SUPPLEMENT TO "ON THE EFFICIENCY OF HIGHLY STRATIFIED EXPERIMENTS"

BY YUEHAO BAI[1,a], JIZHOU LIU[2,b], AZEEM M. SHAIKH[3,c], AND MAX TABORD-MEEHAN[4,d]

[1]Department of Economics, University of Southern California, [a]yuehao.bai@usc.edu

[2]HSBC Business School, Peking University, [b]jizhou.liu@phbs.pku.edu.cn

[3]Department of Economics, University of Chicago, [c]amshaikh@uchicago.edu

[4]Department of Economics, University of Toronto, [d]m.tabordmeehan@utoronto.ca

APPENDIX A: PROOFS OF MAIN RESULTS

A.1. Proof of Theorem 3.1. First note (10) follows from (9) and Lemma A.6. In particular, the second component of the decomposition therein is zero because $E[\psi^*|X, A] = E[\psi^*|X]$. To show (9), we first establish (12), i.e.,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -M^{-1}\frac{1}{\sqrt{n}}\sum_{1\le i\le n} m(X_i, A_i, R_i, \theta_0) + o_P(1) .$$

By the proof of Theorem 5.21 in [14], to show (12), it suffices to show

$$\mathbb{L}_n(\hat{\theta}_n) \xrightarrow{P} 0 , \tag{S.1}$$

where $\mathbb{L}_n(\theta) = (\mathbb{L}_n^{(1)}(\theta), \dots, \mathbb{L}_n^{(d_\theta)}(\theta))'$ for

$$\mathbb{L}_n^{(s)}(\theta) = \frac{1}{\sqrt{n}}\sum_{1\le i\le n}(m_s(X_i, A_i, R_i, \theta) - E_P[m_s(X_i, A_i, R_i, \theta)])$$
$$- \frac{1}{\sqrt{n}}\sum_{1\le i\le n}(m_s(X_i, A_i, R_i, \theta_0) - E_P[m_s(X_i, A_i, R_i, \theta_0)]) .$$

To accomplish this, we study $\mathbb{L}_n^{(s)}(\theta)$ for $1 \le s \le d_\theta$ separately. It follows from Assumption 3.3(c)–(d), Proposition 8.11 in [9], and the arguments to establish (S.24) that

$$\sup_{\theta\in\Theta:\|\theta-\theta_0\|<\delta} |\mathbb{L}_n^{(s)}(\theta)| = \sup_{\theta\in\Theta^*:\|\theta-\theta_0\|<\delta} |\mathbb{L}_n^{(s)}(\theta)| .$$

Therefore, since $\hat{\theta}_n \xrightarrow{P} \theta_0$ by Lemma A.8, to show (S.1) it suffices to argue that for every $\epsilon > 0$ and every sequence $\delta_n \downarrow 0$ [p.89 of 15],

$$\lim_{n\to\infty} P\left\{\sup_{\theta\in\Theta^*:\|\theta-\theta_0\|<\delta_n} \left|\mathbb{L}_n^{(s)}(\theta)\right| > \epsilon\right\} = 0 . \tag{S.2}$$

Following the arguments in the proof of Lemma A.8, we decompose $\mathbb{L}_n^{(s)}(\theta) = \mathbb{L}_{n,1}^{(s)}(\theta) + \mathbb{L}_{n,0}^{(s)}(\theta)$, where

$$\mathbb{L}_{n,1}^{(s)}(\theta) = \frac{1}{\sqrt{n}}\sum_{1\le i\le n} A_i(m_s(X_i, 1, R_i(1), \theta) - m_s(X_i, 1, R_i(1), \theta_0)$$
$$- E[m(X_i, 1, R_i(1), \theta) - m_s(X_i, 1, R_i(1), \theta_0)])$$

$$\mathbb{L}_{n,0}^{(s)}(\theta) = \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} (1 - A_i)(m_s(X_i, 0, R_i(0), \theta) - m_s(X_i, 0, R_i(0), \theta_0) \\ - E[m_s(X_i, 0, R_i(0), \theta) - m_s(X_i, 0, R_i(0), \theta_0)]) \ .$$

Define

$$\rho_Q(\theta, \theta_0) = E_Q[(m_s(X, a, R(a), \theta) - m_s(X, a, R(a), \theta_0) \\ - E_Q[m_s(X, a, R(a), \theta) - m_s(X, a, R(a), \theta_0)])^2]^{1/2} \ .$$

Note by Assumption 3.3(c) that $\rho_Q(\theta, \theta_0)$ is continuous in θ, i.e., as $\|\theta - \theta_0\| \to 0$,

$$\rho_Q(\theta, \theta_0) \le E_Q[(m_s(X, a, R(a), \theta) - m_s(X, a, R(a), \theta_0))^2]^{1/2} \to 0 \ .$$

Fix any sequence $\tilde{\delta}_n \downarrow 0$. For every n, there exists n' such that $\{\theta \in \Theta^* : \|\theta - \theta_0\| < \delta_{n'}\} \subseteq \{\theta \in \Theta^* : \rho_Q(\theta, \theta_0) < \tilde{\delta}_n\}$. By Proposition C.1 in [7],

$$E\bigg[\sup_{\rho_Q(\theta,\theta_0) < \tilde{\delta}_n} |\mathbb{L}_{n,a}^{(s)}(\theta)|\bigg] E\bigg[\sup_{\rho_Q(\theta,\theta_0) < \tilde{\delta}_n} \bigg|\frac{1}{\sqrt{n}} \sum_{1 \le i \le n} (m_s(X_i, 1, R_i(1), \theta) \\ - m_s(X_i, 1, R_i(1), \theta_0) - E[(m_s(X_i, 1, R_i(1), \theta) - m_s(X_i, 1, R_i(1), \theta_0)])\bigg|\bigg] \to 0 \ .$$

where the convergence follows from Assumption 3.3(e) and Corollary 2.3.12 in [15]. We then obtain (S.2) by Markov's inequality.

