Online supplement to: Identification of a Nonseparable Model under Endogeneity using Binary Proxies for Unobserved Heterogeneity

Benjamin Williams *

June 9, 2018

Contents

*George Washington University, Monroe Hall 309, 2115 G Street NW, Washington, DC 20052, bdwilliams@gwu.edu

B Proofs of lemmas in Appendix A

Proof of Lemma A.2 First, if $A \leq z$ then either $B \leq z + x$ or $|A - B| > x$ so

$$\begin{aligned} Pr(A \leq z) &\leq Pr(B \leq z + x) + Pr(|A - B| > x) \\ &\leq Pr(B \leq z' + x + w) + Pr(|A - B| > x). \end{aligned} \tag{B.1}$$

Therefore,

$$\begin{aligned} Pr(A \leq z) - Pr(B \leq z') &\leq Pr(B \leq z' + x + w) - Pr(B \leq z') + y \\ &\leq \bar{f}_B(x + w) + y. \end{aligned} \tag{B.2}$$

Similarly, if $B \leq z - x$ then either $A \leq z$ or $|A - B| > x$, so

$$\begin{aligned} Pr(B \leq z' - x - w) &\leq Pr(B \leq z - x) \\ &\leq Pr(A \leq z) + Pr(|A - B| > x), \end{aligned} \tag{B.3}$$

which implies that

$$\begin{aligned} Pr(A \leq z) - Pr(B \leq z') &\geq -\left(Pr(B \leq z') - Pr(B \leq z' - x - w)\right) - y \\ &\geq -\bar{f}_B(x + w) - y. \end{aligned} \tag{B.4}$$

The result then follows from (B.2) and (B.4). □

Proof of Lemma A.6 First, let $\alpha(x) := \alpha_{\lfloor x \rfloor}$ for any positive real number x. Because $\alpha_k \to 0$ as $k \to \infty$, there exists a sequence of positive numbers $x_N \to 0$ such that $\alpha(A^{1/2}N^{1/2}x_N) \leq x_N$ for all sufficiently large N.[1] Next, to simplify the notation, define $c^*(r) := c_Q^{-1}(c_{\bar{p}}^{-1}(r))$. Because $c^*(r)$ is a continuous, strictly increasing function such that $c^*(0) = 0$, I can define r_N such that $r_N c^*(r_N) = x_N^{1/2}$ for N sufficiently large.[2] Moreover, for this sequence $x_N \to 0$ implies that $r_N \to 0$ since otherwise $r_N c^*(r_N)$ is bounded away from 0.

Define $k_N^r = A^{1/3}N^{1/3}r_N^{2/3}c^*(r_N)^{1/3}x_N^{1/2}$ and $k_N = \lfloor k_N^r \rfloor$. For sufficiently large N, $c^*(r_N) <$

[1] For $N = 1$, there exists x_1 satisfying this inequality is guaranteed since $\alpha(\cdot)$ is a nonincreasing function. Then for each $N > 1$, if $\alpha(A^{1/2}(N-1)^{1/2}x_{N-1}) \leq x_{N-1}$ then $\alpha(A^{1/2}N^{1/2}x_{N-1}) \leq \alpha(A^{1/2}(N-1)^{1/2}x_{N-1}) \leq x_{N-1}$. Therefore, I can define $x_N = \inf\{x \leq x_{N-1} : \alpha(A^{1/2}N^{1/2}x) \leq x\}$. If $\lim_{N\to\infty} x_N \neq 0$ then there is some $\epsilon > 0$ such that $x_N > \epsilon$ infinitely often. By the definition of x_N, this implies that $\alpha(A^{1/2}N^{1/2}\epsilon) > \epsilon$ infinitely often, which is a contradiction since $\lim_{x\to\infty} \alpha(x) = 0$.

[2] Let $f(r) = rc^*(r)$ and consider some $\bar{r} > 0$. Since $0 = f(0) < f(\bar{r})$, and since $x_N \to 0$, for all N sufficiently large, $f(0) \leq \sqrt{x_N} < f(\bar{r})$. By the intermediate value theorem for each such N there must be $0 \leq r_N \leq \bar{r}$ such that $f(r_N) = \sqrt{x_N}$.

1 and $r_N < 1$ so $r_N^{2/3} c^*(r_N)^{1/3} > r_N c^*(r_N)$. Since $\alpha(\cdot)$ is decreasing,

$$\begin{aligned}\alpha_{k_N} = \alpha(k_N^r) &\leq \alpha(A^{1/3} N^{1/3} r_N c^*(r_N) x_N^{1/2}) \\ &= \alpha(A^{1/3} N^{1/3} x_N) \\ &\leq x_N = r_N c^*(r_N) x_N^{1/2}\end{aligned}$$

I have shown that $\frac{\alpha_{k_N}}{r_N c^*(r_N)} \leq x_N^{1/2} \to 0$, as desired. Note also that, since since $r_N = o(1)$ and $x_N = o(1)$, $\frac{k_N^r}{N} = A^{1/3} N^{-2/3} r_N^{2/3} c^*(r_N)^{1/3} x_N^{1/2} \to 0$. Therefore, $k_N = o(N)$. This in turn implies that for any $\kappa > 0$, $\kappa N \geq k_N^r$ for sufficiently large N. Since $\alpha(\cdot)$ is decreasing and $r_N = o(1)$,

$$\begin{aligned}\frac{\alpha(\kappa N)}{c^*(r_N)} &\leq \frac{\alpha(k_N^r)}{c^*(r_N)} \\ &= r_N \frac{\alpha(k_N^r)}{r_N c^*(r_N)} \to 0\end{aligned}$$

Finally, because $log(z) \leq z$,

$$(k_N^r)^2 log\left(\frac{k_N^r}{c^*(r_N) x_N^{3/2}}\right) \leq \left(\frac{(k_N^r)^3}{c^*(r_N) x_N^{3/2}}\right) = ANr_N^2$$

Rearranging this inequality,

$$\frac{k_N^r}{c^*(r_N)} \exp\left(-A\frac{Nr_N^2}{(k_N^r)^2}\right) \leq x_N^{3/2} \to 0$$

The desired result follows since $k_N \leq k_N^r$ implies that

$$\frac{k_N}{c^*(r_N)} \exp\left(-A\frac{Nr_N^2}{k_N^2}\right) \leq \frac{k_N^r}{c^*(r_N)} \exp\left(-A\frac{Nr_N^2}{(k_N^r)^2}\right).$$

□

Before proving Lemmas A.7-A.9 we provide an extension of Lemma A.1 to allow for weak dependence. The result is based on the following version of Azuma's inequality, which is a standard result.

Lemma B.1. *(Azuma's inequality) Suppose $\mathcal{F}_j$ is a filtration and Z_j is a martingale difference with respect to $\mathcal{F}_j$. In addition, suppose $w_{1,J}, \ldots, w_{J,J}$ and $d_{1,J}, \ldots, d_{J,J}$ are constants* *such that* $|w_{j,J}Z_j| \leq d_{j,J}$. *Then, for any* $\epsilon > 0$,

$$Pr\left(\left|\sum_{j=1}^{J} w_{j,J}Z_j\right| > \epsilon\right) \leq 2\exp\left(-\frac{1}{2}\frac{\epsilon^2}{\sum_{j=1}^{J} d_{j,J}}\right)$$

Lemma B.2. *Suppose* $w_{1,J}, \ldots, w_{J,J}$ *are constants with* $w_J = \sum_{j=1}^{J} w_{j,J}$. *If the random variables* $X, \theta, M_1, \ldots, M_J, \ldots$ *satisfy condition (i) of Assumption 2.8 then, for any* $\epsilon > 0$ *and any sequence of integers* $k_J \geq 1$,

$$Pr\left(w_J^{-1}|\sum_{j=1}^{J} w_{j,J}(M_j - E(M_j \mid X, \theta))| > \epsilon\right) \leq 2k_J \exp\left(-\frac{1}{8}\frac{w_J\epsilon^2}{k_J^2}\right) + \frac{2\alpha_{k_J}}{\epsilon}$$

Proof. First, let $\mathcal{F}_j$ denote the sigma algebra generated by $\{X, \theta, M_1, \ldots, M_j\}$ for $j \geq 1$ and the sigma algebra generated by $\{X, \theta\}$ for $j \leq 0$. Then for any $k \geq 1$,

$$\begin{aligned}
&|M_j - E(M_j \mid X, \theta)| \\
&\leq \sum_{s=0}^{k-1} |E(M_j \mid \mathcal{F}_{j-s}) - E(M_j \mid \mathcal{F}_{j-s-1})| + |E(M_j \mid \mathcal{F}_{j-k}) - E(M_j \mid X, \theta)|
\end{aligned}$$

Therefore, taking a sequence of integers $k_J \geq 1$,

$$\begin{aligned}
&Pr(w_J^{-1}|\sum_{j=1}^{J} w_{j,J}(M_j - E(M_j \mid X, \theta))| > \epsilon) \qquad \text{(B.5)} \\
&\leq \sum_{s=0}^{k_J-1} Pr(w_J^{-1}|\sum_{j=1}^{J} w_{j,J}(E(M_j \mid \mathcal{F}_{j-s}) - E(M_j \mid \mathcal{F}_{j-s-1}))| > \epsilon/(2k_J)) \\
&+ Pr(w_J^{-1}|\sum_{j=1}^{J} w_{j,J}(E(M_j \mid \mathcal{F}_{j-k_J}) - E(M_j \mid X, \theta))| > \epsilon/2)
\end{aligned}$$

Clearly $E(M_j \mid \mathcal{F}_{j-s}) - E(M_j \mid \mathcal{F}_{j-s-1})$ is a martingale difference with respect to the filtration $\mathcal{F}_{j-s}$. Therefore, applying Azuma's inequality with $d_{j,J} = w_{j,J}$ (since $0 \leq M_j \leq 1$), the first term can be bounded,

$$\begin{aligned}
&\sum_{s=0}^{k_J-1} Pr(w_J^{-1}|\sum_{j=1}^{J} w_{j,J}(E(M_j \mid \mathcal{F}_{j-s}) - E(M_j \mid \mathcal{F}_{j-s-1}))| > \epsilon/(2k_J)) \\
&\leq 2k_J \exp\left(-\frac{1}{8}\frac{w_J\epsilon^2}{k_J^2}\right) \qquad \text{(B.6)}
\end{aligned}$$

Applying Markov's inequality to the second term of (B.5),

$$
\begin{aligned}
& Pr(w_J^{-1}|\sum_{j=1}^{J} w_{j,J}(E(M_j \mid \mathcal{F}_{j-k_J}) - E(M_j \mid X, \theta))| > \epsilon/2) \\
& \leq \frac{2}{w_J \epsilon} \sum_{j=1}^{J} |w_{j,J}| E\left(|E(M_j \mid \mathcal{F}_{j-k_J}) - E(M_j \mid X, \theta)|\right) \\
& \leq \frac{2}{w_J \epsilon} \alpha_k \sum_{j=1}^{J} w_{j,J} \qquad \text{(B.7)}
\end{aligned}
$$

The desired result follows by combining (B.5)-(B.7). □

Proof of Lemma A.7 Fix $x \in \mathcal{X}$ and $t_0 \in \Theta_0(x)$. I will first show that

$$
\begin{aligned}
& Pr(|\bar{M} - \bar{p}_0(x, t_0)| \leq a r_J/2 \mid X = x) \\
& \geq c_Q^{-1}(c_{\bar{p}}^{-1}(a r_J/2)) - 2\pi_X^{-1} k_J \exp(-\frac{1}{128}\frac{N_J a^2 r_J^2}{k_J^2}) - 8\pi_X^{-1}\frac{\alpha_{k_J}}{a r_J} \qquad \text{(B.8)}
\end{aligned}
$$

where $\pi_X = \inf_{x' \in \mathcal{X}} Pr(X = x')$. First, by iterating expectations and then restricting the range of θ,

$$
\begin{aligned}
& Pr(|\bar{M} - \bar{p}_0(x, t_0)| \leq a r_J/2 \mid X = x) \qquad \text{(B.9)} \\
& = \int Pr(|\bar{M} - \bar{p}_0(x, t_0)| \leq a r_J/2 \mid X = x, \theta = \tau) dF^0_{\theta|X=x}(\tau) \\
& \geq \int_{\tau : |\bar{p}_0(x,\tau) - \bar{p}_0(x,t_0)| \leq a r_J/4} Pr(|\bar{M} - \bar{p}_0(x, t_0)| \leq a r_J/2 \mid X = x, \theta = \tau) dF^0_{\theta|X=x}(\tau) \\
& = \int_{\tau : |\bar{p}_0(x,\tau) - \bar{p}_0(x,t_0)| \leq a r_J/4} \left(1 - Pr(|\bar{M} - \bar{p}_0(x, t_0)| > a r_J/2 \mid X = x, \theta = \tau)\right) dF^0_{\theta|X=x}(\tau) \\
& \geq \int_{\tau : |\bar{p}_0(x,\tau) - \bar{p}_0(x,t_0)| \leq a r_J/4} \left(1 - Pr(|\bar{M} - \bar{p}_0(x, \tau)| > a r_J/4 \mid X = x, \theta = \tau)\right) dF^0_{\theta|X=x}(\tau)
\end{aligned}
$$

where the last inequality follows because $|\bar{M} - \bar{p}_0(x,\tau)| \geq |\bar{M} - \bar{p}_0(x,t_0)| - |\bar{p}_0(x,\tau) - \bar{p}_0(x,t_0)|$.

