_i\left[\,Y_{2i}Ln\left[P_i(\eta)\right]+\left[1-Y_{2i}\right]Ln\left[1-P_i(\eta)\right]\right]\equiv\bar{Q}(\eta)+R, \\ \bar{Q}(\eta) &\equiv \frac{1}{N}\sum_{i=1}^{N}\tau_i\left[\,Y_{2i}Ln\left[P_i(\eta)\right]+\left[1-Y_{2i}\right]Ln\left[1-P_i(\eta)\right]\right] \\ R &\equiv \frac{1}{N}\sum_{i=1}^{N}\left[\,\hat{\tau}_i-\tau_i\right]\left[\,Y_{2i}Ln\left[P_i(\eta)\right]+\left[1-Y_{2i}\right]Ln\left[1-P_i(\eta)\right]\right]. \end{aligned}$$

It can be shown that R vanishes in probability, uniformly in η. From standard uniform convergence arguments:

$$\sup_{\eta}\left|\bar{Q}(\eta)-E\left[\bar{Q}(\eta)\right]\right|\xrightarrow{p}0.$$

Employing the identification condition in (A5), $E\left[\bar{Q}(\eta)\right]$ is uniquely maximized at η_0, which completes the argument.

Theorem 2. Defining $H_0\equiv\nabla_{\eta}^{2}E\left(L\left(\eta_0\right)\right)$:

$$\sqrt{N}\left[\hat{\eta}-\eta_0\right]\xrightarrow{d}N\left(0,-H_0^{-1}\right).$$

Proof of Theorem 2. With the quasi-likelihood defined under X-trimming (see D6) and with $\eta^{+}\ \epsilon\ \left[\hat{\eta},\ \eta_0\right]$, from a standard Taylor series expansion:

$$\begin{aligned} N^{r_p}\left[\hat{\eta}_p-\eta_0\right] &= -\hat{H}\left(\eta^{+}\right)^{-1}N^{r_p}\hat{G}(\eta_0), \\ \hat{H}\left(\eta^{+}\right) &= \nabla_{\eta}^{2}\hat{L}\left(\eta^{+}\right),\quad \hat{G}(\eta_0)=\nabla_{\eta}^{1}\hat{L}\left(\eta_0\right), \end{aligned}$$

[5]In analyzing the first term, it is important to exploit the fact that the δ-adjustment factors behaving as trimming functions in that they control the rate which denominators in various expressions tend to zero [see Klein and Spady (1993, proof of lemma 4, p. 414)]
To analyze the second term, it is important to note that from the assumption of bounded $X\prime s$, it follows that $P\left(\eta_o\right)$ is strictly bounded away from one and zero. It then follows that $P(\eta)$, a conditional expectation of $P\left(\eta_o\right)$, is also strictly bounded away from one and zero. While the assumption of bounded $X\prime s$ could be replaced by tail conditions, this assumption considerably simplifies the argument for the second term. [see Klein and Spady (1993, Proof of Theorem 3, p. 415)].

where we have employed X-trimming. Beginning with the Hessian component, as in the previous theorem define the probability functions: $\hat{P}(\eta)$, $P_N(\eta)$, and $P(\eta)$. From Lemma 3 and arguments very similar to those employed to analyze the averaged likelihood in Theorem 1, it can be shown that:

$$\sup_{\eta} \left| \hat{H}(\eta) - H(\eta) \right| \xrightarrow{p} 0.$$

From standard uniform convergence arguments, $H(\eta)$ converges in probability and uniformly in η to its expectation. It follows that $\hat{H}(\eta^{+})^{-1} = H_0^{-1}(\eta_0) + o_p(1)$. Therefore, a convergence rate for the pilot estimator, $\hat{\eta}_p$, will follow from the rate at which the gradient converges to zero.

In the notation of Lemmas 8A and 8B:

$$N^{r_p}\hat{G}(\eta_0) = N^{r_p}[A_1 + A_2] + N^{r_p}[B_1 + B_2]$$

From Lemmas 8A and 8B, it now follows that:

$$N^{r_p}\left[\hat{\eta}_p - \eta_0\right] = o_p(1), \; r_p > r_3.$$

Employing the $\hat{\eta}_p$ to construct a smooth Index-trimming function, employ the quasi-likelihood under Index-trimming (D7) and a Taylor series expansion to obtain:

$$N^{1/2}[\hat{\eta} - \eta_0] = -\hat{H}\left(\eta^{+}\right)^{-1} N^{1/2}\hat{G}(\eta_0).$$

As above, $\hat{H}(\eta^{+})^{-1} = H_0^{-1}(\eta_0) + o_p(1)$. From Lemmas 8A and 8B:

$$\begin{aligned} N^{1/2}\hat{G}(\eta_0) &= N^{-1/2}\sum \tau_{wi}[Y_i - P_i]\rho_i + o_p(1) \\ &\equiv N^{1/2}G(\eta_0) + o_p(1), \end{aligned}$$

where $G(\eta_0)$ is the gradient term with all estimated functions replaced by their (uniform) probability limits. The theorem now follows from a standard central limit theorem.

Turning to the outcomes equation, recall that it is given as:

$$Y_1 = Z\theta_o + u,$$

$\theta_o \equiv [\beta_o,\ \mu_o]$ and $Z \equiv [X,\ Y_2]$. Then, the IV estimator is given as :

$$\hat{\alpha}_{IV} = \left[\hat{Z}^* (\hat{\eta})' Z\right]^{-1} \hat{Z}^* (\hat{\eta})' Y_1,\ \hat{Z}^* (\eta) \equiv \left[X,\ \hat{P}(\eta)\right]$$

Consistency and asymptotic normality (Theorem 3 of Section 4.2) will now be immediate if the conditions given in the next lemma hold.

Lemma 8: With $Z^* \equiv [X,\ P(\eta_0)]$, under Assumptions (A1-4) and Definitions (D1-5):

$$\begin{aligned} 1) &: \left[\hat{Z}^* (\hat{\eta})' Z - Z^{*\prime} Z\right] /N = o_p(1), \\ 2) &: \sqrt{N} \left[\hat{Z}^* (\hat{\eta})' u - Z^* u\right] /N = o_p(1). \end{aligned}$$

Proof of Lemma 8. The first condition follows from Theorem 2 and Lemma 3. The second condition follows from a standard U-Statistics argument and is to be expected from Newey and McFadden (Handbook of Econometrics, vol. 4, Chapter 36, section 6.2 and Theorem 6.2).