Finally, we derive (9) from (12). Note that

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{1 \le i \le n} m(X_i, A_i, R_i, \theta_0) \\ &= \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} \Big(\eta E[m(X_i, 1, R_i(1), \theta_0)|X_i] + (1 - \eta) E[m(X_i, 0, R_i(0), \theta_0)|X_i] \\ &\qquad + I\{A_i = 1\}(m(X_i, 1, R_i, \theta_0) - E[m(X_i, 1, R_i(1), \theta_0)|X_i]) \\ &\qquad + I\{A_i = 0\}(m(X_i, 0, R_i, \theta_0) - E[m(X_i, 0, R_i(0), \theta_0)|X_i]) \\ &\qquad + (A_i - \eta)(E[m(X_i, 1, R_i(1), \theta_0) - m(X_i, 0, R_i(0), \theta_0)|X_i])\Big) \ . \end{aligned}$$

Let $\Omega(X_i) = E[m(X_i, 1, R_i(1), \theta_0) - m(X_i, 0, R_i(0), \theta_0)|X_i]$ and note that by Assumption 2,

$$E\bigg[\frac{1}{\sqrt{n}} \sum_{1 \le i \le n} (A_i - \eta)\Omega(X_i)\bigg| X^{(n)}\bigg] = 0 \ .$$

Recall $\Omega^{(s)}(X_i)$ is the sth component of $\Omega(X_i)$. Next, it follows from Assumption 3.1, 3.3(f), and equation (12.3) in [12] that for $1 \le s \le d_\theta$,

$$\begin{aligned} \operatorname{Var}\bigg[\frac{1}{\sqrt{n}} \sum_{1 \le i \le n} (A_i - \eta)\Omega^{(s)}(X_i)\bigg| X^{(n)}\bigg] &= \frac{1}{n} \sum_{1 \le j \le n/k} \frac{\ell(k - \ell)}{k - 1} \sum_{i \in \lambda_j} (\Omega_i^{(s)} - \overline{\Omega}_j^{(s)})^2 \\ &\le C^2 \frac{\ell(k - \ell)}{k - 1} \frac{1}{n} \sum_{1 \le j \le n/k} \max_{i, i' \in \lambda_j} \|X_i - X_{i'}\|^2 \ , \end{aligned}$$

where $\overline{\Omega}_j^{(s)} = \frac{1}{k}\sum_{i\in\lambda_j}\Omega^{(s)}(X_i)$, and so the conditional variance converges in probability to zero under Assumption 3.2. It then follows from Markov's inequality and the fact that probabilities are bounded and hence uniformly integrable that

$$\frac{1}{\sqrt{n}}\sum_{1\le i\le n}(A_i-\eta)\Omega(X_i) = o_P(1)\ .$$

Therefore,

$$-M^{-1}\frac{1}{\sqrt{n}}\sum_{1\le i\le n} m(X_i,A_i,R_i,\theta_0) = -M^{-1}\frac{1}{\sqrt{n}}\sum_{1\le i\le n} m^*(X_i,A_i,R_i,\theta_0) + o_P(1)\ ,$$

which, together with (12), implies the desired result in (9). ■

A.2. Proof of Theorem 3.2. By assumption $\widehat{M}_n \xrightarrow{P} M$. Therefore, it suffices to show that

$$\hat{\Sigma}_{1,n} \xrightarrow{P} \Sigma_1 \tag{S.3}$$

$$\hat{\Sigma}_{2,n} \xrightarrow{P} \Sigma_2\ , \tag{S.4}$$

where

$$\begin{aligned}\Sigma_1 &= \eta\operatorname{Var}[m(X_i,1,R_i(1),\theta_0)] + (1-\eta)\operatorname{Var}[m(X_i,0,R_i(0),\theta_0)]\\ \Sigma_2 &= -\eta(1-\eta)\operatorname{Var}\big[E[m(X_i,1,R_i(1),\theta_0)|X_i] - E[m(X_i,1,R_i(1),\theta_0)]\\ &\qquad - (E[m(X_i,0,R_i(0),\theta_0)|X_i] - E[m(X_i,0,R_i(0),\theta_0)])\big]\ .\end{aligned}$$

In what follows, we will show (S.3). The proof of (S.4) will follow from similar steps, but with the calculations below replaced by the ones in the proof of Lemmas C.2–C.3 in [2], along with Assumptions 3.4–3.5. To that end, we show for $1\le s\le d_\theta$,

$$\hat{\mu}_{1,n}^{(s)} := \frac{1}{\eta n}\sum_{1\le i\le n} m_s(X_i,1,R_i,\hat{\theta}_n) \xrightarrow{P} E[m_s(X_i,1,R_i(1),\theta_0)] =: \mu_1^{(s)}\ , \tag{S.5}$$

and similar arguments will establish the results for $a=0$ as well as for the second moments and therefore (S.3). For $\theta\in\Theta$, define

$$\hat{\mu}_{1,n}^{(s)}(\theta) = \frac{1}{\eta n}\sum_{1\le i\le n} m_s(X_i,1,R_i,\theta) \tag{S.6}$$

and note $\hat{\mu}_{1,n}^{(s)} = \hat{\mu}_{1,n}^{(s)}(\hat{\theta}_n)$. Suppose (S.5) doesn't hold. Then, there exists a subsequence $\{n_k\}_{k\ge1}$ and $\epsilon_1,\epsilon_2>0$, such that

$$\lim_{k\to\infty} P\{|\hat{\mu}_{1,n}^{(s)} - \mu_1^{(s)}| > \epsilon_1\} \to \epsilon_2\ . \tag{S.7}$$