Next, applying the law of iterated expectations in reverse,

$$\begin{aligned}
&\int_{\tau:|\bar{p}_0(x,\tau)-\bar{p}_0(x,t_0)|\leq ar_J/4} \left(1 - Pr(|\bar{M} - \bar{p}_0(x,\tau)| > ar_J/4 \mid X = x, \theta = \tau)\right) dF^0_{\theta|X=x}(\tau) \\
&= \int_{\tau:|\bar{p}_0(x,\tau)-\bar{p}_0(x,t_0)|\leq ar_J/4} dF^0_{\theta|X=x}(\tau) \qquad \text{(B.10)} \\
&- \int_{\tau:|\bar{p}_0(x,\tau)-\bar{p}_0(x,t_0)|\leq ar_J/4} Pr(|\bar{M} - \bar{p}_0(x,\tau)| > ar_J/4 \mid X = x, \theta = \tau) dF^0_{\theta|X=x}(\tau) \\
&\geq Pr(|\bar{p}_0(x,\theta) - \bar{p}_0(x,t_0)| \leq ar_J/4 \mid X = x) - Pr(|\bar{M} - \bar{p}_0(x,\theta)| > ar_J/4 \mid X = x)
\end{aligned}$$

By Assumptions 2.10 and A.1,

$$\begin{aligned}
&Pr(|\bar{p}_0(x,\theta) - \bar{p}_0(x,t_0)| \leq ar_J/4 \mid X = x) \qquad \text{(B.11)} \\
&= F^0_{\theta|X=x}(\bar{p}_0^{-1}(\bar{p}_0(x,t_0) + ar_J/4; x)) - F^0_{\theta|X=x}(\bar{p}_0^{-1}(\bar{p}_0(x,t_0) - ar_J/4; x)) \\
&\geq c_Q^{-1}(c_{\bar{p}}^{-1}(ar_J/2))
\end{aligned}$$

and applying Lemma B.2,

$$\begin{aligned}
Pr(|\bar{M} - \bar{p}_0(x,\theta)| > ar_J/4 \mid X = x) &\leq \frac{Pr(|\bar{M} - \bar{p}_0(X,\theta)| > ar_J/4, X = x)}{Pr(X = x)} \qquad \text{(B.12)} \\
&\leq 2\pi_X^{-1} k_J \exp(-\frac{1}{128}\frac{N_J a^2 r_J^2}{k_J^2}) + 8\pi_X^{-1}\frac{\alpha_{k_J}}{ar_J}
\end{aligned}$$

Inequality (B.8) follows from (B.9)-(B.12).

On the other hand, consider the model parameterized by $\gamma \in \Gamma_J$. If $|\bar{M} - \bar{p}_0(x,t_0)| \leq ar_J/2$ then for any τ either $|\bar{p}(x,\tau) - \bar{p}_0(x,t_0)| \leq ar_J$ or $|\bar{M} - \bar{p}(x,\tau)| > ar_J/2$. Therefore,

$$\begin{aligned}
&Pr(|\bar{M} - \bar{p}_0(x,t_0)| \leq ar_J/2 \mid X = x) \qquad \text{(B.13)} \\
&= \int Pr(|\bar{M} - \bar{p}_0(x,t_0)| \leq ar_J/2 \mid X = x, \theta = \tau) dF_{\theta|X=x}(\tau) \\
&\leq \int Pr(|\bar{p}(x,\tau) - \bar{p}_0(x,t_0)| \leq ar_J \mid X = x, \theta = \tau) dF_{\theta|X=x}(\tau) \\
&+ \int Pr(|\bar{M} - \bar{p}(x,\tau)| > ar_J/2 \mid X = x, \theta = \tau) dF_{\theta|X=x}(\tau) \\
&= Pr(|\bar{p}(x,\theta) - \bar{p}_0(x,t_0)| \leq ar_J \mid X = x) + Pr(|\bar{M} - \bar{p}(x,\theta)| > ar_J/2 \mid X = x)
\end{aligned}$$

Using the definition of $\bar{M}$ in equation (A.35) and applying Lemma B.2, with $w_{j,J} = \mathbf{1}(j \in$

$\mathcal{J}_m^J(\eta)$) so that $w_J = N_J$, since γ must also satisfy Assumption 2.8,

$$Pr(|\bar{M} - \bar{p}(x,\theta)| > ar_J/2 \mid X = x) \leq 2\pi_X^{-1} k_J \exp(-\frac{1}{32}\frac{N_J a^2 r_J^2}{k_J^2}) + 4\pi_X^{-1}\frac{\alpha_{k_J}}{ar_J} \tag{B.14}$$

To prove the result by contradiction, suppose that $|\bar{p}(x,\tau) - \bar{p}_0(x,t_0)| > ar_J$ for all $\tau \in \Theta(x)$. Then $Pr(|\bar{p}(x,\theta) - \bar{p}_0(x,t_0)| < ar_J \mid X = x) = 0$ and (B.8), (B.13), and (B.14) together imply that

$$\begin{aligned} & c_Q^{-1}(c_{\bar{p}}^{-1}(ar_J/2)) - 2\pi_X^{-1} k_J \exp(-\frac{1}{128}\frac{N_J a^2 r_J^2}{k_J^2}) - 8\pi_X^{-1}\frac{\alpha_{k_J}}{ar_J} \\ & \leq Pr(|\bar{M} - \bar{p}_0(x,t_0)| \leq ar_J/2 \mid X = x) \\ & \leq 2\pi_X^{-1} k_J \exp(-\frac{1}{32}\frac{N_J a^2 r_J^2}{k_J}) + 4\pi_X^{-1}\frac{\alpha_{k_J}}{ar_J} \end{aligned} \tag{B.15}$$

which implies that

$$\begin{aligned} & 1 - 2\pi_X^{-1}\frac{k_J}{c_Q^{-1}(c_{\bar{p}}^{-1}(ar_J/2))}\exp\left(-\frac{1}{128}\frac{N_J a^2 r_J^2}{k_J^2}\right) - 8\pi_X^{-1}\frac{\alpha_{k_J}}{ar_J c_Q^{-1}(c_{\bar{p}}^{-1}(ar_J/2))} \\ & \leq 2\pi_X^{-1}\frac{k_J}{c_Q^{-1}(c_{\bar{p}}^{-1}(ar_J/2))}\exp\left(-\frac{1}{32}\frac{N_J a^2 r_J^2}{k_J}\right) + 4\pi_X^{-1}\frac{\alpha_{k_J}}{ar_J c_Q^{-1}(c_{\bar{p}}^{-1}(ar_J/2))} \end{aligned} \tag{B.16}$$

which implies a contradiction for large enough J since r_J,k_J, and a were chosen so that the left hand side has a limit of 1 and the right hand side a limit of 0. I can conclude that for all sufficiently large J, $\exists t \in \Theta(x)$ such that $|\bar{p}_0(x,t_0) - \bar{p}(x,t)| \leq ar_J$.

$\square$

Proof of Lemma A.8 Consider any $x \in \mathcal{X}$ and $t \in \Theta(x)$, and let m_0 be such that $|m_0 - \bar{p}(x,t)| \leq r_J/2$.

First, $\bar{M}$ is measurable with respect to $\{M_j : j \in \mathcal{J}_Y^J(\eta)\}$ so that $E(Y \mid |\bar{M} - m_0| \leq r_J, X = x) = E(E(Y \mid X, \theta, \{M_j : j \in \mathcal{J}_Y^J(\eta)\}) \mid |\bar{M} - m_0| \leq r_J, X = x)$. Therefore,

$$\begin{aligned} & \left|G(x,t) - E(Y \mid |\bar{M} - m_0| \leq r_J, X = x)\right| \\ & \leq \left|G(x,t) - E(E(Y \mid X,\theta) \mid |\bar{M} - m_0| \leq r_J, X = x)\right| \\ & + \left|E\left(E(Y \mid X, \theta, \{M_j : j \in \mathcal{J}_Y^J(\eta)\}) - E(Y \mid X,\theta) \mid |\bar{M} - m_0| \leq r_J, X = x\right)\right| \end{aligned} \tag{B.17}$$

For any random variable Z and event A, $E(|Z| \mid A) = \frac{E(|Z|) - E(|Z||A^c)(1-Pr(A))}{Pr(A)} \leq \frac{E(|Z|)}{Pr(A)}$.

Therefore,

$$\begin{aligned}
&\left|E\left(E(Y \mid X, \theta, \{M_j : j \in \mathcal{J}_Y^J(\eta)\}) - E(Y \mid X, \theta) \mid |\bar{M} - m_0| \leq r_J, X = x\right)\right| \\
&\leq \frac{E\left(\left|E(Y \mid X, \theta, \{M_j : j \in \mathcal{J}_Y^J(\eta)\}) - E(Y \mid X, \theta)\right|\right)}{Pr(|\bar{M} - m_0| \leq r_J \mid X = x) Pr(X = x)} \\
&\leq \frac{\alpha_{\tilde{k}_J}}{Pr(|\bar{M} - m_0| \leq r_J \mid X = x) Pr(X = x)}
\end{aligned} \tag{B.18}$$

where the second inequality follows from Assumption 2.8.

Next, by Assumption 2.1, and because γ is observationally equivalent to γ_0,

$$\begin{aligned}
&|G(x,t) - E(E(Y \mid X, \theta) \mid |\bar{M} - m_0| \leq r_J, X = x)| \\
&= \left| \int (G(x,t) - G(x,\tau))\, dF_{\theta||\bar{M}-m_0| \leq r_J, X=x}(\tau) \right| \\
&\leq \int_{\tau:|\bar{p}(x,\tau)-\bar{p}(x,t)| \leq 3r_J} |G(x,\tau) - G(x,t)| dF_{\theta||\bar{M}-m_0| \leq r_J, X=x}(\tau) \\
&+ \int_{\tau:|\bar{p}(x,\tau)-\bar{p}(x,t)| > 3r_J} |G(x,\tau) - G(x,t)| dF_{\theta||\bar{M}-m_0| \leq r_J, X=x}(\tau) \\
&\leq c_G(c_{\bar{p}^{-1}}(3r_J)) + BPr(|\bar{p}(x,\theta) - \bar{p}(x,t)| > 3r_J \mid |\bar{M} - m_0| \leq r_J, X = x)
\end{aligned} \tag{B.19}$$

The first term in the final line follows because $|G(x,\tau) - G(x,t)| \leq c_G(|\tau - t|)$ and because Assumption 2.9 implies strict monotonicity of $\bar{p}(x,\cdot)$ so that

$$\begin{aligned}
|\tau - t| &= |\bar{p}^{-1}(\bar{p}(x,\tau); x) - \bar{p}^{-1}(\bar{p}(x,t); x)| \\
&\leq c_{\bar{p}^{-1}}(|\bar{p}(x,\tau) - \bar{p}(x,t)|)
\end{aligned} \tag{B.20}$$

The second term in the final line of (B.19) follows because $G(x,\cdot)$ is uniformly continuous on a compact subset of $\mathbb{R}$ for each x and $|\mathcal{X}|$ is finite and therefore there is some positive constant $B < \infty$ such that $\sup_{x \in \mathcal{X}, t \in \Theta(x)} |G(x,t)| \leq B/2$.

