Appendix to: Semiparametric Estimation of a Credit Rating Model

Abstract

In this Appendix, I prove the two asymptotic theorems in the paper. Lemma 2.1 - 2.4 establish various convergence rates for the employed "recursive differencing" estimator. Lemma 2.5 establishes the convergence rate for the Hessian matrix associated with the log-likelihood function. Lemma 2.9 proves a "residual property" of the semiparametric probability derivative, which I use in conjunction with the recursive differencing estimator to reduce bias. These intermediate lemmas are instrumental in establishing the asymptotic property of the index parameter estimator.

Contents

1 Proofs of Asymptotic Results **1**

1.1 Proof of Theorem B1 1

1.2 Proof of Theorem B2 2

2 Intermediate Lemmas and Proofs **6**

1 Proofs of Asymptotic Results

1.1 Proof of Theorem B1

Proof. To show consistency, recall that θ maximize the estimated average log-likelihood function $\hat{Q}(\theta)$. From Lemma C.3, $\sup_\theta|\hat{Q}(\theta) - \tilde{Q}(\theta)| \xrightarrow{p} 0$, where $\tilde{Q}(\theta)$ is obtained from $\hat{Q}(\theta)$ by replacing all estimated functions with their probability limits. From standard argument, $\tilde{Q}(\theta)$ converges uniformly to its expectation $\mathrm{E}[\tilde{Q}(\theta)]$. Then, consistency would follow as long as $\mathrm{E}[\tilde{Q}(\theta)]$ is uniquely maximized at θ_0.

To show θ_0 uniquely maximize $\mathrm{E}[\tilde{Q}(\theta)]$, I apply the law of iterated expectation on $\mathrm{E}[\tilde{Q}(\theta)]$ by conditioning on the true index $V_i(\theta_0) = (F_i\theta_0^F, B_i\theta_0^B, MFOI_i)$:

$$\begin{aligned} E[\tilde{Q}(\theta)] &= \frac{1}{N} E_{V_i}\{E \sum_{i=1}^{N}\sum_{k=1}^{7} Y_i^k Ln(\hat{P^k}(V_i(\theta)))|V_i(\theta_0)\} \quad (1)\\ &= \frac{1}{N} E \sum_{i=1}^{N}\sum_{k=1}^{7} P_{0i}^k Ln(\hat{P^k}(V_i(\theta))), \quad P_{0i}^k \equiv E[Y_i^k|V_i(\theta_0)] \quad (2)\end{aligned}$$

If θ_0 maximize $\mathcal{L}_i(\theta) \equiv \sum_{k=1}^{7} P_{0i} Ln(\hat{P^k}(V_i(\theta)))$ for each i, it will clearly maximize $E[\tilde{Q}(\theta)]$. Let $P_k(\theta)$ denote the estimated rating probability inside this quantity, $\hat{P^k}(V_i(\theta))$. For each k from 1 to 7, take the derivative of $\mathcal{L}_i(\theta)$ with respect to P_k gives

$$\nabla_k \mathcal{L}_i(\theta) = \frac{P_{0i}^k}{P_k(\theta)} - \frac{1-\sum_{j\neq k} P_{0i}^j}{1-\sum_{j\neq k} P_j(\theta)}, \quad k = 1, 2, \cdots, 7 \quad (3)$$

When $\hat{P^k}(V_i(\theta)) = P_{i0}^k$, it can be seen that $\nabla_k \mathcal{L}_i(\theta) = 0$ for all k from 1 to 7. From Theorem 1 and 2 in Klein and Spady (1993), $\hat{P^k}(V_i(\theta)) = P_{i0}^k \iff \theta = \theta_0$. □

1.2 Proof of Theorem B2

Proof. From a Taylor expansion of the estimated gradient on $\hat{\theta}$ and the fact that the estimated gradient is zero evaluated at θ_0, we have

$$\sqrt{N}(\hat{\theta}-\theta_0) = -\hat{H}(\theta^+)^{-1}\sqrt{N}\hat{G}(\theta_0) \quad \theta^+ \in (\theta_0, \hat{\theta}) \quad (4)$$

where $\hat{G}(\theta) = \nabla_{\theta'}\widehat{Q}_2(\theta)$, $\hat{H}(\theta) = \nabla_{\theta'\theta}\widehat{Q_2}(\theta)$ for any θ in its support. Under the condition of Lemma 2.5, $\hat{H}(\theta^+)$ converges in probability to $H_0 \equiv E[H(\theta_0)]$. We let $P_{ki} \equiv P^k(V_i)$ to simplify the notation,