Because $\hat{\theta}_{n_k} \xrightarrow{P} \theta_0$ by Theorem 3.1, there exists a further subsequence, which we still denote by $\{n_k\}_{k\ge1}$ by an abuse of notation, along which $\hat{\theta}_{n_k}\to\theta_0$ with probability one. Along that subsequence, Lemma A.9 implies that $\hat{\mu}_{1,n}^{(s)} \xrightarrow{P} \mu_1^{(s)}$, in contradiction to (S.7). Therefore, (S.5) holds, and the theorem follows as discussed above. ■

A.3. Proofs for Section 4.1. Recall that P_n denotes the distribution of the observed data $(X^{(n)}, A^{(n)}, R^{(n)})$, and Q denotes the marginal distribution of the vector $(R_i(1), R_i(0), X_i)$. Note that any treatment assignment mechanism $A^{(n)}$ satisfying Assumption 2.1 can be represented as a function of $X^{(n)}$ and some additional exogenous randomization device $U_n \in \mathbf{R}$. Let $p_n^{U_n}$ denote the density function for U_n with respect to a dominating measure μ^U. In what follows, we consider a family $\{Q_t : t \in \mathbf{R}^{d_\theta}\}$ of marginal distributions indexed by t, and let q_t^X denote the density function for X_i with respect to a dominating measure μ^X, $q_t^{R(a)|X}(r|x)$ denote the conditional density of $R_i(a)$ given X_i with respect to a dominating measure μ^R. With some abuse of notation, continue letting $P_{t,n}$ denote the distribution of $(U_n, X^{(n)}, R^{(n)})$. We require that $Q_0 = Q$ and $P_{0,n} = P_n$ and define $q^X = q_0^X$ and $q^{R(a)|X} = q_0^{R(a)|X}$. As a consequence, the density function of $P_{t,n}$ is given by

$$\ell_n = p_n^U(U_n) \prod_{1\le i\le n} q_t^X(X_i) \prod_{1\le i\le n} \prod_{a\in\{0,1\}} q_t^{R(a)|X}(R_i|X_i)^{I\{A_i=a\}} . \tag{S.8}$$

Because the density $p_n^{U_n}$ does not depend on t, and in general we will only concern ourselves with the ratio of likelihoods at different values of t (so that $p_n^{U_n}$ in the ratio will cancel), in what follows we suppress the dependence on n and simply denote the distribution $P_{t,n}$ by P_t.

We consider parametric submodels $\{P_t : t \in \mathbf{R}^{d_\theta}\}$, where $P_0 = P$, such that the following holds for some $g = (g^X, g^{R(1)|X}, g^{R(0)|X})$, each component of which is a d_θ-dimensional function:

(a) As $t \to 0$,

$$\int \frac{1}{\|t\|^2}\Big(q_t^X(x)^{1/2} - q^X(x)^{1/2} - \frac{1}{2} q^X(x)^{1/2} t' g^X(x)\Big)^2 d\mu^X(x) \to 0 . \tag{S.9}$$

(b) For $a \in \{0,1\}$, as $t \to 0$,

$$\frac{1}{\|t\|^2}\iint \Big(q_t^{R(a)|X}(r|x)^{1/2} - q^{R(a)|X}(r|x)^{1/2} - \frac{1}{2} q^{R(a)|X}(r|x)^{1/2} t' g^{R(a)|X}(r|x)\Big)^2 \times d\mu^R(r) q^X(x) d\mu^X(x) \to 0 . \tag{S.10}$$

In what follows, we will index a parametric submodel by its associated function g, denoted by $P_{t,g}$, to emphasize the role of g. Similarly we denote the density of $Q_{t,g}$ by $q_{t,g}$. When writing expectations and variances, we suppress the subscripts P and Q whenever doing so does not lead to confusion. For completeness, we document the following properties of score functions which satisfy (S.9)–(S.10):

LEMMA A.1. *For a parametric submodel $\{P_{t,g} : t \in \mathbf{R}^{d_\theta}\}$ with $P_{0,g} = P$ that satisfies* (S.9)–(S.10),

(a) $E[g^X(X) g^X(X)'] < \infty$.
(b) $E[g^X(X)] = 0$.
(c) $E[g^{R(a)|X}(R(a)|X) g^{R(a)|X}(R(a)|X)'] < \infty$ *and hence* $I^{R(a)|X}(X) < \infty$ *with probability one under* Q.
(d) $E[g^{R(a)|X}(R(a)|X)|X] = 0$ *with probability one under* Q.