Next, recall that it is assumed that $|m_0 - \bar{p}(x,t)| \leq r_J/2$. Thus, if $|\bar{p}(x,\theta) - \bar{p}(x,t)| > 3r_J$ and $|\bar{M} - m_0| \leq r_J$ then

$$|\bar{M} - \bar{p}(x,\theta)| \geq |\bar{p}(x,\theta) - \bar{p}(x,t)| - |\bar{M} - m_0| - |m_0 - \bar{p}(x,t)| > r_J \tag{B.21}$$

and therefore

$$\begin{aligned}
&Pr(|\bar{p}(x,\theta) - \bar{p}(x,t)| > 3r_J \mid |\bar{M} - m_0| \leq r_J, X = x) \\
&\leq Pr(|\bar{M} - \bar{p}(x,\theta)| > r_J \mid |\bar{M} - m_0| \leq r_J, X = x)
\end{aligned} \tag{B.22}$$

and then

$$\begin{aligned}
&Pr(|\bar{M} - \bar{p}(x,\theta)| > r_J \mid |\bar{M} - m_0| \le r_J, X = x) \\
&= \frac{Pr(|\bar{M} - \bar{p}(X,\theta)| > r_J, |\bar{M} - m_0| \le r_J, X = x)}{Pr(|\bar{M} - m_0| \le r_J, X = x)} \\
&\le \frac{Pr(|\bar{M} - \bar{p}(X,\theta)| > r_J)}{Pr(|\bar{M} - m_0| \le r_J, X = x)} \qquad \text{(B.23)}
\end{aligned}$$

Using the definition of $\bar{M}$ in equation (A.35) and applying Lemma B.2, with $w_{j,J} = \mathbf{1}(j \in \mathcal{J}_m^J(\eta))$ so that $w_J = N_J$, since γ must also satisfy Assumption 2.8,

$$Pr(|\bar{M} - \bar{p}(X,\theta)| > r_J) \le 2k_J \exp(-\frac{1}{8}\frac{N_J r_J^2}{k_J^2}) + 2\frac{\alpha_{k_J}}{r_J} \qquad \text{(B.24)}$$

Combining this with equations (B.17)-(B.23),

$$\begin{aligned}
&\left|G(x,t) - E(Y \mid |\bar{M} - m_0| \le r_J, X = x)\right| \qquad \text{(B.25)} \\
&\le \frac{\alpha_{\tilde{k}_J} + 2Bk_J \exp(-\frac{1}{8}\frac{N_J r_J^2}{k_J^2}) + 2B\frac{\alpha_{k_J}}{r_J}}{Pr(|\bar{M} - m_0| \le r_J, X = x)} + c_G(c_{\bar{p}^{-1}}(3r_J))
\end{aligned}$$

The desired result follows because $Pr(|\bar{M} - m_0| \le r_J, X = x) = Pr(|\bar{M} - m_0| \le r_J \mid X = x)Pr(X = x)$, $Pr(X = x) \ge \pi_X := \inf_{x' \in \mathcal{X}} Pr(X = x')$, and

$$Pr(|\bar{M} - m_0| \le r_J \mid X = x) \ge c_Q^{-1}(c_{\bar{p}}^{-1}(r_J/2)) - 2\pi_X^{-1}k_J \exp(-\frac{1}{128}\frac{N_J r_J^2}{k_J^2}) - 8\pi_X^{-1}\frac{\alpha_{k_J}}{r_J} \qquad \text{(B.26)}$$

The proof is concluded by proving (B.26). Since $|m_0 - \bar{p}(x,t)| \le r_J/2$, $Pr(|\bar{M} - m_0| \le r_J \mid X = x) \ge Pr(|\bar{M} - \bar{p}(x,t)| \le r_J/2 \mid X = x)$. Following the arguments in lines (B.9) and (B.10) of the proof of Lemma A.7,

$$\begin{aligned}
&Pr(|\bar{M} - \bar{p}(x,t)| \le r_J/2 \mid X = x) \qquad \text{(B.27)} \\
&\ge Pr(\bar{p}(x,\theta) - \bar{p}(x,t)| \le r_J/4 \mid X = x) - Pr(|\bar{M} - \bar{p}(x,\theta)| > r_J/4 \mid X = x) \qquad \text{(B.28)}
\end{aligned}$$

By Assumptions 2.10 and A.1,

$$\begin{aligned}
&Pr(|\bar{p}(x,\theta)-\bar{p}(x,t)| \leq r_J/4 \mid X=x) \\
&= F_{\theta|X=x}(\bar{p}^{-1}(\bar{p}(x,t)+r_J/4;x)) - F_{\theta|X=x}(\bar{p}^{-1}(\bar{p}(x,t)-r_J/4;x)) \\
&\geq c_Q^{-1}(c_{\bar{p}}^{-1}(r_J/2))
\end{aligned} \tag{B.29}$$

and applying Lemma B.2,

$$Pr(|\bar{M}-\bar{p}(x,\theta)| > r_J/4 \mid X=x) \leq 2\pi_X^{-1} k_J \exp(-\frac{1}{128}\frac{N_J r_J^2}{k_J^2}) + 8\pi_X^{-1}\frac{\alpha_{k_J}}{r_J} \tag{B.30}$$

This proves inequality (B.26) and completes the proof. □

Proof of Lemma A.9 For any $x' \in \mathcal{X}$ and $m_0' \in [0,1]$, define $T(m_0',x';r_J) := E(M_{j_0} \mid |\bar{M}-m_0'| \leq r_J, X=x')$. I will show below that for any $t_0' \in \Theta_0(x')$, if $|\bar{p}_0(x',t_0')-m_0'| < r_J/2$ then

$$\begin{aligned}
&|T(m_0',x';r_J)-p_{j_0,0}(t_0')| \\
&\leq \delta_J := \frac{\alpha_{\tilde{k}_J} + 2Bk_J\exp(-\frac{1}{8}\frac{N_J r_J^2}{k_J^2}) + 2B\frac{\alpha_{k_J}}{r_J}}{\pi_X c_Q^{-1}(c_{\bar{p}}^{-1}(r_J/2)) - 2k_J\exp(-\frac{1}{128}\frac{N_J r_J^2}{k_J^2}) - 8\frac{\alpha_{k_J}}{r_J}} + c_{p_{j_0}}(c_{\bar{p}^{-1}}(3r_J))
\end{aligned} \tag{B.31}$$

where $\pi_X := \inf x' \in \mathcal{X} Pr(X=x)$. This in turn implies that

$$\begin{aligned}
Pr(|T(\bar{M},X;r_J)-p_{j_0,0}(\theta)| > \delta_J) &\leq Pr(|\bar{p}_0(X,\theta)-\bar{M}| > r_J/2) \\
&\leq \rho_J := 2k_J\exp(-\frac{1}{32}\frac{N_J r_J^2}{k_J^2}) + 4\frac{\alpha_{k_J}}{r_J}
\end{aligned} \tag{B.32}$$

where the second line follows from Lemma B.2.

Since γ is observationally equivalent to γ_0, the exact same argument shows that for any $t' \in \Theta(x')$, if $|\bar{p}(x',t')-m_0'| \leq r_J/2$ then

$$|T(m_0',x';r_J)-p_{j_0}(t')| \leq \delta_J \tag{B.33}$$

and hence

$$Pr(|T(\bar{M},X;r_J)-p_{j_0}(\theta)| \geq \delta_J) \leq Pr(|\bar{p}(X,\theta)-\bar{M}| \geq r_J/2) \leq \rho_J \tag{B.34}$$

It can also be concluded that for $m_0 = \bar{p}_0(x, t_0)$, since, by assumption, $|m_0 - \bar{p}(x,t)| \leq r_J/2$,

$$\begin{aligned} &|p_{j_0,0}(t_0) - p_{j_0}(t)| \\ &\leq |T(m_0, x; r_J) - p_{j_0,0}(t_0)| + |T(m_0, x; r_J) - p_{j_0}(t)| \\ &\leq 2\delta_J \end{aligned} \tag{B.35}$$

Then (B.32) implies that

$$\begin{aligned} &Pr(T(\bar{M}, X; r_J) \leq p_{j_0,0}(t_0)) \\ &\leq Pr(p_{j_0,0}(\theta) \leq p_{j_0,0}(t_0) + \delta_J) + Pr(|T(\bar{M}, X; r_J) - p_{j_0,0}(\theta)| > \delta_J) \\ &\leq p_{j_0,0}^{-1}(p_{j_0,0}(t_0) + \delta_J) + \rho_J \end{aligned} \tag{B.36}$$

since $Pr(p_{j_0,0}(\theta) \leq p_{j_0,0}(t_0) + \delta_J) = Pr\left(\theta \leq p_{j_0,0}^{-1}\left(p_{j_0,0}(t_0) + \delta_J\right)\right)$ and, by Assumption 2.3, this equals $p_{j_0,0}^{-1}\left(p_{j_0,0}(t_0) + \delta_J\right)$. Next, (B.32) similarly implies that

$$\begin{aligned} &Pr(T(\bar{M}, X; r_J) \leq p_{j_0,0}(t_0)) \\ &\geq Pr(p_{j_0,0}(\theta) \leq p_{j_0,0}(t_0) - \delta_J) - Pr(|T(\bar{M}, X; r_J) - p_{j_0,0}(\theta)| > \delta_J) \\ &\geq p_{j_0,0}^{-1}(p_{j_0,0}(t_0) - \delta_J) - \rho_J \end{aligned} \tag{B.37}$$

Then $p_{j_0,0}(t_0) - \delta_J \leq p_{j_0,0}(t_0) \leq p_{j_0,0}(t_0) + \delta_J$ implies that

$$p_{j_0,0}^{-1}(p_{j_0,0}(t_0) - \delta_J) \leq t_0 \leq p_{j_0,0}^{-1}(p_{j_0,0}(t_0) + \delta_J). \tag{B.38}$$

So (B.36) and (B.37) together imply that

$$|Pr(T(\bar{M}, X; r_J) \leq p_{j_0,0}(t_0)) - t_0| \leq c_{p_{j_0}^{-1}}(2\delta_J) + 2\rho_J \tag{B.39}$$

Similarly, (B.34) and (B.35) imply that

$$\begin{aligned} &Pr(T(\bar{M}, X; r_J) \leq p_{j_0,0}(t_0)) \\ &\leq Pr(p_{j_0}(\theta) \leq p_{j_0,0}(t_0) + \delta_J) + Pr(|T(\bar{M}, X; r_J) - p_{j_0}(\theta)| > \delta_J) \\ &\leq Pr(p_{j_0}(\theta) \leq p_{j_0}(t) + 3\delta_J) + Pr(|T(\bar{M}, X; r_J) - p_{j_0}(\theta)| > \delta_J) \\ &\leq p_{j_0}^{-1}(p_{j_0}(t) + 3\delta_J) + \rho_J, \end{aligned} \tag{B.40}$$

where the second inequality follows because $p_{j_0,0}(t_0) \leq p_{j_0}(t) + 2\delta_J$ by (B.35), and that and that

$$\begin{aligned}
&Pr(T(\bar{M}, X; r_J) \leq p_{j_0,0}(t_0)) \\
&\geq Pr(p_{j_0}(\theta) \leq p_{j_0,0}(t_0) - \delta_J) - Pr(|T(\bar{M}, X; r_J) - p_{j_0}(\theta)| > \delta_J) \qquad \text{(B.41)}\\
&\geq Pr(p_{j_0}(\theta) \leq p_{j_0}(t) - 3\delta_J) - Pr(|T(\bar{M}, X; r_J) - p_{j_0}(\theta)| > \delta_J) \\
&\geq p_{j_0}^{-1}(p_{j_0}(t) - 3\delta_J) - \rho_J
\end{aligned}$$

(B.40) and (B.41) together imply that

$$|Pr(T(\bar{M}, X; r_J) \leq p_{j_0,0}(t_0)) - t| \leq c_{p_{j_0}^{-1}}(6\delta_J) + 2\rho_J \qquad \text{(B.42)}$$

Then (B.39) and (B.42) imply that $|t - t_0| \leq 2c_{p_{j_0}^{-1}}(6\delta_J) + 4\rho_J$, the desired result.

It remains to show that (B.31) holds for any $x' \in \mathcal{X}$, any $m_0' \in [0, 1]$ and any $t_0' \in \Theta_0(x')$ for which $|\bar{p}_0(x', t_0') - m_0'| < r_J/2$. The proof of this is almost identical to the proof of Lemma A.8 so I will provide only a sketch.