$$\begin{aligned} \sqrt{N}\hat{G}(\theta_0) &\equiv \frac{\sqrt{N}}{N}\sum_{i=1}^{N}\widehat{\tau_{iv}}\sum_k \frac{Y_i^k}{\widehat{P_{ki}^*}}\frac{\partial \widehat{P_{ki}^*}}{\partial\theta}|_{\theta=\theta_0} \quad (5)\\ &= \frac{\sqrt{N}}{N}\sum_{i=1}^{N}\widehat{\tau_{iv}}\sum_k \frac{Y_i^k-\widehat{P_{ki}^*}}{\widehat{P_{ki}^*}}\frac{\partial \widehat{P_{ki}^*}}{\partial\theta}|_{\theta=\theta_0} \quad (6)\\ &= \frac{\sqrt{N}}{N}\sum_{i=1}^{N}\widehat{\tau_{iv}}[\frac{Y_i^1-\widehat{P_{1i}^*}}{\widehat{P_{1i}^*}}\frac{\partial \widehat{P_{1i}^*}}{\partial\theta}|_{\theta=\theta_0} + \frac{Y_i^2-\widehat{P_{2i}^*}}{\widehat{P_{2i}^*}}\frac{\partial \widehat{P_{2i}^*}}{\partial\theta}|_{\theta=\theta_0} + \ldots] \quad (7)\end{aligned}$$

Noting that $\sum_k Y_i^k = 1$ and $\sum_k \nabla_\theta P_{ki} = 0$ for all i. The first equality above follows from the fact that $\sum_k \frac{\partial \widehat{P^*_{ki}}}{\partial\theta} = \frac{\partial \sum_k \widehat{P^*_{ki}}}{\partial\theta} = \frac{\partial 1}{\partial\theta} = 0$. As the first term in (7) and all remaining terms have the same structure, it suffices to analyze the first term. From standard argument in Pakes and Pollard (1989) and Klein and Spady (1993), the estimated trimming function $\widehat{\tau_{iv}}$ can be replaced by the truth asymptotically. For any $k \in \{1, 2..., 7\}$, with $\widehat{w_{ki}} \equiv \tau_{iv} \frac{\partial \widehat{P^*_{ki}}}{\widehat{P^*_{ki}}\partial\theta}|_{\theta=\theta_0}$, a representative term in the estimated gradient is:

$$
\begin{aligned}
\frac{\sqrt{N}}{N}\sum_{i=1}^{N}[Y_i^k - \widehat{P^*_{ki}}]\widehat{w_{ki}} &= \underbrace{\frac{\sqrt{N}}{N}\sum_{i=1}^{N}[Y_i^k - P_{ki}]\widehat{w_{ki}}}_{D_1} - \underbrace{\frac{\sqrt{N}}{N}\sum_{i=1}^{N}[\widehat{P^*_{ki}} - P_{ki}]\widehat{w_{ki}}}_{D_2} && (8)\\
&= \underbrace{\frac{\sqrt{N}}{N}\sum_{i=1}^{N}[Y_i^k - P_{ki}]w_{ki}}_{D_1^*} + \underbrace{\frac{\sqrt{N}}{N}\sum_{i=1}^{N}[\widehat{P^*_{ki}} - P_{ki}]w_{ki}}_{D_2^*} && (9)\\
&+ \quad (D_1 - D_1^*) + (D_2 - D_2^*) && (10)
\end{aligned}
$$

By making use of the fact that the residual $Y_i^k - P_{ki}$ has zero conditional expectation, $D_1 - D_1^* \equiv N^{-1/2}\sum_i (Y_i^k - P_{ki})(\widehat{w_{ki}} - w_{ki}) = o_p(1)$ through a mean-square convergence results similar to Klein and Spady (1993). For the convergence of double-sum in Lemma 2.7, $D_2 - D_2^* = o_p(1)$.

Recall from D.4 that the recursive differencing estimator has the form:

$$
\widehat{P^*}_k(v) \quad \equiv \quad \frac{\frac{1}{N-1}\sum_j [Y_j^K - \hat{\Delta}_j(v)]\mathbf{K}_h(V_j - v)}{\frac{1}{N-1}\sum_j \mathbf{K}_h(V_j - v)} = \hat{f}/\hat{g} \tag{11}
$$

with $\hat{g}$ being and estimated density derivative and $\widehat{\Delta}_j^k(v) = \widehat{P}^K(V_j) - \widehat{I}^K(v)$ is an estimate of the localization error. Note that

$$
\begin{aligned}
D_2^* &= \frac{\sqrt{N}}{N}\sum_{i=1}^{N}[\widehat{P^*_{ki}} - P_{ki}]\frac{\hat{g}_i}{g_i}w_{ki} + \underbrace{\frac{\sqrt{N}}{N}\sum_{i=1}^{N}[\widehat{P^*_{ki}} - P_{ki}][1 - \frac{\hat{g}_i}{g_i}]w_{ki}}_{o_p(1)\ from\ Lemma\ 2.7} && (12)\\
&= \frac{\sqrt{N}}{N}\sum_{i=1}^{N}[\widehat{P_{ki}} - P_{ki}]\frac{\hat{g}_i}{g_i}w_{ki} && (13)\\
&\equiv \sqrt{N}U_N && (14)
\end{aligned}
$$