PROOF. (a) and (b) follow from Lemma 14.2.1 in [12]. (c) follows from the same lemma. In order to show (d), fix $t_n \to 0$. Note (S.10) and Markov's inequality imply that along a subsequence t_{n_k},

$$\frac{1}{\|t_{n_k}\|^2}\int\Big(q_{t_{n_k}}^{R(a)|X}(r|x)^{1/2}-q^{R(a)|X}(r|x)^{1/2}-\frac{1}{2}q^{R(a)|X}(r|x)^{1/2}t_{n_k}'g^{R(a)|X}(r|x)\Big)^2 \times d\mu^R(r)\to 0$$

for Q-almost every x. Along that subsequence, another application of Lemma 14.2.1 in [12] implies (d). ∎

Define the information of X as $I^X = E[g^X(X)g^X(X)']$. Define the conditional information of $R(a)$ given $X=x$ as

$$I^{R(a)|X}(x)=E[g^{R(a)|X}(R(a)|X)g^{R(a)|X}(R(a)|X)'|X=x]\ .$$

Further define $I = I^X + \eta E[I^{R(1)|X}(X)] + (1-\eta)E[I^{R(0)|X}(X)]$. We restrict ourselves to parametric submodels that satisfy (S.9)–(S.10) for a g that satisfies the following conditions. These submodels exist by Lemma A.2 below.

CONDITION A.1. The function g satisfies that

(a) $E[g^X(X)]=0$ and $\mathrm{Var}[g^X(X)]<\infty$.
(b) For $a\in\{0,1\}$, $E[g^{R(a)|X}(R(a)|X)|X]=0$, and $\mathrm{Var}[g^{R(a)|X}(R(a)|X)]<\infty$ with probability one.
(c) I is nonsingular.

LEMMA A.2. *For any g that satisfies Condition A.1, there exists a parametric submodel $\{P_{t,g}: t\in\mathbf{R}^{d_\theta}\}$ such that (S.9)–(S.10) hold.*

PROOF. We use a vector version of the construction in in Example 25.16 in [14]. Let $k(x)$ be any strictly positive function that is bounded from above and away from zero with a bounded derivative such that $k(0)=k'(0)=1$; for example, take $k(x)=2(1+e^{-2x})^{-1}$. Define

$$q_t^X(x)=C(t)q^X(x)k(t'g^X(x))\ ,$$

where $C(t)=\big(\int q^X(x)k(t'g^X(x))d\mu^X(x)\big)^{-1}$, so that $q_t^X(x)$ is a probability density function. Differentiating both sides of $C(t)\int q^X(x)k(t'g^X(x))d\mu^X(x)=1$ at $t=0$, we get that $\frac{\partial}{\partial t}\big|_{t=0}C(t)=0$. It can then be verified through direct calculation that

$$\frac{\partial}{\partial t}\bigg|_{t=0}\log q_t^X(x)=g^X(x)\ .$$

The quadratic mean differentiability requirement in (S.9) follows from Lemma 7.6 in [14]. Next, for each $x\in\mathbf{R}^{d_x}$, we define

$$q_t^{R(1)|X}(r|x)=C(t)q^{R(1)|X}(r|x)k(t'g^{R(1)|X}(r|x))\ .$$

As above, it can be verified through direct calculation that

$$\frac{\partial}{\partial t}\bigg|_{t=0}\log q_t^{R(1)|X}(r|x)=g^{R(1)|X}(r|x)\ .$$

To show (S.10) for $a=1$ (and symmetric arguments apply for $a=0$), we modify the arguments in the proof of Lemma 7.6 in [14]. Define $s_t(r|x)=q_t^{R(1)|X}(r|x)^{1/2}$ and denote the ℓth component by $s_t^{(\ell)}(r|x)$ for $1\le\ell\le d_\theta$. By the mean-value theorem, we have

$s_t(r|x) - s(r|x) = \int_0^1 t'\dot{s}_{ut}(r|x)du$, where $\dot{s}_t = \frac{\partial}{\partial t}s_t$, so that it follows from Jensen's inequality that

$$\frac{1}{\|t\|^2}\iint \left(s_t(r|x) - s_0(r|x) - t'\dot{s}_0(r|x)\right)^2 d\mu^R(r)q^X(x)d\mu^X(x)$$
$$\leq \frac{1}{\|t\|^2}\iiint_0^1 \left(t'(\dot{s}_{ut}(r|x) - \dot{s}_0(r|x))\right)^2 dud\mu^R(r)q^X(x)d\mu^X(x)$$
$$\leq \iiint_0^1 \|\dot{s}_{ut}(r|x) - \dot{s}_0(r|x)\|^2 dud\mu^R(r)q^X(x)d\mu^X(x) \tag{S.11}$$

where the first inequality follows from Jensen's inequality and the second by Cauchy-Schwarz. It then suffices to show (S.11) goes to zero as $t \to 0$. Analyzing componentwise, it suffices to show that for $1 \leq \ell \leq d_\theta$, as $t \to 0$,

$$\iiint_0^1 (\dot{s}_{ut}^{(\ell)}(r|x) - \dot{s}_0^{(\ell)}(r|x))^2 dud\mu^R(r)q^X(x)d\mu^X(x) \to 0\,. \tag{S.12}$$

The integrand in (S.12) obviously converges to zero as $t \to 0$ by continuous differentiability of q_t. Recall that $\int \dot{s}_{ut}^{(\ell)}(r|x)^2 d\mu^R(r) = \frac{1}{4}[I_{ut}^{R(1)|X}(x)]_{(\ell,\ell)}$, where $I_{ut}^{R(1)|X}(x)$ is the conditional information for $a = 1$ given x at $P_{ut,g}$. Therefore, it follows from Fubini's theorem that

$$\iiint_0^1 \dot{s}_0^{(\ell)}(r|x)^2 dud\mu^R(r)q^X(x)d\mu^X(x) = \frac{1}{4}\int [I_0^{R(1)|X}(x)]_{(\ell,\ell)}q^X(x)d\mu^X(x)$$
$$\iiint_0^1 \dot{s}_{ut}^{(\ell)}(r|x)^2 dud\mu^R(r)q^X(x)d\mu^X(x) = \frac{1}{4}\iint_0^1 [I_{ut}^{R(1)|X}(x)]_{(\ell,\ell)}duq^X(x)d\mu^X(x)\,.$$