First, $\bar{M}$ is measurable with respect to $\{M_j : |j - j_0| > \tilde{k}_J\}$ so that $E(M_{j_0} \mid |\bar{M} - m_0'| \leq r_J, X = x') = E(E(M_{j_0} \mid X, \theta, \{M_j : |j - j_0| > \tilde{k}_J\}) \mid |\bar{M} - m_0'| \leq r_J, X = x')$. Therefore,

$$\begin{aligned}
&\left|p_{j_0,0}(t_0') - E(M_{j_0} \mid |\bar{M} - m_0'| \leq r_J, X = x')\right| \\
&= \left|p_{j_0,0}(t_0') - E(E(M_{j_0} \mid X, \theta) \mid |\bar{M} - m_0'| \leq r_J, X = x')\right| \qquad \text{(B.43)}\\
&+ \left|E\left(E(M_{j_0} \mid X, \theta, \{M_j : |j - j_0| > \tilde{k}_J\}) - E(M_{j_0} \mid X, \theta) \mid |\bar{M} - m_0'| \leq r_J, X = x'\right)\right|
\end{aligned}$$

Next, under Assumption 2.8,

$$\begin{aligned}
&\left|E\left(E(M_{j_0} \mid X = x', \theta, \{M_j : j \in \mathcal{J}_{M_{j_0}}^J\}) - E(M_{j_0} \mid X = x', \theta) \mid |\bar{M} - m_0'| \leq r_J, X = x'\right)\right| \\
&\leq \frac{E\left(\left|E(M_{j_0} \mid X = x', \theta, \{M_j : j \in \mathcal{J}_{M_{j_0}}^J\})) - E(M_{j_0} \mid X = x', \theta)\right|\right)}{Pr(|\bar{M} - m_0'| \leq r_J, X = x')} \qquad \text{(B.44)}\\
&\leq \frac{\alpha_{\tilde{k}_J}}{\pi_X c_Q^{-1}(c_{\bar{p}}^{-1}(r_J/2)) - 2k_J \exp(-\frac{1}{128}\frac{N_J r_J^2}{k_J^2}) - 8\frac{\alpha_{k_J}}{r_J}}
\end{aligned}$$

and, by Assumption 2.4, $E(M_{j_0} \mid X = x', \theta) = p_{j_0,0}(\theta)$ so that

$$
\begin{aligned}
&|p_{j_0,0}(t_0') - E(E(M_{j_0} \mid X = x', \theta) \mid |\bar{M} - m_0'| \leq r_J, X = x')| \\
&\leq \left| \int (p_{j_0,0}(t_0') - p_{j_0,0}(\tau))\, dF^0_{\theta \mid |\bar{M} - m_0'| \leq r_J, X = x'}(\tau) \right| \\
&\leq c_{p_{j_0}}(c_{\bar{p}^{-1}}(3r_J)) + BPr(|\bar{p}_0(x', \theta) - \bar{p}_0(x', t_0')| > 3r_J \mid |\bar{M} - m_0'| \leq r_J, X = x') \qquad \text{(B.45)} \\
&\leq c_{p_{j_0}}(c_{\bar{p}^{-1}}(3r_J)) + BPr(|\bar{M} - \bar{p}_0(x', \theta)| > r_J \mid |\bar{M} - m_0'| \leq r_J, X = x') \\
&\leq c_{p_{j_0}}(c_{\bar{p}^{-1}}(3r_J)) + \frac{\alpha_{\tilde{k}_J} + 2Bk_J \exp(-\frac{1}{8b} \frac{N_J r_J^2}{k_J}) + 2B \frac{\alpha_{k_J}}{r_J}}{\pi_X c_Q^{-1}(c_{\bar{p}}^{-1}(r_J/2)) - 2k_J \exp(-\frac{1}{128} \frac{N_J r_J^2}{k_J^2}) - 8 \frac{\alpha_{k_J}}{r_J}}
\end{aligned}
$$

where the third inequality follows because $|\bar{p}_0(x', t_0') - m_0'| < r_J/2$.

□

C Asymptotic theory for estimators

This section presents the proof of Theorem 3.2.

C.1 Additional assumptions

First, the regularity conditions omitted in the text are stated. The first set of regularity conditions strengthen the conditions of Assumption 3.3.

Take η and $\mathcal{J}_m^J$ as given by Assumption 3.2. I now assume that $N_J := |\mathcal{J}_m^J| = c_0 J$ for some constant $0 < c_0 < 1$. Moreover, $an^r \leq J \leq bn^r$ for some $a, b, r > 0$. Let $\bar{M} = N_J^{-1} \sum_{j \in \mathcal{J}_m^J} M_j$ and $\bar{p} = N_J^{-1} \sum_{j \in \mathcal{J}_m^J} p_j$.

Assumption C.1.

The functions p_{j_0} and $\bar{p}(x, \cdot)$ are differentiable and there are positive constants, m and M, such that

i. $m < inf_{t \in [0,1]} p_{j_0}'(t)$ *and* $\sup_{t \in [0,1]} p_{j_0}(t) < M$

ii. $m < inf_{t \in [0,1]} \bar{p}'(x, t)$ *and* $\sup_{t \in [0,1]} \bar{p}'(x, t) < M$

Assumption C.2. *The distribution of $\Theta \mid X = x$ admits a density with respect to Lebesgue measure on $\Theta_x = [\underline{\theta}_x, \bar{\theta}_x]$, denoted $f_{\theta|X}$ that satisfies*

$$
\underline{f}_{\theta|X} \leq f_{\theta|X}(t \mid x) \leq \bar{f}_{\theta|X}
$$

for some constants $\underline{f}_{\theta|X} > 0$ and $\bar{f}_{\theta|X} < \infty$.

Assumption C.3. $|Y| < \bar{Y}$ *and for each* x, $G(x,t)$ *is locally Lipschitz continuous at* t.

Assumption C.4. $\alpha_k \leq \alpha_0 \exp(-\alpha_1 t)$

Assumption C.5.

i. $\underline{\kappa}\mathbf{1}(|u| \leq 1/2) \leq K(u) \leq \bar{\kappa}\mathbf{1}(|u| \leq 1)$

ii. $\underline{\lambda}\mathbf{1}(|u| \leq 1/2) \leq L(u) \leq \bar{\lambda}\mathbf{1}(|u| \leq 1)$

In addition, to derive the asymptotic results it is assumed that $\{X_i, \theta_i, U_i, M_i\}_{i=1}^n$ is an i.i.d. sample and I impose the following conditions on the convergence rate of the bandwidths.

Assumption C.6.

i. $h_{1n}, h_{2n} \to 0$ *and* $nh_{1n}, nh_{2n} \to \infty$

ii. $\sqrt{log(J)J^{-1}} = o(h_{1n})$

iii. $h_{1n} + \sqrt{log(n)n^{-1}h_{1n}^{-1}} = o(h_{2n})$

The first condition is a standard assumption required to ensure that both the first stage estimator and the infeasible version of the second stage estimator are consistent. The second and third conditions ensure that the error in estimating the covariates in each stage ($\bar{M}_i - \bar{p}(X_i, \theta_i)$ in the first stage and $\hat{\theta}_i - \theta_i$ in the second stage) is sufficiently small relative to the bandwidth so that the bias is not affected. Indeed, it is shown in Lemma C.1 that $\max_{1 \leq i \leq n} |\hat{\theta}_i - \theta_i| = O_p(h_{1n} + \sqrt{log(n)n^{-1}h_{1n}^{-1}})$.

C.2 Proof of Theorem 3.2

Proof of Theorem 3.2. Let $\psi_n = 1/\sqrt{nh_{2n}} + h_{2n}$ and suppose n is large enough that $h_{2n}/2$ is less than the radius of the neighborhood around t assumed by Assumptions C.3. First, for any $\varepsilon^* > 0$,

$$Pr(|\hat{G}(x,t) - G(x,t)| > \varepsilon^* \psi_n) \leq Pr(|\hat{G}(x,t) - G(x,t)| > \varepsilon^* \psi_n, \max_{1 \leq i \leq n} |\hat{\theta}_i - \theta_i| < h_{2n}/4)$$
$$+ Pr(\max_{1 \leq i \leq n} |\hat{\theta}_i - \theta_i| \geq h_{2n}/4)$$

where by Lemma C.1 and Assumption C.6 the second term converges to 0 and thus it remains to bound the first term.

Let $u_i = Y_i - G(X_i, \theta_i)$. Then if $\max_{1\le i\le n} |\hat{\theta}_i - \theta_i| < h_{2n}/4$,

$$
\begin{aligned}
&|\hat{G}(x,t) - G(x,t)| \\
&\le \frac{\frac{1}{n}\sum_{i=1}^n L_{h_{2n}}(\hat{\theta}_i - t, X_i - x)|G(X_i,\theta_i) - G(x,t)|}{\frac{1}{n}\sum_{i=1}^n L_{h_{2n}}(\hat{\theta}_i - t, X_i - x)} + \frac{\left|\frac{1}{n}\sum_{i=1}^n L_{h_{2n}}(\hat{\theta}_i - t, X_i - x)u_i\right|}{\frac{1}{n}\sum_{i=1}^n L_{h_{2n}}(\hat{\theta}_i - t, X_i - x)} \\
&\le \max_{1\le i\le n} \mathbf{1}(|\hat{\theta}_i - t| \le h_{2n})|G(x,\theta_i) - G(x,t)| + \frac{\left|\frac{1}{nh_{2n}}\sum_{i=1}^n L(h_{2n}^{-1}(\hat{\theta}_i - t))\mathbf{1}(X_i = x)u_i\right|}{\frac{1}{nh_{2n}}\sum_{i=1}^n L(h_{2n}^{-1}(\hat{\theta}_i - t))\mathbf{1}(X_i = x)} \\
&\le \Delta(x,t)(h_{2n} + \max_{1\le i\le n}|\hat{\theta}_i - \theta_i|) + \frac{\left|\frac{1}{nh_{2n}}\sum_{i=1}^n L(h_{2n}^{-1}(\hat{\theta}_i - t))\mathbf{1}(X_i = x)u_i\right|}{\frac{1}{nh_{2n}}\sum_{i=1}^n L(h_{2n}^{-1}(\hat{\theta}_i - t))\mathbf{1}(X_i = x)} \\
&\le \frac{5}{4}\Delta(x,t)h_{2n} + \frac{\left|\frac{1}{nh_{2n}}\sum_{i=1}^n L(h_{2n}^{-1}(\hat{\theta}_i - t))\mathbf{1}(X_i = x)u_i\right|}{\frac{1}{nh_{2n}}\sum_{i=1}^n L(h_{2n}^{-1}(\hat{\theta}_i - t))\mathbf{1}(X_i = x)}
\end{aligned}
$$

where the second inequality follows from Assumption C.5 and the third follows from Assumption C.3 for sufficiently large n since $|\theta_i - t| \le |\hat{\theta}_i - \theta_i| + |\hat{\theta}_i - t| < \frac{5}{4}h_{2n}$. Then it remains to show that $\hat{D}(x,t)^{-1}\hat{N}(x,t) = O_p(1/\sqrt{nh_{2n}})$, or more precisely, that

$$
Pr(\hat{D}(x,t)^{-1}\hat{N}(x,t) > \varepsilon^*/\sqrt{nh_{2n}}, \max_{1\le i\le n}|\hat{\theta}_i - \theta_i| < h_{2n}/4)
$$

can be made arbitrarily small by choosing ε^* large enough, where $\hat{N}(x,t) := \frac{1}{nh_{2n}}\sum_{i=1}^n L(h_{2n}^{-1}(\hat{\theta}_i - t))\mathbf{1}(X_i = x)u_i$ and $\hat{D}(x,t) := \frac{1}{nh_{2n}}\sum_{i=1}^n L(h_{2n}^{-1}(\hat{\theta}_i - t))\mathbf{1}(X_i = x)$.

First, let $\hat{N}^*(x,t) = E(\hat{N}(x,t) \mid \{\theta_i, X_i, \{M_{i,j} : j \in \mathcal{J}_Y^J(\eta J)\}\})$. Then by Assumption 3.1, $E(|\hat{N}^*(x,t)|) \le h_{2n}^{-1}\bar{\lambda}\alpha_{\eta J}$. So by Markov's inequality,

$$
\begin{aligned}
&Pr(|\hat{N}^*(x,t)| > \varepsilon/2, \max_{1\le i\le n}|\hat{\theta}_i - \theta_i| < h_{2n}/4) \\
&\le Pr(|\hat{N}^*(x,t)| > \varepsilon/2) \\
&\le \frac{2\bar{\lambda}\alpha_{\eta J}}{\varepsilon h_{2n}}
\end{aligned}
$$

Next, $E(\hat{N}(x,t) - \hat{N}^*(x,t)) = 0$ and, while $\max_{1\le i\le n}|\hat{\theta}_i - \theta_i| < h_{2n}/4$, $Var(\hat{N}(x,t) - \hat{N}^*(x,t)) \le 4\bar{\lambda}^2\bar{Y}^2 Pr(|\theta_i - t| < (5/4)h_{2n})/(nh_{2n}^2)$.