where $\widehat{P_{ki}}$ is the standard Nadaraya Watson estimator: $\widehat{P_{ki}} = \frac{\frac{1}{N-1}\sum_j Y_j^k \mathbf{K}_h(V_j - V_i)}{\frac{1}{N-1}\sum_j \mathbf{K}_h(V_j - V_i)} = \widehat{f}/\widehat{g}$ Lemma 2.8 establishes the equivalence between (12) and (13). In below, we show that this term is a degenerate $U-$statistic. Let

$$U_N \equiv \frac{1}{N}\sum_{i=1}^{N}(\widehat{P_{ki}} - P_{ki})w_{ki}\frac{\hat{g}_i}{g_i} = \frac{1}{N}\sum_{i=1}^{N}(\hat{f}_i - P_{ki}\hat{g}_i)\frac{w_{ki}}{g_i} \tag{15}$$

With $\widehat{f}_i = \frac{1}{N-1}\sum_j Y_j^k \mathbf{K}_h(V_j - V_i)$ and $\widehat{g}_i = \frac{1}{N-1}\sum_j \mathbf{K}_h(V_j - V_i)$,

$$U_N = \frac{1}{N}\sum_{i=1}^{N}(\frac{1}{N-1}\sum_j Y_j^k \mathbf{K}_h(V_j - V_i) - P_{ki}\frac{1}{N-1}\sum_j \mathbf{K}_h(V_j - V_i))\frac{w_{ki}}{g_i} \tag{16}$$

$$= \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}[Y_j^k - P_{ki}]\mathbf{K}_h(V_j - V_i)\frac{w_{ki}}{g_i} \tag{17}$$

$$= \frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}\rho_{ij}^* \tag{18}$$

$$\text{with} \quad \rho_{ij} \equiv [Y_j^k - P_{ki}]\mathbf{K}_h(V_j - V_i)\frac{w_{ki}}{g_i} \tag{19}$$

$$\text{and} \quad \rho_{ij}^* = \frac{\rho_{ij} + \rho_{ji}}{2} \tag{20}$$

Note that by construction, ρ_{ij}^* is now symmetric with respect to i, j. After applying the standard approximation theory of U-statistics (Powell et al., 1989, Serfling, 2009), we obtain $\sqrt{N}(U_N - \hat{U}_N) = o_p(1)$ where $\hat{U}_N$ is a "projection" of U_N:

$$\sqrt{N}\hat{U}_N \equiv \sqrt{N}\sum_{i=1}^{N}(E[\rho_{ji}|V_i] + E[\rho_{ij}|V_i])/N = T_1 + T_2 \tag{21}$$

Each term in T_2 is zero from the law of iterated expectation and the residual property of

w_{ki}:

$$\begin{aligned}
E[\rho_{ij}|V_i] &= E[(Y_j^k - P_{ki})\mathbf{K}_h(V_j - V_i)\frac{w_{ki}}{g_i}|V_i] && (22)\\
&= E[(Y_j^k - P_{ki})\mathbf{K}_h(V_j - V_i)\tau_{iv}\frac{1}{P_{ki}g_i}\frac{\partial P_{ki}}{\partial \theta}|_{\theta=\theta_0}|V_i] && (23)\\
&= E_{V_j}\{E[(Y_j^k - P_{ki})\mathbf{K}_h(V_j - V_i)\tau_{iv}\frac{1}{P_{ki}g_i}\frac{\partial P_{ki}}{\partial \theta}|_{\theta=\theta_0}|V_i, V_j]\} && (24)\\
&= (Y_j^k - P_{ki})\mathbf{K}_h(V_j - V_i)\tau_{iv}\frac{1}{P_{ki}g_i}E_{V_j}\underbrace{\{E[\frac{\partial P_{ki}}{\partial \theta}|_{\theta=\theta_0}|V_i, V_j]\}}_{=0 \text{ from Lemma 2.9}} && (25)\\
&= 0 && (26)
\end{aligned}$$

It can be shown that for each term in T_1:

$$\begin{aligned}
E[term_i] &\equiv E_{V_i}\{E[\rho_{ji}|V_i]\} = E[\rho_{ji}] = E_{V_j}\{E[\rho_{ji}|V_j]\} = 0 && (27)\\
Var[term_i] &= O(1) && (28)
\end{aligned}$$