To apply Vitali's theorem [Proposition 2.29 in 14], it suffices to show that

$$\iint_0^1 [I_{ut}^{R(1)|X}(x)]_{(\ell,\ell)}duq^X(x)d\mu^X(x) \to \int [I_0^{R(1)|X}(x)]_{(\ell,\ell)}q^X(x)d\mu^X(x)$$

as $t \to 0$. To do so, we fix any arbitrarily small $\delta > 0$ and note that at least for t small enough, $\|ut\| \leq \delta$ for $u \in [0,1]$, so we can apply the dominated convergence theorem with

$$[I_{ut}^{R(1)|X}(x)]_{(\ell,\ell)} \leq \sup_{\|h\|\leq\delta}[I_h^{R(1)|X}(x)]_{(\ell,\ell)}\,,$$

as long as

$$\begin{aligned}&\iint_0^1 \sup_{\|h\|\leq\delta}[I_h^{R(1)|X}(x)]_{(\ell,\ell)}duq^X(x)d\mu^X(x)\\ &= \int \sup_{\|h\|\leq\delta}[I_h^{R(1)|X}(x)]_{(\ell,\ell)}q^X(x)d\mu^X(x) < \infty\,.\end{aligned} \tag{S.13}$$

To show (S.13), we calculate the conditional information as

$$E\left[\left(\frac{\frac{\partial}{\partial h_\ell}C(h)}{C(h)} + \frac{k'(h'g^{R(1)|X}(R(1)|X))}{k(h'g^{R(1)|X}(R(1)|X))}g_\ell^{R(1)|X}(R(1)|X)\right)^2 \middle| X = x\right]\,.$$

Note that k' is bounded above and k is bounded below, $C(h)$ is continuously differentiable with $C(0) = 1$, and so is bounded for $\|h\| \leq \delta$. Therefore, an application of the Cauchy-Schwarz inequality implies the previous expectation is bounded by a constant plus a constant multiple of $[I_0^{R(1)|X}(x)]_{(\ell,\ell)}$. The desired conclusion in (S.13) then follows because $E[I_0^{R(1)|X}(X)] < \infty$, and the proof is complete. ■

For $t \in \mathbf{R}^{d_\theta}$, the log-likelihood ratio between $P_{t/\sqrt{n},g}$ and $P_0 = P$ is

$$L_{t,n}(g) = \frac{1}{n} \sum_{1 \le i \le n} \log \frac{q^X_{t/\sqrt{n},g}(X_i)}{q^X(X_i)} + \frac{1}{n} \sum_{1 \le i \le n} \sum_{a \in \{0,1\}} I\{A_i = a\} \log \frac{q^{R(a)|X}_{t/\sqrt{n},g}(R_i|X_i)}{q^{R(a)|X}(R_i|X_i)} .$$

The following lemma establishes an expansion of the log-likelihood ratio and local asymptotic normality of $\{P_{t/\sqrt{n},g}\}$.

LEMMA A.3. *Suppose the treatment assignment mechanism satisfies Assumption 2.1 and the path satisfies* (S.9)–(S.10) *for g satisfying Condition A.1. Then,*

$$L_{t,n}(g) = \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} t' s_g(X_i, A_i, R_i) - \frac{1}{2} t' I^X t - \frac{1}{2n} \sum_{1 \le i \le n} \sum_{a \in \{0,1\}} I\{A_i = a\} t' I^{R(a)|X}(X_i) t + o_P(1) ,$$

where

$$s_g(x, a, r) = g^X(x) + I\{a = 1\} g^{R(1)|X}(r|x) + I\{a = 0\} g^{R(0)|X}(r|x) \tag{S.14}$$

and $I = I^X + \eta E_Q[I^{R(1)|X}(X_i)] + (1-\eta) E_Q[I^{R(0)|X}(X_i)]$. If in addition the assignment mechanism satisfies Assumption 4.1, then, under P_0,

$$L_{t,n}(g) \xrightarrow{d} N\Big(-\frac{1}{2} t' I t, t' I t\Big) ,$$

PROOF. The first result follows from Theorem 3.1 of [1]. The second result follows from Lemma A.6 given Assumption 4.1 and the assumption that each component of $I^{R(a)|X}(x)$ is integrable, noting that $E[s_g(X, 1, R(1)) - s_g(X, 0, R(0))|X] = 0$. ■

We emphasize that Lemma A.7 implies

$$\sum_{1 \le i \le n} s_g(X_i, A_i, R_i)$$

is the sum of n identically distributed, although possibly dependent, random variables. Therefore, in what follows, quantities like $E_P[s_g]$ are well defined.

Let $\theta(P) \in \mathbf{R}^{d_\theta}$ be a parameter of interest. Further suppose that there exists a $d_\theta \times 1$ vector of functions $\psi^* \in L^2(P)$ such that for each g satisfying Condition A.1, for all $t \in \mathbf{R}^{d_\theta}$, as $n \to \infty$,

$$\sqrt{n}(\theta(P_{t/\sqrt{n},g}) - \theta(P)) \to E_P[\psi^* s_g' t] . \tag{S.15}$$

In Lemma A.5 below, we provide explicit conditions which guarantee this is possible when $\theta(P)$ is defined by (3).