Therefore, using Chebyschev's inequality, for any $\varepsilon > 0$,

$$
\begin{aligned}
&Pr(\hat{N}(x,t) > \varepsilon, \max_{1\le i\le n} |\hat{\theta}_i - \theta_i| < h_{2n}/4) \\
&\le Pr(|\hat{N}(x,t) - \hat{N}^*(x,t)| > \varepsilon/2, \max_{1\le i\le n} |\hat{\theta}_i - \theta_i| < h_{2n}/4) \\
&+ Pr(|\hat{N}^*(x,t)| > \varepsilon/2, \max_{1\le i\le n} |\hat{\theta}_i - \theta_i| < h_{2n}/4) \\
&\le \frac{25\bar{\lambda}^2\bar{Y}^2}{\varepsilon^2 n h_{2n}} + \frac{2\bar{\lambda}\alpha_{\eta J}}{\varepsilon h_{2n}}
\end{aligned}
$$

Next, if $\max_{1\le i\le n} |\hat{\theta}_i - \theta_i| < h_{2n}/4$ then

$$
\begin{aligned}
\hat{D}(x,t) &\ge \underline{\lambda}\frac{1}{nh_{2n}}\sum_{i=1}^{n}\mathbf{1}(|\hat{\theta}_i - t| < h_{2n}/2)\mathbf{1}(X_i = x) \\
&\ge \underline{\lambda}\frac{1}{nh_{2n}}\sum_{i=1}^{n}\mathbf{1}(|\theta_i - t| < h_{2n}/4)\mathbf{1}(X_i = x) := D^*(x,t)
\end{aligned}
$$

where the first inequality is due to Assumption C.5 and the second line is due to the fact that $|\hat{\theta}_i - \theta_i| < h_{2n}/4$ and $|\theta_i - t| < h_{2n}/4$ together imply that $|\hat{\theta}_i - t| < h_{2n}/2$, by the triangle inequality. Then,

$$
\begin{aligned}
\frac{1}{4}h_{2n}\underline{f}_{\theta|X}Pr(X_i = x) &\le E(\mathbf{1}(|\theta_i - t| < h_{2n}/4)\mathbf{1}(X_i = x) \\
&\le \frac{1}{2}h_{2n}\bar{f}_{\theta|X}Pr(X_i = x)
\end{aligned}
$$

by Assumption C.2. Therefore,

$$
\begin{aligned}
&E(D^*(x,t)) \ge \underline{\lambda}h_{2n}^{-1}Pr(|\theta_i - t| < h_n/4 \mid X_i = x)Pr(X_i = x) \\
&\ge \frac{1}{4}\underline{\lambda}\underline{f}_{\theta|X}Pr(X_i = x)
\end{aligned}
$$

and

$$
\begin{aligned}
&Var(D^*(x,t)) \le \underline{\lambda}(nh_{2n}^2)^{-1}Pr(|\theta_i - t| < h_n/4 \mid X_i = x)Pr(X_i = x) \\
&\le \underline{\lambda}(nh_{2n})^{-1}\frac{1}{2}\bar{f}_{\theta|X}Pr(X_i = x)
\end{aligned}
$$

So for $d = \frac{1}{8}\underline{\lambda}\underline{f}_{\theta|X}Pr(X_i = x)$,

$$
\begin{aligned}
& Pr(\hat{D}(x,t) < d, \max_{1\leq i\leq n} |\hat{\theta}_i - \theta_i| < r_t) \\
& \leq Pr(D^*(x,t) < d) \\
& \leq Pr(|D^*(x,t) - E(D^*(x,t))| > d) \\
& \leq d^{-2}\underline{\lambda}(nh_{2n})^{-1}\frac{1}{2}\bar{f}_{\theta|X}Pr(X_i = x) \to 0
\end{aligned}
$$

where the last inequality follows from Chebyschev's inequality and the convergence follows from Assumption C.6.

Therefore,

$$
\begin{aligned}
& Pr(\hat{D}(x,t)^{-1}\hat{N}(x,t) > \varepsilon^*/\sqrt{nh_{2n}}, \max_{1\leq i\leq n} |\hat{\theta}_i - \theta_i| < h_{2n}/4) \\
& \leq Pr(|\hat{N}(x,t)| > d\varepsilon^*/\sqrt{nh_{2n}}, \max_{1\leq i\leq n} |\hat{\theta}_i - \theta_i| < h_{2n}/4) + Pr(\hat{D}(x,t) < d, \max_{1\leq i\leq n} |\hat{\theta}_i - \theta_i| < h_{2n}/4) \\
& \leq \frac{25\bar{\lambda}^2\bar{Y}^2}{\varepsilon^{*2}d^2} + \frac{2\bar{\lambda}\alpha_{\eta J}n^{1/2}}{d\varepsilon^* h_{2n}^{1/2}} + o(1)
\end{aligned}
$$

which implies the desired result, since the second term is $o(1)$ by Assumption C.4. □

C.3 Lemmas

Here I state and prove the lemmas used in the proof of Theorem 3.2 above.

Lemma C.1. $\max_{1\leq i\leq n} |\hat{\theta}_i - \theta_i| = O_p(r_n)$ *where*

$$
r_n = h_{1n} + \sqrt{\log(n)n^{-1}h_{1n}^{-1}}
$$

Proof. Define $\rho_n = \rho_0 r_n$ for some $\rho_0 > 0$. I next show that for ρ_0 sufficiently large $Pr(\max_{1\leq i\leq n} |\hat{\theta}_i - \theta_i| > \rho_n) \to 0$.

First,

$$
|\hat{\theta}_i - \theta_i| = |\hat{F}_{\hat{q}(X,\bar{M})}(\hat{q}(X,\bar{M})) - \theta_i| \leq \sum_{r=1}^{3} |\hat{T}_{ri}|
$$

where

$$\hat{T}_{1i} := \hat{F}_{\hat{q}(X,\bar{M})}(\hat{q}(X,\bar{M})) - F_{\hat{q}(X,\bar{M})}(\hat{q}(X,\bar{M}))$$
$$\hat{T}_{2i} := F_{\hat{q}(X,\bar{M})}(\hat{q}(X,\bar{M})) - F_{p_{j_0}(\theta)}(\hat{q}(X_i,\bar{M}_i))$$
$$\hat{T}_{3i} := F_{p_{j_0}(\theta)}(\hat{q}(X_i,\bar{M}_i)) - \theta_i$$

First, for any $\delta > 0$,

$$\begin{aligned}|\hat{T}_{2i}| &\leq \sup_{x\in[0,1]} |F_{\hat{q}(X,\bar{M})}(x) - F_{p_{j_0}(\theta)}(x)| \\ &\leq Pr(|\hat{q}(X,\bar{M}) - p_{j_0}(\theta)| > \delta) + \sup_{x\in[0,1]} \{|F_{p_{j_0}(\theta)}(x+\delta) - F_{p_{j_0}(\theta)}(x)| \\ &+ |F_{p_{j_0}(\theta)}(x) - F_{p_{j_0}(\theta)}(x-\delta)|\} \\ &\leq Pr(|\hat{q}(X,\bar{M}) - p_{j_0}(\theta)| > \delta) + \frac{2\delta}{m}\end{aligned}$$

where the third inequality follows from Assumption C.2. Take $\delta = \varepsilon^* m \rho_n$. Then by Lemma C.2, if ε^* is sufficiently large,

$$|\hat{T}_{2i}| \leq c_0 n^{-c_1} + 2\varepsilon^* \rho_n$$

for any $c_1 > 0$. So, for n large enough, $Pr(|\hat{T}_{2i}| > \rho_n/3) = 0$.

Second, by Assumption C.1, $|\hat{T}_{3i}| = |F_{p_{j_0}(\theta)}(\hat{q}(X_i,\bar{M}_i)) - F_{p_{j_0}(\theta)}(p_{j_0}(\theta_i))| \leq \frac{1}{m}|\hat{q}(X_i,\bar{M}_i) - p_{j_0}(\theta_i)|$. So, for n large enough,

$$Pr(|\hat{T}_{3i}| > \rho_n/3) \leq Pr(|\hat{q}(X_i,\bar{M}_i) - p_{j_0}(\theta_i)| > m\rho_0 r_n/3)$$

By Lemma C.2, ρ_0 can be chosen large enough so that this is bounded by an arbitrarily large power of n^{-1} for sufficiently large n.

Third, $|\hat{T}_{1i}| \leq \sup_{x\in[0,1]} |\hat{F}_{\hat{q}(X,\bar{M})}(x) - F_{\hat{q}(X,\bar{M})}(x)|$. Since $\hat{F}_{\hat{q}(X,\bar{M})}$ is the empirical distribution function of the random variable $\hat{q}(X,\bar{M})$, by the Dvoretsky-Kiefer-Wolfowitz inequality (see, e.g., p. 268 in Van der Vaart, 2000),

$$Pr(|\hat{T}_{1i}| > \rho_n/3) \leq 2\exp(-\frac{2}{9}\rho_n^2 n) \leq 2n^{-\frac{2}{9}\rho_0^2}$$

Putting these three results together, if ρ_0 is large enough, for all sufficiently large n

$$Pr(\max_{1\le i\le n}|\hat{\theta}_i-\theta_i|>\rho_n)\le n\sum_{r=1}^{3}Pr(|\hat{T}_{ri}|>\rho_n/3)\to 0$$

$\square$

Lemma C.2. *If $\delta_n=O(h_{1n}+\sqrt{\log(n)n^{-1}h_{1n}^{-1}})$ then for any $c_1>0$ there exist $\varepsilon^*>0$, $c_0>0$ and such that for sufficiently large n,*

$$Pr(|\hat{q}(X,\bar{M})-p_{j_0}(\theta)|>\varepsilon^*\delta)\le c_0n^{-c_1}$$

Proof. Recall that $q(x,m)=p_{j_0}(\bar{p}^{-1}(m;x))$.

First, let $\eta_{j_0,i}:=M_{j_0,i}-p_{j_0}(\theta_i)$. Because $p_{j_0}(\theta_i)=q(x,\bar{p}(x,\theta_i))$, $p_{j_0}(\theta_i)-q(x,m)=(q(x,\bar{p}(x,\theta_i))-q(x,\bar{M}_i))+(q(x,\bar{M}_i)-q(x,m))$. So

$$\begin{aligned}
&|\hat{q}(x,m)-q(x,m)|\\
&\le\frac{\frac{1}{nh_{1n}}\sum_{i=1}^n K\left(h_{1n}^{-1}(\bar{M}_i-m)\right)\mathbf{1}(X_i=x)|p_{j_0}(\theta_i)-q(x,m)|}{\frac{1}{nh_{1n}}\sum_{i=1}^n K\left(h_{1n}^{-1}(\bar{M}_i-m)\right)\mathbf{1}(X_i=x)}\\
&+\frac{\left|\frac{1}{nh_{1n}}\sum_{i=1}^n K\left(h_{1n}^{-1}(\bar{M}_i-m)\right)\mathbf{1}(X_i=x)\eta_{j_0,i}\right|}{\frac{1}{nh_{1n}}\sum_{i=1}^n K\left(h_{1n}^{-1}(\bar{M}_i-m)\right)\mathbf{1}(X_i=x)}\\
&\le\max_{1\le i\le n}\mathbf{1}(X_i=x)|q(x,\bar{M}_i)-q(x,\bar{p}(x,\theta_i))|\\
&+\max_{1\le i\le n}\mathbf{1}(|\bar{M}_i-m|<h_{1n})|q(x,\bar{M}_i)-q(x,m)|\\
&+\frac{\left|\frac{1}{nh_{1n}}\sum_{i=1}^n K\left(h_{1n}^{-1}(\bar{M}_i-m)\right)\mathbf{1}(X_i=x)\eta_{j_0,i}\right|}{\frac{1}{nh_{1n}}\sum_{i=1}^n K\left(h_{1n}^{-1}(\bar{M}_i-m)\right)\mathbf{1}(X_i=x)}\\
&\le\frac{M}{m}\max_{1\le i\le n}\mathbf{1}(X_i=x)|\bar{M}_i-\bar{p}(x,\theta_i)|+\frac{M}{m}h_{1n}+\frac{|\hat{N}(x,m)|}{\hat{D}(x,m)}
\end{aligned}$$

where the second inequality follows from Assumption C.5 and the third follows from Assumption C.1 and

$$\hat{N}(x,m)=\frac{1}{nh_{1n}}\sum_{i=1}^n K\left(h_{1n}^{-1}(\bar{M}_i-m)\right)\mathbf{1}(X_i=x)\eta_{j_0,i}$$

$$\hat{D}(x,m)=\frac{1}{nh_{1n}}\sum_{i=1}^n K\left(h_{1n}^{-1}(\bar{M}_i-m)\right)\mathbf{1}(X_i=x)$$

Therefore, taking ε^* large enough that $\frac{M}{m}h_{1n} \leq \varepsilon^*\delta_n/6$,

$$\begin{aligned}
&Pr(|\hat{q}(x,m) - q(x,m)| > \varepsilon^*\delta_n/2) \\
&\leq nPr\left(|\bar{M}_i - \bar{p}(x,\theta_i)| > \frac{m}{M}\varepsilon^*\delta_n/6 \mid X_i = x\right) + Pr\left(\frac{|\hat{N}(x,m)|}{\hat{D}(x,m)} > \varepsilon^*\delta_n/6\right) \\
&\leq 2nk_J \exp\left(-\frac{m^2\varepsilon^{*2}}{18M^2}\frac{N_J\delta_n^2}{k_J}\right) + \frac{12M\alpha_{k_J}}{m\varepsilon^*\delta_n} + Pr\left(\frac{|\hat{N}(x,m)|}{\hat{D}(x,m)} > \varepsilon^*\delta_n/6\right)
\end{aligned}$$

where the last inequality follows for any $k_J \geq 1$ by applying Lemma A.2. Take $k_J = k_0 \log(N_J)$ for some $k_0 > 0$. By Assumption C.6, for n large enough $J_n\delta_n^2 > \rho_0 \log(J_n)^2$. Therefore, because $N_J = \eta J_n$, ε^* can be chosen so that the first term here is bounded by an arbitrarily large power of n^{-1}. By Assumption C.4, k_0 can be chosen so that the second term is also bounded by an arbitrarily large power of n^{-1} for sufficiently large n.