T_1 is therefore $o_p(1)$ from standard sample mean property. Since $\sqrt{N}\hat{U}_N$ converges in probability to zero, $\sqrt{N}U_N$ and therefore D_2^* are both $o_p(1)$. As such, the only term remain in the gradient expression in (8) is $D_1^* = \frac{\sqrt{N}}{N}\sum_{i=1}^N [Y_i^k - P_{ki}]w_{ki}$, so

$$\sqrt{N}\hat{G}(\theta_0) = \sqrt{N}\sum_{i=1}^N \tau_{iv}\frac{Y_i^k - P_i^k}{P_i^k}\frac{\partial P_i^k}{\partial \theta}/N + o_p(1) \quad (29)$$

Then, referring to the expression in (4), $\sqrt{N}(\hat{\theta} - \theta_0)$ has an asymptotic linear form:

$$\begin{aligned}
\sqrt{N}(\hat{\theta} - \theta_0) &= \sqrt{N}\sum_{i=1}^N H_0^{-1}G_i(\theta_0)/N + o_p(1) && (30)\\
G_i(\theta_0) &\equiv \sum_{k=1}^7 \tau_{iv}\frac{Y_i^k - P^k}{P^k}\frac{\partial P^k}{\partial \theta}|_{\theta=\theta_0} && (31)
\end{aligned}$$

We apply the Lindberg-Levy CLT and obtain:

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, H_0^{-1}E[G_i(\theta_0)G_i'(\theta_0)]H_0^{-1}) \quad (32)$$

□

2 Intermediate Lemmas and Proofs

Lemma 2.1 (Uniform Convergence of Estimates to Expectations)**.** *Let w be a K dimensional vector and assume that $m(w)$ is a sample average of terms $m(w; z_i)$, where z_i are i.i.d. Assume that with $h \to 0$, we have uniformly over N:*

$$h^{r+1}|m(w;z_i)| < c, \quad r+1>0 \quad \textit{and} \quad h^s|\partial m(w;z_i)/\partial w| < c, \quad s>0$$

Let $E[m(w)]$ be the expectation of $m(w)$ taken over the distribution of z_i. Then, with w in a compact set and for any $\alpha > 0$:

$$N^{(1-\alpha)/2}h^{r+1}sup|m(w) - E[m(w)]| \xrightarrow{p} 0 \quad a.s.$$

Proof. See the proof of Lemma 2.1 on pp 411 in Klein and Spady (1993). The proof utilizes important implication in Hoeffding (1963) and Bhattacharya (1967). □

Lemma 2.2 (Convergence Property for Density Estimator and Its Derivatives)**.** *Let $\hat{g}$ be the estimated index density defined in D.3 and $\nabla^r_\theta \hat{g}$ be the r-th density derivative with respect to θ, when $r = 0, 1, 2$, then:*

(2.2.1) $sup_{v,\theta} E\{(\nabla^r_\theta \hat{g}(v) - E[\nabla^r_\theta \hat{g}(v)])^2]\} = O_p(\frac{1}{Nh^{2r+3}})$

(2.2.2) $sup_{v,\theta}|E[\nabla^r_\theta \hat{g}(v)] - \nabla^r_\theta g(v)| = O_p(h^2)$

Proof. The order of bias and variance for estimated kernel density and density derivatives are fairly standard in the literature, see Hansen (2009) for an discussion. □

Lemma 2.3 (Convergence Properties of Estimated Probability after Recursive Differencing)**.** *The following convergence properties hold for the conditional probability estimator defined above:*

(1) $sup_v E\{(\hat{g}(v)[\widehat{P}^*_k(v) - E[\widehat{P}^*_k(v)])^2\}|_{\theta=\theta_0} = O_p(\frac{1}{Nh^3})$

(2) $sup_v|E[\hat{g}(v)(\widehat{P}^*_k(v) - P^k(v))]|_{\theta=\theta_0} = O_p(h^4)$

(3) $sup_{v,\theta}\nabla^t_\theta|\widehat{P}^*_k(v) - P^k(v)| = O_p(a_N)$*, where* $a_N = (\frac{lnN}{Nh^{3+2t}})^{1/2} + h^4$ *with* $t = 0, 1, 2.$

Proof. See Theorem 1 and Lemma 11 in Shen and Klein (2019) for the proofs of part (1) and (2). In particular, they demonstrated that a lower order of bias can be achieved after implementing the recursive differencing strategy, without causing the order of variance to go up. As illustrated in the first two results, the order of the variance here is the same compared to that with a regular kernel in Lemma 2.2, while a lower order bias is obtained (h^4 vs h^2). For the uniform convergence rate in (3), the first component of a_N is derived following Lemma B.1 of Newey (1994) whereas the second component comes from the bias. □