We call an estimator $\tilde{\theta}_n$ for $\theta(P)$ regular if for all g satisfying Condition A.1 and $t \in \mathbf{R}^{d_\theta}$,

$$\sqrt{n}(\tilde{\theta}_n - \theta(P_{t/\sqrt{n},g})) \xrightarrow{P_{t/\sqrt{n},g}} L \tag{S.16}$$

for a fixed probability measure L.

The following lemma establishes a convolution theorem for regular estimators:

LEMMA A.4. *Suppose θ satisfies (S.15). Let $\tilde{\theta}_n$ be a regular estimator for θ. Further suppose that $\psi^* = s_g$ for some function g satisfying Condition A.1. Then,*

$$L = N(0, E_P[\psi^* \psi^{*\prime}]) * B \ ,$$

where B is a fixed probability measure.

PROOF. In what follows, for each g satisfying Condition A.1, we consider the linear subspace given by

$$\mathcal{M}_g = \{t' s_g : t \in \mathbf{R}^{d_\theta}\} \ .$$

Note that $t' s_g$ appears in the expansion of the log-likelihood ratio between $P_{t/\sqrt{n},g}$ and P. To align our setting with Theorem 3.11.2 in [15], we first characterize the adjoint map (viewed as a mapping into $\mathcal{M}_g$) of the function $v \mapsto E[\psi^* v] \in \mathbf{R}^{d_\theta}$, where $v \in \mathcal{M}_g$. To that end, implicitly identifying each $b \in \mathbf{R}^{d_\theta}$ with the functional $b^* : \mathbf{R}^{d_\theta} \to \mathbf{R}$ given by $x \mapsto b'x$, we construct a $w(b) \in \mathbf{R}^{d_\theta}$ such that

$$b' E_P[\psi^* s_g' t] = E_P[w(b)' s_g s_g' t]$$

for all $t \in \mathbf{R}^{d_\theta}$, where we note $w(b)' s_g \in \mathcal{M}_g$, and is thus the output of the adjoint map when applied to the functional b^*. Because $E[s_g s_g']$ is invertible, we immediately obtain

$$w(b) = E[s_g s_g']^{-1} E[s_g \psi^{*\prime}] b \ .$$

Here we use the assumption that I is nonsingular. It then follows from the local asymptotic normality established in Lemma A.3 and Theorem 3.11.2 in [15] that

$$L = N(0, V_g) * B_g \ ,$$

where B_g is a fixed probability measure and $b' V_g b = E[(w(b)' s_g)^2]$, so that we have

$$V_g = E_P[\psi^* s_g'] E_P[s_g s_g']^{-1} E[s_g \psi^{*\prime}] \ .$$

Furthermore, by a standard projection argument, in particular the fact that the second moment of $\psi^* - E_P[\psi^* s_g'] E_P[s_g s_g']^{-1} s_g$ is positive semi-definite, it can be shown that V_g is maximized in the matrix sense when $s_g = \psi^*$. Note this maximum is attained by our assumption that $\psi^* = s_g$ for some g satisfying Condition A.1. The conclusion then follows. ■

To apply Lemma A.4 to the setting in Section 4.1, we we establish the form of ψ^* in (S.15) for the parameter $\theta_0 = \theta(P)$ defined by (3). Define $\eta(X_i) = P\{A_i = 1 | X_i\}$. Note that

$$\begin{aligned} 0 &= E_P[m(X_i, A_i, R_i, \theta(P))] \\ &= E_Q[m(X, 1, R(1), \theta(P)) \eta(X)] + E_Q[m(X, 0, R(0), \theta(P))(1 - \eta(X))] \ . \end{aligned} \tag{S.17}$$

LEMMA A.5. *Suppose the treatment assignment mechanism satisfies Assumptions 2.1 and 4.1. Fix a function g that satisfies Condition A.1. Suppose (S.9)–(S.10) holds. Fix $t \in \mathbf{R}^{d_\theta}$ and consider a one-dimensional submodel $\{P_{t/\sqrt{n},g}\}$ such that*

$$E_{Q_{t/\sqrt{n}}}[m(X, a, R(a), \theta(P))^2] = O(1) \tag{S.18}$$

$$E_{Q^X}[E_{Q^{R(a)|X}_{t/\sqrt{n}}}[m(X, a, R(a), \theta(P))^2 | X]] = O(1) \tag{S.19}$$

$$E_{Q^X_{t/\sqrt{n}}}[E_{Q^{R(a)|X}}[m(X, a, R(a), \theta(P))^2 | X]] = O(1) \tag{S.20}$$

as $n \to \infty$ and $\theta(P_{t/\sqrt{n},g})$ is uniquely determined by (S.17). Then, $\theta(P_{t/\sqrt{n},g})$ defined by (S.17) satisfies

$$\begin{aligned}
&\sqrt{n}(\theta(P_{t/\sqrt{n},g}) - \theta(P)) \\
&\to M^{-1} E_P[m(X_i, A_i, R_i, \theta(P))(g^X(X_i) \\
&\qquad + I\{A_i = 1\} g^{R(1)|X}(R_i|X_i) + I\{A_i = 0\} g^{R(0)|X}(R_i|X_i))']t \\
&= E_P[\psi^*(X_i, A_i, R_i, \theta(P))(g^X(X_i) \\
&\qquad + I\{A_i = 1\} g^{R(1)|X}(R_i|X_i) + I\{A_i = 0\} g^{R(0)|X}(R_i|X_i))']t \,,
\end{aligned}$$

where

$$\begin{aligned}
&\psi^*(X_i, A_i, R_i, \theta(P)) \\
&= -M^{-1}\Big(\eta(X_i) E_Q[m(X_i, 1, R_i(1), \theta(P))|X_i] \\
&\qquad + (1 - \eta(X_i)) E_Q[m(X_i, 0, R_i(0), \theta(P))|X_i] \\
&\qquad + I\{A_i = 1\}(m(X_i, 1, R_i, \theta(P)) - E_Q[m(X_i, 1, R_i(1), \theta(P))|X_i]) \\
&\qquad + I\{A_i = 0\}(m(X_i, 0, R_i, \theta(P)) - E_Q[m(X_i, 0, R_i(0), \theta(P))|X_i])\Big) \,.
\end{aligned}$$