Next, let $\hat{N}^*(x,m) = E(\hat{N}(x,m) \mid \{X_i, \theta_i, \{M_{i,j} : |j - j_0| > \eta J\}\})$. Then

$$Pr(|\hat{N}^*(x,m)| > \varepsilon/2) \leq \frac{\alpha_{\eta J}}{\varepsilon h_{1n}}$$

and since, by Assumption C.5, $|\hat{N}(x,m) - \hat{N}^*(x,m)| \leq \bar{\kappa}$, Bernstein's inequality can be applied to conclude that

$$Pr(|\hat{N}(x,m) - \hat{N}^*(x,m)| > \varepsilon/2) \leq \exp\left(-\frac{\frac{1}{2}nh_{1n}^2\varepsilon^2}{E((K\left(h_{1n}^{-1}(\bar{M}_i - m)\right)^2 \mathbf{1}(X_i = x)\eta_{j_0,i}^{*2}) + \frac{1}{3}\bar{\kappa}h_{1n}\varepsilon}\right)$$

where $\eta_{j_0,i}^* = \eta_{i,j_0} - E(\eta_{i,j_0} \mid X_i, \theta_i, \{M_{i,j} : |j - j_0| > \eta J\})$. This can be bounded further since

$$\begin{aligned}
E((K\left(h_{1n}^{-1}(\bar{M}_i - m)\right)^2 \mathbf{1}(X_i = x)\eta_{j_0,i}^{*2}) &\leq \frac{\bar{\kappa}^2}{4}Pr(|\bar{M}_i - m| \leq h_{1n} \mid X = x) \\
&\leq \bar{\kappa}^2 C h_{1n}
\end{aligned}$$

where the first inequality follows from Assumption C.5 and the second follows, for sufficiently large n, from Lemma C.3. Therefore, there are positive constants a and b such that

$$Pr(\hat{N}(x,m) > \varepsilon) \leq \exp\left(-\frac{anh_{1n}\varepsilon^2}{1 + b\varepsilon}\right) + \frac{\alpha_{\eta J}}{\varepsilon h_{1n}}$$

Next, there is a constant $d > 0$ such that $E(\hat{D}(x,m)) > d$ for sufficiently large n. To see this, by Assumption C.5 and Lemma C.3, $E(\hat{D}(x,m)) \geq \underline{\kappa}h_{1n}^{-1}Pr(|\bar{M} - m| < h_{1n}/2 \mid X =$

$x))Pr(X=x) \geq \underline{\kappa} c Pr(X=x)$ for sufficiently large n. Then

$$
\begin{aligned}
&Pr\left(\frac{|\hat{N}(x,m)|}{\hat{D}(x,m)} > \varepsilon^* \delta_n/6\right) \\
&\leq Pr(|\hat{N}(x,m)| > \varepsilon^* \delta_n d/6) + Pr(|\hat{D}(x,m) - E(\hat{D}(x,m))| > d/2) \\
&\leq \exp\left(-\frac{a n h_{1n}(\varepsilon^* \delta_n d/6)^2}{1 + b(\varepsilon^* \delta_n d/6)}\right) + \exp\left(-\frac{a' n h_{1n}(d/2)^2}{1+b'(d/2)}\right) + \frac{\alpha_{\eta J}}{\varepsilon h_{1n}}
\end{aligned}
$$

where the third inequality follows for some positive constants a', b' by applying Bernstein's inequality to $Pr(|\hat{D}(x,m) - E(\hat{D}(x,m))| > d/2)$ since (i) $\hat{D}(x,m) - E(\hat{D}(x,m))$ has mean 0 and is bounded and (ii) $E(\hat{D}(x,m) - E(\hat{D}(x,m)))^2 \leq E(\hat{D}(x,m)^2) \leq \bar{\kappa}^2 Pr(|\bar{M}_i - m| \leq h_{1n})$. The second and third terms are bounded by an arbitrarily large power of n^{-1} for sufficiently large n. And the first term is as well is ε^* is chosen large enough.

Therefore, I can conclude that

$$
J \sup_{x \in \mathcal{X}, m \in [\bar{p}(x, \underline{\theta}_x) - \gamma_n, \bar{p}(x, \bar{\theta}_x) + \gamma_n]} Pr(|\hat{q}(x,m) - q(x,m)| > \varepsilon^* \delta_n/2) \tag{C.1}
$$

bounded by an arbitrarily large power of n^{-1} for sufficiently large n as well.

Next,

$$
\begin{aligned}
&Pr(|\hat{q}(X_i, \bar{M}_i) - p_{j_0}(\theta_i)| > \varepsilon^* \delta_n) \\
&\leq Pr(|\hat{q}(X_i, \bar{M}_i) - q(X_i, \bar{M}_i) > \varepsilon^* \delta_n/2) + Pr(|q(X_i, \bar{M}_i) - p_{j_0}(\theta_i)| > \varepsilon^* \delta_n/2) \\
&\leq d_x J \sup_{x \in \mathcal{X}, m \in [\bar{p}(x, \underline{\theta}_x) - \gamma_n, \bar{p}(x, \bar{\theta}_x) + \gamma_n]} Pr(|\hat{q}(x,m) - q(x,m)| > \varepsilon^* \delta_n/2) \\
&+ Pr(\bar{M}_i \notin [\bar{p}(X_i, \underline{\theta}_{X_i}) - \gamma_n, \bar{p}(X_i, \bar{\theta}_{X_i}) + \gamma_n]) + Pr(|\bar{M}_i - \bar{p}(X_i, \theta_i)| > \frac{m \varepsilon^* \delta_n}{2M})
\end{aligned}
$$

where I have used that the support of X has d_x points and the support of $\bar{M}$ has J points and that $|q(X_i, \bar{M}_i) - p_{j_0}(\theta_i)| \leq \frac{M}{m} |\bar{M}_i - \bar{p}(X_i, \theta_i)|$. The second term is bounded by $Pr(|\bar{M}_i - \bar{p}(X_i, \theta_i)| > \gamma_n)$. Take $\gamma_n = \sqrt{\frac{\bar{\kappa}}{2}} (\log(J)/J)^{1/2}$, which satisfies the assumptions of Lemma C.3 by Assumption C.6. Then I can apply Lemma A.2 again to conclude that all three terms here are bounded by an arbitrarily large power of n^{-1} for sufficiently large n, which is the desired result. □

Lemma C.3. *There exist* $0 < c < C$ *such that for sufficiently large* n, $c h_{1n} \leq Pr(|\bar{M} - m| \leq h_{1n}/2 \mid X = x)$ *and* $Pr(|\bar{M} - m| \leq h_{1n} \mid X = x) \leq C h_{1n}$ *for all* (x, m) *such that* $x \in \mathcal{X}$ *and* $m \in [\bar{p}(x, \underline{\theta}_x) - \gamma_n, \bar{p}(x, \bar{\theta}_x) + \gamma_n]$ *where* $\gamma_n = o(h_{1n})$.

Proof. First,

$$
\begin{aligned}
&Pr(|\bar{M}_i - m| \leq h_{1n} \mid X_i = x) \qquad\qquad (C.2)\\
&\leq Pr(|\bar{p}(x,\theta_i) - m| \leq 2h_{1n} \mid X_i = x) + Pr(|\bar{M}_i - \bar{p}(x,\theta_i)| > h_{1n} \mid X_i = x)
\end{aligned}
$$

The first term is largest when $m \in [\bar{p}(x,\underline{\theta}_x) + 2h_{1n}, \bar{p}(x,\bar{\theta}_x) - 2h_{1n}]$ in which case it is equal to $Pr(|\bar{p}(x,\theta_i) - \bar{p}(x,t)| \leq 2h_{1n} \mid X_i = x)$ for some $t \in \Theta_x$, which is bounded by

$$
\begin{aligned}
Pr(|\theta_i - t| \leq 2h_{1n}/m \mid X_i = x) &= F_{\theta|X}(t + 2h_{1n}/m) - F_{\theta|X}(t - 2h_{1n}/m)\\
&\leq 4Bh_{1n}/m
\end{aligned}
$$

for some $B > 0$ by Assumption C.2. For any $s > 0$, the second term in (C.2) is bounded by $Pr(|\bar{M}_i - \bar{p}(x,\theta_i)| > s(log(J_n)/J_n)^{1/2} \mid X_i = x)$ for sufficiently large n. But by Lemma A.2, s can be chosen large enough so that this is $o(h_{1n})$.

Next,

$$
\begin{aligned}
&Pr(|\bar{M}_i - m| \leq h_{1n}/2 \mid X_i = x) \qquad\qquad (C.3)\\
&\geq Pr(|\bar{M} - m| \leq h_{1n}/2, |\bar{p}(x,\theta_i) - m| \leq h_{1n}/4 \mid X_i = x)\\
&= \int_{t:|\bar{p}(x,t)-m|\leq h_{1n}/4} (1 - Pr(|\bar{M} - m| > h_{1n}/2 \mid X_i = x, \theta_i = t)) f_{\theta|X}(t \mid x) dt \qquad (C.4)\\
&\geq \int_{t:|\bar{p}(x,t)-m|\leq h_{1n}/4} (1 - Pr(|\bar{M} - \bar{p}(x,t)| > h_{1n}/4 \mid X_i = x, \theta_i = t)) f_{\theta|X}(t \mid x) dt \qquad (C.5)
\end{aligned}
$$

Applying Lemma A.2 again, the integrand is no less than $1 + o(h_{1n})$. So for sufficiently large n,

$$
\begin{aligned}
&Pr(|\bar{M}_i - m| \leq h_{1n}/2 \mid X_i = x)\\
&\geq \frac{1}{2} Pr(|\bar{p}(x,\theta_i) - m| \leq h_{1n}/4 \mid X_i = x)(1 + o(h_{1n}))\\
&= \left(F_{\bar{p}(x,\theta)|X}(m + h_{1n}/4 \mid x) - F_{\bar{p}(x,\theta)|X}(m - h_{1n}/4 \mid x)\right)(1 + o(h_{1n}))
\end{aligned}
$$

Now, if $m - h_{1n}/4 > \bar{p}(x,\underline{\theta}_x)$ and $m + h_{1n}/4 < \bar{p}(x,\bar{\theta}_x)$ then this is no smaller than $\underline{f}_{\theta|X} h_{1n}/(2M)$ by Assumptions C.1 and C.2. If $m - h_{1n}/4 < \bar{p}(x,\underline{\theta}_x)$ then it must be the case that $m + h_{1n}/4 > \bar{p}(x,\underline{\theta}_x)$ for sufficiently large n because $m + h_{1n}/4 - \bar{p}(x,\underline{\theta}_x) > h_{1n}/4 - \gamma_n$

and $\gamma_n = o(h_{1n})$. Therefore, in that case,

$$\begin{aligned}
&F_{\bar{p}(x,\theta)|X}(m + h_{1n}/4 \mid x) - F_{\bar{p}(x,\theta)|X}(m - h_{1n}/4 \mid x) \\
&= F_{\bar{p}(x,\theta)|X}(m + h_{1n}/4 \mid x) - F_{\bar{p}(x,\theta)|X}(\bar{p}(x, \underline{\theta}_x) \mid x) \\
&\geq F_{\bar{p}(x,\theta)|X}(\bar{p}(x, \underline{\theta}_x) + h_{1n}/4 - \gamma_n \mid x) - F_{\bar{p}(x,\theta)|X}(\bar{p}(x, \underline{\theta}_x) \mid x) \\
&\geq \underline{f}_{\theta|X}(h_{1n}/4 - \gamma_n)/M
\end{aligned}$$

The same argument applies for the case where $m + h_{1n}/4 > \bar{p}(x, \bar{\theta}_x)$. The desired result follows with $\tilde{c} = \underline{f}_{\theta|X}/(8M)$. □

D Additional details regarding Section 3

The ability estimates, $\hat{\theta}_i$, rely crucially on the exclusion restriction, Assumption 2.4. Typically this could be justified based on the content of the questions if at least one question plausibly depends only on skills taught before the lowest schooling level observed. The nature of the available data precludes this because I do not have access to the text of the questions asked for each item. However, assuming that the exclusion is satisfied for *some* item, it is possible to identify this item j_0 under the additional assumption that schooling has a nonnegative effect on responses to the test questions. I implement this idea, as I will now describe, to obtain an estimate, $\widehat{j_0} = 4$.