Lemma 2.4 (Mean-square Convergence of Estimated Probability)**.** *For recursive differencing estimator $\hat{P}^*_k(v)$ defined in the main text, for $h = 0.97\sigma N^{-r}$ and under the conditions of Lemma 2.1- 2.3,*

$$\frac{1}{N}\sum_i (\hat{P}^*_{ki}(v) - P_{ki}(v))^2 = O_p(N^{-1}h^{-3}) + O_p(h^8) \tag{33}$$

Proof. With $\hat{f}^*_i(v) \equiv \hat{g}_i(v)\hat{P}^*_{ki}(v)$, note that

$$\begin{aligned} \frac{1}{N}\sum_i (\hat{P}^*_{ki}(v) - P_{ki}(v)) &= \frac{1}{N}\sum_i (\frac{\hat{f}^*_i(v)}{\hat{g}_i(v)} - P_{ki}(v)) &(34)\\ &\leq sup(\frac{1}{\hat{g}_i})\frac{1}{N}\sum_i (\hat{f}^*_i(v) - P_{ki}(v)\hat{g}_i(v)) &(35)\end{aligned}$$

Since we use the trimming function defined in D.5 to keep $\hat{g}_i$ away from zero, $1/inf(\hat{g}_i)$ is bounded from above. Due to the recursive differencing structure in $\hat{f}^*(v)$ that is explained in Lemma 2.3, for some constant B $= O(1)$,

$$\begin{aligned} \frac{1}{N}\sum_i (\hat{P}^k_i(v) - P^k_i(v))^2 &\leq B^2\frac{1}{N}\sum_i (\hat{f}^*_i - P^k_i(v)\hat{g}_i)^2 &(36)\\ &= B^2\frac{1}{N}\sum_i (\hat{f}^*_i - E[\hat{f}^*_i] + E[\hat{f}^*_i] - P^k_i(v)\hat{g}_i)^2 \\ &= \underbrace{B^2\frac{1}{N}\sum_i (\hat{f}^*_i - E[\hat{f}^*_i])^2}_{\text{variance of } \hat{f}^*_i} + \underbrace{B^2\frac{1}{N}\sum_i (E[\hat{f}^*_i] - P_{ki}(v)\hat{g}_i)^2}_{\text{squared bias of } \hat{f}^*_i} \\ &+ \underbrace{B^2\frac{2}{N}\sum_i (\hat{f}^*_i - E[\hat{f}^*_i])(E[\hat{f}^*_i] - P^k_i(v)\hat{g}_i)}_{o_p(1)} \\ &= O^2_p(1)(O_p(N^{-1}h^{-3}) + O_p(h^8)) &(37)\end{aligned}$$

$\square$

Lemma 2.5 (Convergence of Hessian)**.** *Assume the kernel bandwidth* $h = 0.97\sigma N^{-r}$ *and* $1/16 < r < 1/7$*. Then, under the conditions of lemma 2.1 and with* $\theta^+ \in [\hat{\theta}, \theta_0]$*,*

$$\hat{H}(\theta^+)^{-1} \xrightarrow{p} H_0 = E[H(\theta_0)]$$

Proof. Given that the Hessian is a continuous function on θ, the desired argument $\hat{H}(\theta^+) \xrightarrow{p} E[H(\theta_0)]$ would follow if (a) $\theta^+ \xrightarrow{p} \theta_0$ and (b) $\sup|\hat{H}(\theta) - E[H(\theta)]| \xrightarrow{p} 0$. Condition (b) implies that $\hat{H}(\theta_0) \xrightarrow{p} E[H(\theta_0)]$. If (a) holds, then by the continuous mapping theorem we have $\hat{H}(\theta^+) \xrightarrow{p} \hat{H}(\theta_0) \xrightarrow{p} E[H(\theta_0)]$. Condition (a) directly follows from consistency because θ^+ is some intermediate point between θ_0 and $\hat{\theta}$. To show (b), note that:

$$\begin{aligned} sup|\hat{H}(\theta) - E[H(\theta)]| &\leq sup|\hat{H}(\theta) - H(\theta)| + sup|H(\theta) - E[H(\theta)]| && (38) \\ &\leq T_1 + T_2 && (39) \end{aligned}$$

$T_2 \xrightarrow{p} 0$ from Lemma 2.1. Note that the Hessian, by definition, is the second derivative of the quasi-log-likelihood function:

$$H(\theta_0) \equiv \frac{1}{N}\sum_i \left(\frac{Y_i^k}{P_{k,i}}\nabla_{\theta'\theta}P_{k,i} - \frac{Y_i^k}{P_{k,i}^2}\nabla_\theta P_{k,i}\right)$$

To make T_1 uniformly converge to 0, we need $\nabla_\theta^t \hat{P}_i^k$ uniformly converge to its associated estimand for $t = 0, 1, 2$, with the rate at $t = 2$ being the slowest. From Lemma 2.2, for the case of $t = 2$, the bandwidth parameter r must be chosen in a way that $\frac{lnN}{N^{1-7r}} = o(1)$. From the L'Hôpital's rule, $\lim \frac{ln(N)}{N^a} = 0$ for any positive a. Therefore, $r < 1/7$ is sufficient to guarantee the convergence of Hessian: $\hat{H}(\theta^+)^{-1} \xrightarrow{p} H_0 = E[H(\theta_0)]$ $\square$