PROOF. In what follows, we only use the property that the quadratic mean derivative of $P_{t/\sqrt{n},g}$ is given by $s_g' t$. Therefore, for ease of notation we consider a generic one-dimensional submodel $\{P_\nu : \nu \in [-\epsilon, \epsilon]\}$ that satisfies (S.9)–(S.10) for some $g = (g^X, g^{R(1)|X}, g^{R(0)|X})$, each component of which is a one-dimensional function. (S.17) implies

$$\begin{aligned}
0 = &\int m(x, 1, r, \theta(P_\nu)) q_\nu^{R(1)|X}(r|x) d\mu^R(r) \eta(x) q_\nu^X(x) d\mu^X(x) \\
&\qquad + \int m(x, 0, r, \theta(P_\nu)) q_\nu^{R(0)|X}(r|x) d\mu^R(r)(1 - \eta(x)) q_\nu^X(x) d\mu^X(x)
\end{aligned}$$

Note that

$$\begin{aligned}
&\int m(x, 1, r, \theta(P)) q_\nu^{R(1)|X}(r|x) d\mu^R(r) \eta(x) q_\nu^X(x) d\mu^X(x) \\
&\qquad - \int m(x, 1, r, \theta(P)) q^{R(1)|X}(r|x) d\mu^R(r) \eta(x) q^X(x) d\mu^X(x) \\
&= \gamma_1(\nu) + \gamma_2(\nu) + \gamma_3(\nu) + \gamma_4(\nu) \,,
\end{aligned}$$

where

$$\begin{aligned}
\gamma_1(\nu) = &\int m(x, 1, r, \theta(P)) \big(q_\nu^{R(1)|X}(r|x)^{1/2} - q^{R(1)|X}(r|x)^{1/2}\big) q_\nu^{R(1)|X}(r|x)^{1/2} d\mu^R(r) \\
&\qquad \times \eta(x) q_\nu^X(x) d\mu^X(x) \\
\gamma_2(\nu) = &\int m(x, 1, r, \theta(P)) \big(q_\nu^{R(1)|X}(r|x)^{1/2} - q^{R(1)|X}(r|x)^{1/2}\big) q^{R(1)|X}(r|x)^{1/2} d\mu^R(r) \\
&\qquad \times \eta(x) q_\nu^X(x) d\mu^X(x) \\
\gamma_3(\nu) = &\int m(x, 1, r, \theta(P)) q^{R(1)|X}(r|x) d\mu^R(r)
\end{aligned}$$

$$\times \eta(x)\big(q_\nu^X(x)^{1/2} - q^X(x)^{1/2}\big)q_\nu^X(x)^{1/2}d\mu^X(x)$$

$$\gamma_4(\nu) = \int m(x,1,r,\theta(P))q^{R(1)|X}(r|x)d\mu^R(r) \times \eta(x)\big(q_\nu^X(x)^{1/2} - q^X(x)^{1/2}\big)q^X(x)^{1/2}d\mu^X(x)\,.$$

It follows from the Cauchy-Schwarz inequality that

$$\begin{aligned}
&\frac{1}{\nu}\gamma_4(\nu) - \int m(x,1,r,\theta(P))q^{R(1)|X}(r|x)d\mu^R(r) \\
&\qquad\qquad \times \eta(x)\frac{1}{2}g^X(x)q^X(x)^{1/2} \times q^X(x)^{1/2}d\mu^X(x) \\
&\le \int \Big(m(x,1,r,\theta(P))^2 q^{R(1)|X}(r|x)d\mu^R(r)\eta(x)^2 q^X(x)d\mu^X(x)\Big)^{1/2} \\
&\times \Big(\int q^{R(1)|X}(r|x)d\mu^R(r) \\
&\qquad\qquad \times \Big(\frac{1}{\nu}(q_\nu^X(x)^{1/2} - q^X(x)^{1/2}) - \frac{1}{2}g^X(x)q^X(x)^{1/2}\Big)^2 d\mu^X(x)\Big)^{1/2} \\
&\to 0
\end{aligned}$$

by the assumption that $E_P[m(X,a,R(a),\theta(P))^2] < \infty$, the facts that $0 \le \eta(x) \le 1$, $\int q^{R(1)|X}(r|x)d\mu^R(r) = 1$, and (S.9). Similar arguments implies as $\nu \to 0$,

$$\frac{1}{\nu}\gamma_1(\nu) - \int m(x,1,r,\theta(P))\frac{1}{2}g^{R(1)|X}(r|x)q^{R(1)|X}(r|x)d\mu^R(r)\eta(x)q^X(x)d\mu^X(x) \to 0$$

because $E_{P_\nu}[m(X,a,R(a),\theta(P))^2] = O(1)$ as $\nu \to 0$. The limits of $\gamma_2(\nu)$ and $\gamma_3(\nu)$ can be derived following similar arguments using (S.19)–(S.20). Combining all previous results yields

$$\begin{aligned}
&\frac{\partial}{\partial \nu}E_{P_\nu}[m(X,A,R,\theta(P))]\Big|_{\nu=0} \\
&= E_Q[m(X,1,R(1),\theta(P))(g^X(X) + g^{R(1)|X}(R|X))\eta(X)] \\
&\qquad + E_Q[m(X,0,R(0),\theta(P))(g^X(X) + g^{R(0)|X}(R|X))(1-\eta(X))] \\
&= E_P[m(X,A,R,\theta(P))(g^X(X) + I\{A=1\}g^{R(1)|X}(R) + I\{A=0\}g^{R(0)|X}(R))]\,.
\end{aligned}$$