First, evidence of an effect of education on the test score can be inferred without imposing Assumption 2.4. The proof of Theorem 2.1 includes the intermediate result,

$$E(Y \mid \bar{M} = m, X = x) = G(x, \bar{p}^{-1}(m; x)) + \delta_J,$$

where $\bar{p}^{-1}(m; x)$ is the inverse in t of the function $\bar{p}(x, t)$ and $\delta_J = o(1)$. Applying this here for each s, $E(\tilde{M}_s \mid \bar{M} = m, X = x) = p_s(x, \bar{p}^{-1}(m; x)) + \delta_J$. Define $q_s(x, m) = p_s(x, \bar{p}^{-1}(m; x))$. If $p_j(1, t) = p_j(0, t)$ for all t and all j then $q_j(1, m) = q_j(0, m)$ for all m and all j.[3] Let $\hat{q}_s(x, m)$ denote a Nadaraya-Watson kernel estimator of $E(\tilde{M}_s \mid \bar{M} = m, X = x)$ and $\hat{\sigma}_s^2(m)$ a consistent estimate of the asymptotic variance of $\hat{q}_s(1, m) - \hat{q}_s(0, m)$. Let $m_1, \ldots, m_R$ denote distinct points in the support of $\bar{M}$ and define the test statistic

$$T_s = \sum_{r=1}^{R} \frac{(\hat{q}_s(1, m_r) - \hat{q}_s(0, m_r))^2}{\sigma_s^2(m_r)}$$

[3]The converse is not true. For example, if $p_j(x, t) = p_1(x, t)$ for all j then $q_j(x, m) = m$ for all j and all x. Thus, the test proposed here will not have power against alternatives in this direction.

Under conditions ensuring that $\hat{q}_s(x, m)$ is an asymptotically normal estimator for $q_s(x, m)$, $T_s \to_d \chi^2_S$ if $q_s(1, m) = q_s(0, m)$. Let $\hat{p}_s$ denote the p-value associated with this test based on the χ^2_S distribution. To test the null hypothesis that college attendance does not have an effect on any items on the test, while controlling the familywise error rate, I use a Bonferroni correction. I reject the null at a significance level α if $\min_s \hat{p}_s \leq \frac{\alpha}{30}$. Table A.1 reports the results of this Chi-squared test for three different definitions of X_i. The null of no effect is rejected in each of the three cases. For the remaining results I focus on the second case, where X_i indicates whether the individual had completed high school or not at the time of the test.

To identify j_0, the item satisfying the exclusion restriction, I first observe that if $p_j(1, t) \geq p_j(0, t)$ for all t then $p_j^{-1}(\pi; 1) \leq p_j^{-1}(\pi; 0)$ for all m, where $p_j^{-1}(\pi; x)$ denotes the inverse in t of $p_j(x, t)$. Let $\psi_{j,k}(\pi) = p_k(1, p_j^{-1}(\pi; 1)) - p_k(0, p_j^{-1}(\pi; 0))$. If the exclusion restriction holds then $j_0 = \arg\min_k \max_{j \neq k} \int \psi_{j,k}(\pi)\omega_j(\pi)d\pi$ for any strictly positive weights $\omega_j(\pi)$.[4,5]

To implement this, the function $p_k(x, p_j^{-1}(\pi; x))$ can be estimated in two steps. First, compute $\hat{q}_j(x, m)$ as described above (excluding both $\tilde{M}_j$ and $\tilde{M}_k$ from $\bar{M}$). Second, compute a Nadaraya-Watson kernel estimator of $E(\tilde{M}_{ik} \mid X_i = x, \hat{q}_j(X_i, \bar{M}_i) = \pi)$. Computing this for both $x = 0$ and $x = 1$ and taking the difference, I obtain $\hat{\psi}_{j,k}(\pi)$ for π in the common support of $\hat{q}_j(1, \bar{M}_i) \mid X_i = 1$ and $\hat{q}_j(0, \bar{M}_i) \mid X_i = 0$. I then calculate

$$\widehat{j_0} = \arg\min_k \max_{j \neq k} n_j^{-1} \sum_{i=1}^{n} \hat{\psi}_{j,k}(\hat{q}_{ij}) \mathbf{1}(\hat{q}_{ij} \in \hat{\mathcal{S}}_j)$$

where $\hat{q}_{ij} = \hat{q}_j(X_i, \bar{M}_i)$, $\hat{\mathcal{S}}_j$ denotes an estimate of the common support and $n_j = \sum_{i=1}^n \mathbf{1}(\hat{q}_{ij} \in \hat{\mathcal{S}}_j)$. Figure A.1 plots $\max_{j \neq k} n_j^{-1} \sum_{i=1}^n \hat{\psi}_{j,k}(\hat{q}_{ij}) \mathbf{1}(\hat{q}_{ij} \in \hat{\mathcal{S}}_j)$ for each k.[6] It is evident from the graph that $\widehat{j_0} = 4$.

E Computation of the identified set

In this section I provide further details on the method described in Section 5 for computing the identified set. In E.1, I define the identified set for the model of Section 5. In E.2, I describe the proposed procedure and prove that the approximation is valid and in E.3, I describe computation of bounds on the identified set.

[4]This follows because $\psi_{j,j_0}(\pi) \leq 0$ for all j and $\psi_{j_0,k}(\pi) \geq 0$ for all k.

[5]If the exclusion restriction also holds for other items then j_0 is not the unique minimizer. In that case, all $k^* \in \arg\min_k \max_{j \neq k} \int \psi_{j,k}(\pi)\omega_j(\pi)d\pi$ satisfy the exclusion restriction.

[6]I exclude $k = 1, 2$ from consideration because these items are correctly answered at a rate exceeding 90% and would not sufficiently discriminate among different ability levels. That is, the corresponding response functions are too flat.

E.1 The identified set

The identified set can be defined as follows. First, stack the response functions into the length $J+1$ vector $\mathbf{p}(\cdot,\cdot)$. The triple

$$\gamma = (G(\cdot,\cdot), \mathbf{p}(\cdot,\cdot), F_{\theta|X}(\cdot \mid \cdot))$$

contains the "reduced form" parameters of the model. Let Γ denote the parameter space, that is, the space of all triples γ satisfying Assumptions 2.3, 2.4 and 5.2. For any γ, let

$$\mu_{c\mathcal{J}x}(\gamma) = \int G(x,t)^c p_{\mathcal{J}}(x,t) dF_{\theta|X}(t \mid x) \tag{E.1}$$

for any $c \in \{0,1\}$ and any $\mathcal{J} \subseteq \{1,\ldots,J+1\}$, and let $m_{c\mathcal{J}x}$ denote the observed moments, $E(Y^c \prod_{j\in\mathcal{J}} M_j \mid X = x)$.

Then the space of parameter values satisfying these moment conditions is

$$\mathcal{I} = \{\gamma \in \Gamma : \mu_{c\mathcal{J}x}(\gamma) = m_{c\mathcal{J}x} \text{ for all } c = 0,1, \mathcal{J} \subseteq \{1,\ldots,J\}, x \in \mathcal{X}\} \tag{E.2}$$

The identified set for any object of interest that can be written as $\tau = \tau(\gamma) \in \mathbb{R}$ is defined simply as $\tau(\mathcal{I})$. Further, I can define bounds

$$\underline{\tau} = \inf_{\gamma\in\mathcal{I}} \tau(\gamma), \qquad \bar{\tau} = \sup_{\gamma\in\mathcal{I}} \tau(\gamma). \tag{E.3}$$

E.2 Approximation method

Suppose that $\mathcal{X} = \{0,1\}$ and that $0 \le Y \le 1$. Since $\theta \sim Uniform(0,1)$, $F_{\theta|X}(t \mid 0) = \frac{t-\pi F_{\theta|X}(t|1)}{1-\pi}$ where $\pi = Pr(X=1)$. Then Γ is a subspace of the space of $2(J+1)$ functions from $[0,1]$ to $[0,1]$, $(G(0,\cdot), G(1,\cdot), p_1(0,\cdot), \ldots, p_J(1,\cdot), F_{\theta|X}(\cdot \mid 1))$, where the last function must satisfy $\pi F(t) \le t$. To approximate the identified set I further restrict Γ so that each of the first $2J+1$ functions is Lipschitz continuous with Lipschitz constant L and the last, $F_{\theta|X}(t \mid 1)$, admits a density function that is Lipschitz continuous with Lipschitz constant L.

Next, define a partition, $0 = t_0 \le t_1 \ldots \le t_S = 1$. Let $\mathcal{F}_S$ denote the space of length S vectors v such that $0 \le v_1 \le \ldots \le v_S \le 1$ and $\max_s |v_s - v_{s-1}| \le L|t_s - t_{s-1}|$. Then let Γ^S be the space of all $(\mathbf{v}, w)$ where $\mathbf{v} = (v_1, \ldots, v_{2J+1}) \in \mathcal{F}_S^{2J+1}$ and w is in the S-dimensional unit simplex and satisfies $\max_s w_s \le (w_s - w_{s-1})/\pi$ and $\max_s |w_s - w_{s-1}| \le L(t_s - t_{s-1})^2$. Let $\mathcal{I}^S$ denote the all $\gamma^S = (G_0^S, G_1^S, p_{10}^S, p_{11}^S, \ldots, p_{J0}^S, p_{J1}^S, \Delta^S) \in \Gamma^S$ such that $\mathcal{M}(\gamma^S) \le \epsilon_S$

where

$$\mu^S_{c\mathcal{J}x}(\gamma^S) := \sum_{s=1}^{S} (G^S_{x,s})^c p^S_{\mathcal{J}x,s} \left(\Delta^S x + \frac{(t_s - t_{s-1}) - \pi\Delta^S}{1-\pi}(1-x) \right)$$

and

$$\mathcal{M}(\gamma^S) = \sum_{c,\mathcal{J},x} w_{c,\mathcal{J},x}(m_{c\mathcal{J}x} - \mu^S_{c\mathcal{J}x}(\gamma^S))^2$$

Next, for any $\gamma = (\mathbf{v}, w) \in \Gamma^S$, I can define a unique $\bar{\gamma} = (\bar{\mathbf{v}}, F_w) \in \Gamma$ such that $\bar{\mathbf{v}}(t_s) = \mathbf{v}_s$ and $F_w(t_s) - F_w(t_{s-1}) = w_s$. I can do this via a linear interpolation between the points t_s and t_{s+1} and imposing $F_w(0) = 0$. Then define $\bar{\mathcal{I}}^S = \{\bar{\gamma}^S : \gamma^S \in \mathcal{I}^S\}$. Note that for any $\bar{\gamma}^S$, with a slight abuse of notation, we can say that $\mathcal{M}(\bar{\gamma}^S) \leq \epsilon_S$ as well.

The approximate bounds for a parameter $\tau(\gamma)$ are given by

$$\begin{aligned} \tau^l_S &= \min_{\gamma^S \in \mathcal{I}^S} \tau_S(\gamma^S) \\ \tau^u_S &= \max_{\gamma^S \in \mathcal{I}^S} \tau_S(\gamma^S) \end{aligned} \tag{E.4}$$

where τ_S is an approximate version of the mapping τ that is defined on the partition $\{t_s\}$ instead of on $[0,1]$.

Lastly, define the Hausdorff distance between two spaces,

$$d_H(\mathcal{A}, \mathcal{B}) = \max\{\sup_{a\in\mathcal{A}} \inf_{b\in\mathcal{B}} d(a,b), \sup_{b\in\mathcal{B}} \inf_{a\in\mathcal{A}} d(a,b)\}$$

where d is a metric defined on a space containing $\mathcal{A}$ and $\mathcal{B}$.

Theorem E.1. *Let $\delta_S = \max_{1\leq s\leq S}(t_s - t_{s-1})$. If $\delta_S \to 0$ and $\delta_S^2/\epsilon_S \to 0$ then $d_H(\mathcal{I}, \bar{\mathcal{I}}^S) \to 0$. As a result, if $\tau(\gamma)$ is a continuous functional and $\sup_{\gamma^S\in\Gamma^S_L} |\tau_S(\gamma^S) - \tau(\bar{\gamma}^S)| \to 0$, then $\tau^l_S \to \tau^l$ and $\tau^u_S \to \tau^u$.*

Proof. First, if $\gamma \in \mathcal{I}$ then define $\gamma^S \in \mathcal{I}^S$ by restricting $(G(0,\cdot), G(1,\cdot), p_1(0,\cdot), \ldots, p_J(1,\cdot))$ to the points $t_0, \ldots, t_S$ defining the partition and by defining $\Delta^S_s = \int_{t_{s-1}}^{t_s} f_{\theta|X}(t \mid 1)dt$.