Lemma 2.6 (Double Convergence)**.** *Let* a_i*,*b_i *be some iid quantity, and* $\hat{a}_i$*,*$\hat{b}_i$ *be their estimator respectively. If* $\frac{1}{N}\sum_i(\hat{a}_i - a_i)^2 = O_p(N^{-\alpha_1})$*,* $\frac{1}{N}\sum_i(\hat{b}_i - b_i)^2 = O_p(N^{-\alpha_2})$*, then* $\frac{1}{N}\sum_i(\hat{a}_i - a_i)(\hat{b}_i - b_i) = O_p(N^{-\alpha_2-\alpha_1})$

Proof. The proof follow directly from the Cauchy-Schwarz:

$$\frac{1}{N}\sum_i(\hat{a}_i - a_i)(\hat{b}_i - b_i) \leq \sqrt{\frac{1}{N}\sum_i(\hat{a}_i - a_i)^2}\sqrt{\frac{1}{N}\sum_i(\hat{b}_i - b_i)^2} \tag{40}$$

$\square$

Lemma 2.7 (Convergence rate for double sums). *Let $h = 0.97\sigma N^{-r}, 1/12 < r < 1/8$ and the recursive differencing estimator $\hat{P}^*_{ki}$ defined in the main text. For $\hat{M}_i$ being (i) $\hat{g}_i$ or (ii) $\widehat{w_{ki}} \equiv \tau_{iv} \frac{\partial \widehat{P^*_{ki}}}{\widehat{P^*_{ki}} \partial \theta}|_{\theta=\theta_0}$ for all k,*

$$\sqrt{N} \frac{1}{N} \sum_i (P_{ki} - \hat{P}^*_{ki})(M_i - \hat{M}_i) = o_p(1)$$

Proof. Apply Lemma 2.6,

$$\sqrt{N} \frac{1}{N} \sum_i (P_{ki} - \hat{P}^*_{ki})(M_i - \hat{M}_i) \leq \sqrt{N} \sqrt{\frac{1}{N} \sum_i (\hat{P}^*_{ki} - P_{ki})^2} \sqrt{\frac{1}{N} \sum_i (\hat{M}_i - M_i)^2} \tag{41}$$

To show (i), replace M_i with g_i. From Lemma 2.2 that

$$\begin{aligned}
\frac{1}{N} \sum_i (\hat{g}_i - g_i)^2 &\leq \frac{1}{N} \sum_i (\hat{g}_i - E[g_i] + E[g_i] - g_i)^2 &(42)\\
&\leq sup_{v,\theta} |E\{\hat{g}(v) - E[g(v)]\}^2| + sup^2_{v,\theta} |E[g(v)] - g(v)| &(43)\\
&= O_p(\frac{1}{Nh^3}) + O_p(h^4) &(44)
\end{aligned}$$

Note from Lemma 2.4 that

$$\frac{1}{N} \sum_i (\hat{P}^*_{ki} - P_{ki})^2 = O_p(\frac{1}{Nh^3}) + O_p(h^8) \tag{45}$$

As such,

$$\begin{aligned}
\sqrt{N} \frac{1}{N} \sum_i (P_{ki} - \hat{P}^*_{ki})(g_i - \hat{g}_i) &\leq \sqrt{N(O_p(\frac{1}{Nh^3}) + O_p(h^8))(O_p(\frac{1}{Nh^3}) + O_p(h^4)))} \\
&= o_p(N^{-1/2}) \quad \text{given that } 1/12 < r < 1/6 &(46)
\end{aligned}$$

To show (ii), from Lemma 10 in Shen and Klein (2019),

$$\frac{1}{N} \sum_i (w - \hat{w})^2 = O(h^8) + O(1/Nh^5) \tag{47}$$

Therefore the double sum will be $o_p(1)$, if $\left\{\begin{matrix}[O_p(h^4)+O_p(1/N^{1/2}h^{3/2})]\times \\ [O_p(h^4)+O_p(1/N^{1/2}h^{5/2})]\end{matrix}\right\} = o_p(N^{-1/2})$. This condition is satisfied with $1/16 < r < 1/8$. □

Lemma 2.8. *With $\widehat{P_{ki}}$ being a standard Nadaraya-Watson estimator defined in D.4 and $D_2^* = \frac{\sqrt{N}}{N}\sum_{i=1}^N[\widehat{P_{ki}^*} - P_{ki}]\frac{\hat{g}_i}{g_i}w_{ki}$ defined in Theorem 2, $D_2^* - D_2^{**} = o_p(1)$ where $D_2^{**} = \frac{\sqrt{N}}{N}\sum_{i=1}^N[\widehat{P_{ki}} - P_{ki}]\frac{\hat{g}_i}{g_i}w_{ki}$.*

Proof. We proceed by first defining two intermediate objects that will simplify the analysis:

$$\begin{aligned}
\widehat{f^0}(v,\theta_0) &\equiv \hat{g}_i\widehat{P_{ki}} = \frac{1}{N}\sum_j Y_j^k K_h(V_j - v) \equiv \frac{1}{N}\sum f_{0j}(v,\theta_0) \\
\widehat{f^1}(v,\theta_0) &\equiv \hat{g}_i\widehat{P_{ki}^*} = \frac{1}{N}\sum_j [Y_j^k - \Delta_j^k(v)]K_h(V_j - v) \equiv \frac{1}{N}\sum_j f_{1j}(v,\theta_0)
\end{aligned}$$

To establish the equivalence result, it is sufficient to show that for each k:

$$\begin{aligned}
D_2^* - D_2^{**} &= \frac{\sqrt{N}}{N}\sum_i [f^0\widehat{(v_i,\theta_0)} - f^1\widehat{(v_i,\theta_0)}]w_i \\
&\leq \sqrt{N}\sup_v |[\hat{f}^0(v,\theta_0) - \hat{f}^1(v,\theta_0)]w_i| = o_p(1) \qquad (48)
\end{aligned}$$

Using a "residual property" of $\nabla_\theta E_i^k|_{\theta=\theta_0}$ provided in Lemma 2.9, it can be shown that $E[f_{0j}(v,\theta_0)w_i] = E[f_{1j}(v,\theta_0)w_i] = 0$. Therefore, with $G_n(v)$ as the empirical CDF and $G(v)$ the true cumulative density of the index V_j at θ_0, we have

$$\begin{aligned}
[\hat{f}^0(v) - \hat{f}^1(v)]w_i &= \hat{f}^0(v)w_i - E[\hat{f}^0(v)w_i] - \hat{f}^1(v)w_i + E[\hat{f}^1(v)w_i] \\
&= \int_{V_j} f_{0j}(v,\theta_0)w_i d[G_n(v) - G(v)] - \int_{V_j} f_{1j}(v,\theta_0)w_i d[G_n(v) - G(v)] \\
&= \int_{V_j} [(f_{0j}(v,\theta_0) - f_{1j}(v,\theta_0))w_i][dG_n(v) - dG(v)] \qquad (49) \\
&= \int_{V_j} [\Delta_j^k(v)K_h(V_j - v)w_i][dG_n(v) - dG(v)] \qquad (50) \\
&= \Delta_j^k(v)K_h(V^* - v)w(v)[G_n(v) - G(v)]|_{V^*\in\partial\Omega} \qquad (51) \\
&- \int_{V_j\in\Omega} [G_n(v) - G(v)]d[\Delta_j^k(v)K_h(V_j - v)w(v)] \qquad (52)
\end{aligned}$$

The last step follows from integrating-by-parts. From A.5, the support of the index V_j is compact. Let $\partial\Omega$ denotes the boundary of this compact support. The first boundary term vanishes because the kernel function $K_h(V^* - v)$ decays very fast when V^* is evaluated at boundary and v is a fixed point. For the second term, one can factor $\sup_v |G_n(v) - G(v)|$[1] outside of the integral. Then, since $\int_{V_j \in \Omega} d[\Delta_j^k(v) K(V_j - v) w(v)]$ is $o_p(1)$, the result claimed in (48) follows. That is, $sup_v |(\hat{f}^0(v, \theta_0) - \hat{f}^1(v, \theta_0)) \tau_i w_i|| = o_p(N^{-1/2})$.

□

Lemma 2.9 (Residual Property)**.** *Under the index assumption:* $Pr[Y_i^k = 1|X_i] = Pr[Y_i^k = 1|V_i(\theta_0)]$*, we have* $E[\nabla_\theta Pr[Y_i^k = 1|V(\theta)|_{\theta=\theta_0}]] = 0$.

Proof. This property is stated and proved in page 401-403 in Klein and Spady (1993), and the authors thank Whitney Newey for mentioning a key idea in a private communication. For what it is worth, I reiterate the proof using the notation employed in this paper.