Then

$$m_{c\mathcal{J}x} - \mu^S_{yc\mathcal{J}x}(\gamma^S) = \sum_{s=1}^{S} \int_{t_{s-1}}^{t_s} \left(G(x,t)^c p_{\mathcal{J}}(x,t) - G(x,t_s)^c p_{\mathcal{J}}(x,t_s)\right) f_{\theta|X}(t \mid 1)dt$$

and by monotonicity, $G(x,t_{s-1})^c p_{\mathcal{J}}(x,t_{s-1}) \leq G(x,t)^c p_{\mathcal{J}}(x,t) \leq G(x,t_s)^c p_{\mathcal{J}}(x,t_s)$ so

$$\int_{t_{s-1}}^{t_s} \left(G(x,t)^c p_{\mathcal{J}}(x,t) - G(x,t_s)^c p_{\mathcal{J}}(x,t_s)\right) f_{\theta|X}(t \mid 1)dt \leq 0$$

and

$$\begin{aligned}&\int_{t_{s-1}}^{t_s} \left(G(x,t)^c p_{\mathcal{J}}(x,t) - G(x,t_s)^c p_{\mathcal{J}}(x,t_s)\right) f_{\theta|X}(t \mid 1)dt \\ &\geq -\left((G(x,t_s))^c p_{\mathcal{J}}(x,t_s) - (G(x,t_{s-1}))^c p_{\mathcal{J}}(x,t_{s-1})\right) \int_{t_{s-1}}^{t_s} f_{\theta|X}(t \mid 1)dt\end{aligned}$$

Noting that $\pi f_{\theta|X}(t \mid 1) \leq 1$, by assumption, I can conclude that

$$\begin{aligned}|m_{c\mathcal{J}x} - \mu^S_{yc\mathcal{J}x}(\gamma^S)| &\leq \delta_S \left((G(x,t_S))^c p_{\mathcal{J}}(x,t_S) - (G(x,t_0))^c p_{\mathcal{J}}(x,t_0)\right) \\ &\leq \delta_S\end{aligned}$$

Therefore, since $\sum_{c,\mathcal{J},x} |w_{c,\mathcal{J},x}| = 1$, $\mathcal{M}(\gamma^S) \leq \delta_S^2$. By assumption, $\delta_S/\epsilon_S \to 0$ and therefore, $\mathcal{M}(\gamma^S) \leq \epsilon_S$ for sufficiently large S. Hence $\gamma^S \in \mathcal{I}^S$ for S large enough. For any γ^S, the linear interpolation, $\bar{\gamma}^S$, must satisfy $d(\gamma^S, \bar{\gamma}^S) \leq L\delta_S$, implying that

$$\sup_{\gamma \in \mathcal{I}} \inf_{\bar{\gamma}^S \in \bar{\mathcal{I}}^S} d(\gamma, \bar{\gamma}^S) \leq \delta_S \to 0$$

Next, suppose $\sup_{\bar{\gamma}^S \in \bar{\mathcal{I}}^S} \inf_{\gamma \in \mathcal{I}} d(\gamma, \bar{\gamma}^S) \not\to 0$. Then there is a subsequence, $\{\bar{\gamma}^{S_k}\}$, such that $S_k \to \infty$ as $k \to \infty$ but $\inf_{\gamma \in \mathcal{I}} d(\gamma, \bar{\gamma}^{S_k}) > \rho$ for all k. However, because Γ is compact, this subsequence has a further subsequence along which it converges to some $\gamma^\infty \in \Gamma$.

Then, because $\mathcal{M}(\bar{\gamma}^S) \leq \epsilon_S$, it must be the case that $|m_{c\mathcal{J}x} - \mu^S_{c\mathcal{J}x}(\bar{\gamma}^S)| \to 0$ for each $c \in \{0,1\}$, $\mathcal{J}$ and $x \in \{0,1\}$. So, using the fact that, by definition of γ^∞, $d(\bar{\gamma}^{S_{k_m}}, \gamma^\infty) \to 0$

$$\begin{aligned}&\left| m_{c\mathcal{J}x} - \int_0^1 (G^\infty(x,t))^c p^\infty_{\mathcal{J}}(x,t)dt \right| \\ &\leq |\mathcal{P}_{y,\mathcal{J}} - \mu^{S_{k_m}}_{c\mathcal{J}x}(\bar{\gamma}^{S_{k_m}})| + \left| \mu^{S_{k_m}}_{c\mathcal{J}x}(\bar{\gamma}^{S_{k_m}}) - \int_0^1 (G^{S_{k_m}}(x,t))^c p^{S_{k_m}}_{\mathcal{J}}(x,t)dt \right| + d(\bar{\gamma}^{S_{k_m}}, \gamma^\infty) \\ &\longrightarrow 0,\end{aligned}$$

which implies that $\left| m_{c\mathcal{J}x} - \int_0^1 (G^\infty(x,t))^c p^\infty_{\mathcal{J}}(x,t)dt \right| = 0$ and, hence, that $\gamma^\infty \in \mathcal{I}$. So $\inf_{\gamma \in \mathcal{I}} d(\gamma, \bar{p}^{S_{k_m}}) \leq d(\gamma^\infty, \bar{\gamma}^{S_{k_m}}) \to 0$, which is a contradiction. Thus the first part of the theorem has been proven.

The second conclusion of the theorem follows because

$$\begin{aligned}\sup_{\gamma^S \in \mathcal{I}^S} \tau_S(\gamma^S) &\leq \sup_{\gamma \in \mathcal{I}} \tau(\gamma) + \sup_{\gamma^S \in \Gamma^S} |\tau_S(\gamma^S) - \tau(\bar{\gamma}^S)| + d_H(\tau(\mathcal{I}), \tau(\bar{\mathcal{I}}^S)) \\ &\longrightarrow \sup_{\gamma \in \mathcal{I}} \tau(\gamma)\end{aligned}$$

and

$$\begin{aligned}\inf_{\gamma^S \in \mathcal{I}^S} \tau_S(\gamma^S) &\leq \inf_{\gamma \in \mathcal{I}} \tau(\gamma) - \sup_{\gamma^S \in \Gamma^S} |\tau_S(\gamma^S) - \tau(\bar{\gamma}^S)| - d_H(\tau(\mathcal{I}), \tau(\bar{\mathcal{I}}^S)) \\ &\longrightarrow \inf_{\gamma \in \mathcal{I}} \tau(\gamma)\end{aligned}$$

□

E.3 Estimation of bounds

Following Chernozhukov et al. (2013), I replace the construction (E.4) with the computationally simpler problem,

$$\begin{aligned}\tau_S^l &= \min\{\tau : Q^S(\tau) \leq \epsilon\} \\ \tau_S^u &= \max\{\tau : Q^S(\tau) \leq \epsilon\}\end{aligned} \tag{E.5}$$

where

$$\begin{aligned}Q^S(\tau) &= min_{\gamma^S \in \Gamma^S} \mathcal{M}(\gamma^S) \\ &\text{s.t. } \tau(\gamma^S) = \tau\end{aligned} \tag{E.6}$$

To estimate the bounds, replace $\pi = Pr(X = 1)$ with $\hat{\pi} = n^{-1} \sum_{i=1}^{n} X_i$ and

$$\hat{\mathcal{M}}(\gamma^S) = \sum_{c,\mathcal{J},x} w_{c,\mathcal{J},x} (\hat{m}_{c\mathcal{J}x} - \mu^S_{c\mathcal{J}x}(\gamma^S))^2$$

where $\hat{m}_{c\mathcal{J}x}$ is the sample analogue of the moments $m_{c\mathcal{J}x}$. Then

$$\begin{aligned}\tau_S^l &= \min\{\tau : \hat{Q}^S(\tau) \leq \epsilon\} \\ \tau_S^u &= \max\{\tau : \hat{Q}^S(\tau) \leq \epsilon\}\end{aligned} \tag{E.7}$$

where

$$\hat{Q}^S(\tau) = min_{\gamma^S \in \Gamma^S} \hat{\mathcal{M}}(\gamma^S) \tag{E.8}$$
$$\text{s.t. } \tau(\gamma^S) = \tau$$

References

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): "Average and quantile effects in nonseparable panel models," *Econometrica*, 81, 535–580.

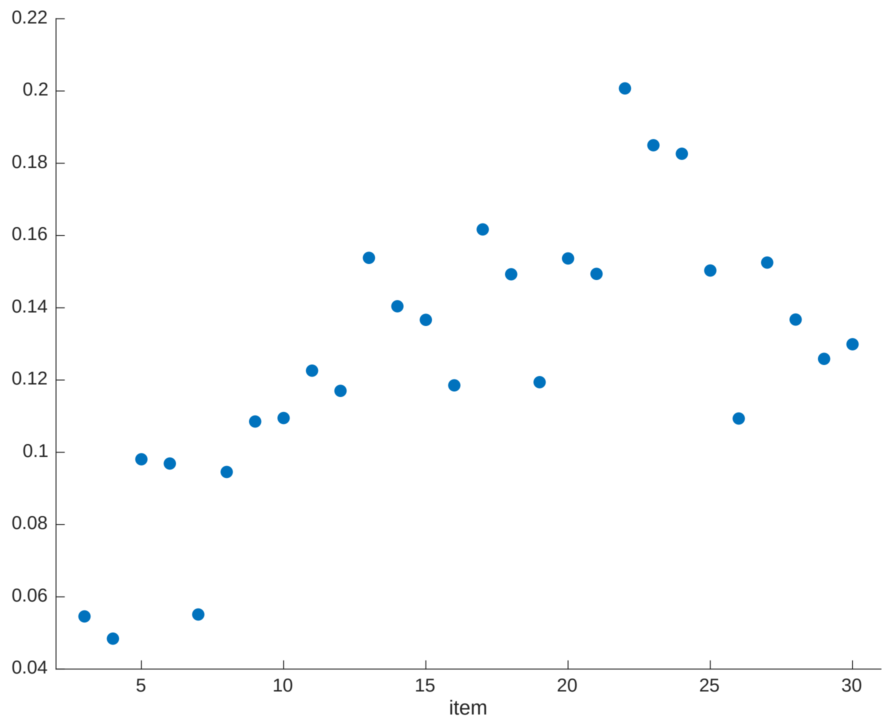
Van der Vaart, A. W. (2000): *Asymptotic statistics*, vol. 3, Cambridge university press.

Table 1. Chi-squared test of no schooling effect

A. ≥ 11 years				B. ≥ 12 years				C. Some college			
item	adj. p-value	item	adj. p-value	item	adj. p-value	item	adj. p-value	item	adj. p-value	item	adj. p-value
1	27.76	16	8.52	1	22.55	16	16.47	1	3.84	16	6.90
2	8.54	17	0.00***	2	14.19	17	0.05**	2	13.05	17	1.44
3	7.87	18	16.34	3	16.36	18	5.77	3	8.06	18	0.00***
4	11.79	19	1.91	4	2.89	19	11.08	4	1.37	19	2.48
5	12.41	20	2.95	5	28.81	20	10.99	5	3.05	20	13.88
6	1.13	21	1.49	6	15.74	21	1.89	6	4.81	21	2.06
7	11.13	22	0.16	7	18.65	22	0.01***	7	6.82	22	0.00***
8	9.26	23	26.54	8	6.47	23	0.42	8	2.99	23	0.00***
9	2.22	24	18.51	9	4.70	24	4.16	9	17.04	24	0.00***
10	4.13	25	17.87	10	2.18	25	29.19	10	5.87	25	17.28
11	21.30	26	7.74	11	29.28	26	5.94	11	14.17	26	0.17
12	24.89	27	19.98	12	23.87	27	14.91	12	2.83	27	5.81
13	13.23	28	29.55	13	1.59	28	19.47	13	0.00***	28	0.00***
14	8.54	29	28.71	14	25.59	29	15.70	14	0.12	29	0.12
15	0.43	30	9.45	15	1.29	30	3.81	15	0.12	30	2.09

Notes: For each item a chi-squared test was performed as described in the text. Each p-value is adjusted by multiplying by 30 (the Bonferonni correction) and the adjusted p-values are reported in this table. The headings for the three panels indicate how the education variable was defined. For example, for the results in Panel A, the education variable indicated whether the individual had completed at least 11 years of education. In each case the entire sample of 1,929 individuals was used. *,**, and *** indicate significance at a 10%, 5%, and 1% level.

Figure A.1: Choice of j_0 among items in the AR component of the ASVAB



Notes: This figure plots $\max_{j \neq k} n_j^{-1} \sum_{i=1}^{n} \hat{\psi}_{j,k}(\hat{q}_{ij}) \mathbf{1}(\hat{q}_{ij} \in \hat{\mathcal{S}}_j)$ for each k from 3 to 30, as described in online appendix D. The Epanechnikov kernel was used in constructing the estimates $\hat{q}_j$, as well as $\hat{\psi}_{j,k}$. Estimates are based on a sample of size $1,927$ from the NLSY79.