For notational convenience, let $V(\theta) = [F_1 + F'\theta^F, B_1 + B'\theta^B, MFOI]$ denote the index value at θ and $G(\cdot|v;\theta)$ as the distribution of $x = [F, B, MFOI]$ conditional on $V(\theta) = v$. We write the conditional rating probability P_k as a function of θ,

$$\begin{aligned} P_k(v;\theta) &\equiv Pr[Y^k = 1|V(\theta)] && (53)\\ &= \int H_k(v;\theta_0) dG(x|v;\theta) && (54)\\ \text{where}\quad H_k(v;\theta_0) &\equiv Pr(Y^k = 1|x) = Pr(Y^k = 1|V(\theta_0)) && (55) \end{aligned}$$

With $\delta(x;\theta) \equiv v(x;\theta_0) - v(x;\theta)$, differentiate $P_k(v;\theta)$ with respect to θ when $v = v(t;\theta)$ for fixed t,

$$\nabla_\theta P_k(v(t;\theta);\theta) = \frac{\partial}{\partial\theta} \int H_k(\delta(x;\theta) + v(x;\theta)) G(x|v(t;\theta);\theta) \qquad (56)$$

When evaluated at $\theta = \theta_0$, based on the law of total differentiation,

$$\begin{aligned} \nabla_\theta P_k(v(t;\theta);\theta)|_{\theta=\theta_0} &= D_1 + D_2 && (57)\\ D_1(t;\theta_0) &\equiv \frac{\partial}{\partial\theta}[\int H_k(\delta(x;\theta_0) + v(x;\theta)) G(x|v(t;\theta);\theta)]|_{\theta=\theta_0} \\ D_2(t;\theta_0) &\equiv \frac{\partial}{\partial\theta}[\int H_k(\delta(x;\theta) + v(x;\theta_0)) G(x|v(t;\theta);\theta)]|_{\theta=\theta_0} \end{aligned}$$

[1] This term is $O_p(N^{-1/2})$ according to Nadaraya (1965) and Eddy and Hartigan (1977).

For $D_1(t;\theta_0)$, since $\delta(x,\theta_0)=0$,

$$\begin{aligned} D_1(t;\theta_0) &= \frac{\partial}{\partial\theta}[\int H_k(v(x;\theta))G(x|v(t;\theta);\theta)]|_{\theta=\theta_0} \\ &= \frac{\partial}{\partial\theta}[\int H_k(v(t;\theta))G(x|v(t;\theta);\theta)]|_{\theta=\theta_0} \\ &= \frac{\partial}{\partial\theta}[H_k(v(t;\theta))\int G(x|v(t;\theta);\theta)]|_{\theta=\theta_0} \\ &= \frac{\partial}{\partial\theta}H_k(v(t;\theta)|_{\theta=\theta_0} \end{aligned} \tag{58}$$

For D_2, note that $\delta(x,\theta_0)=0$, so $\frac{\partial\delta}{\partial\theta}=-\frac{\partial v(x,\theta)}{\partial\theta}$. We may differentiate within the integral to obtain

$$\begin{aligned} D_2(t;\theta_0) &= [\int \frac{\partial}{\partial\theta}H_k(\delta(x;\theta)+v(x;\theta_0))G(x|v(t;\theta);\theta)]|_{\theta=\theta_0} \\ &= -E[D_1(x;\theta_0)|v(x;\theta_0)=v(t;\theta_0)] \end{aligned} \tag{59}$$

where the above expectation is taken with respect to x conditioned on $v(x;\theta_0)=v(t;\theta_0)$. From (57-59),

$$E[\nabla_\theta P_k(v(t;\theta);\theta)|v(t;\theta_0)] = E[D_1(t;\theta_0)|v(t;\theta_0)] - E[D_1(t;\theta_0)|v(t;\theta_0)] = 0 \tag{60}$$

□

References

PK Bhattacharya. Estimation of a probability density function and its derivatives. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 373–382, 1967.

William F Eddy and JA Hartigan. Uniform convergence of the empirical distribution function over convex sets. *The Annals of Statistics*, pages 370–374, 1977.

Bruce E Hansen. Lecture notes on nonparametrics. *Lecture notes*, 2009.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

Roger W Klein and Richard H Spady. An efficient semiparametric estimator for binary response models. *Econometrica: Journal of the Econometric Society*, pages 387–421, 1993.

EA Nadaraya. On non-parametric estimates of density functions and regression curves. *Theory of Probability & Its Applications*, 10(1):186–190, 1965.

Whitney K Newey. Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, pages 233–253, 1994.

Ariel Pakes and David Pollard. Simulation and the asymptotics of optimization estimators. *Econometrica: Journal of the Econometric Society*, pages 1027–1057, 1989.

James L Powell, James H Stock, and Thomas M Stoker. Semiparametric estimation of index coefficients. *Econometrica: Journal of the Econometric Society*, pages 1403–1430, 1989.

Robert J Serfling. *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons, 2009.

Chan Shen and Roger Klein. Recursive differencing for estimating semiparametric models. 2019. URL https://economics.rutgers.edu/downloads-hidden-menu/faculty-cv-s/1824-shen-and-klein-2019/file.