On the other hand, by definition

$$M = \frac{\partial}{\partial \theta'}E_P[m(X,A,R,\theta)]\Big|_{\theta=\theta(P)}\,.$$

The formula for the derivative therefore follows from the implicit function theorem (in particular, because we have assumed the existence of $\theta(P_\nu)$ along the path, it follows from the last part of the proof of Theorem 3.2.1 in [10]). The second equality follows from Lemma A.7 together with Condition A.1. ■

Finally, to prove Theorem 4.1 we require the following additional regularity condition:

ASSUMPTION A.1. For every function g satisfying Condition A.1 and every $t \in \mathbf{R}^{d_\theta}$ there exists a submodel $P_{t/\sqrt{n},g}$ for which (S.18)–(S.20) hold as $n \to \infty$, and $\theta(P_{t/\sqrt{n},g})$ is uniquely determined by (S.17).

This assumption guarantees that every element satisfying Condition A.1 has a corresponding path for which we can apply Lemma A.5. A similar assumption appears in [5] (see their Assumption 4.1(iv)). Note that a simple sufficient condition for the first part of Assumption A.1 is that $m(x,a,r,\theta(P))$ is a bounded function in (x,r) on the support of $(X,R(a))$. The second part of Assumption A.1 can be verified easily in specific examples (see, for instance, Examples 2.1–2.5 in the main text). Alternatively, Assumption A.1 could be avoided by assuming that we can differentiate under the integral in the final step of the proof of Lemma A.5, from which we would immediately obtain the expression for the pathwise derivative. See, for instance, [13] and [4].

PROOF OF THEOREM 4.1. First note θ satisfies (S.15) because of Lemma A.5 and Assumption A.1. The result then follows from Lemma A.4 upon noting that $\psi^* = s_g$ for some g that satisfies Condition A.1. ■

To study regular estimators, we need the following lemma.

LEMMA A.6. *Suppose the treatment assignment mechanism satisfies Assumptions 2.1, 4.1, and 4.2. Let $f(x,a,r)$ be a vector-valued function such that $E[f(X,A,R)]=0$ and $E[f^2(X,a,R(a))]<\infty$ for $a\in\{0,1\}$. Then,*

$$\frac{1}{\sqrt{n}}\sum_{1\le i\le n} f(X_i,A_i,R_i) \xrightarrow{d} N(0,V_f)\ ,$$

where $V_f = V_{1,f}+V_{2,f}+V_{3,f}$ for

$$V_{1,f} = \mathrm{Var}[E[f(X,A,R)|X]] = \mathrm{Var}[\eta E[f(X,1,R(1))|X] + (1-\eta)E[f(X,0,R(0))|X]]$$

$$V_{2,f} = V^{\mathrm{imb}}_{E[f(X,1,R(1))-f(X,0,R(0))|X]}$$

$$V_{3,f} = E[\eta\,\mathrm{Var}[f(X,1,R(1))|X] + (1-\eta)\,\mathrm{Var}[f(X,0,R(0))|X]]\ .$$

PROOF. Note that

$$\mathbb{C}_n := \frac{1}{\sqrt{n}}\sum_{1\le i\le n} f(X_i,A_i,R_i) = \mathbb{C}_{1,n}+\mathbb{C}_{2,n}+\mathbb{C}_{3,n}\ ,$$

where

$$\begin{aligned}
\mathbb{C}_{1,n} &= \frac{1}{\sqrt{n}}\sum_{1\le i\le n} E[f(X_i,A_i,R_i)|X_i]\\
&= \frac{1}{\sqrt{n}}\sum_{1\le i\le n} (\eta E[f(X_i,1,R_i(1))|X_i] + (1-\eta)E[f(X_i,0,R_i(0))|X_i])\\
\mathbb{C}_{2,n} &= \frac{1}{\sqrt{n}}\sum_{1\le i\le n} (A_i-\eta)E[f(X_i,1,R_i(1))-f(X_i,0,R_i(0))|X_i]\\
\mathbb{C}_{3,n} &= \frac{1}{\sqrt{n}}\sum_{1\le i\le n} \big(A_i(f(X_i,1,R_i(1)) - E[f(X_i,1,R_i(1))|X_i])\\
&\qquad + (1-A_i)(f(X_i,0,R_i(0)) - E[f(X_i,0,R_i(0))|X_i])\big)\ .
\end{aligned}$$

Note that $\mathbb{C}_{1,n}$ has zero mean because $E[E[f(X_i,A_i,R_i)|X_i]] = E[f(X_i,A_i,R_i)] = 0$. Further note that $\mathbb{C}_{1,n} = E[\mathbb{C}_n|X^{(n)}]$, $\mathbb{C}_{2,n} = E[\mathbb{C}_n|X^{(n)},A^{(n)}] - E[\mathbb{C}_n|X^{(n)}]$, and $\mathbb{C}_{3,n} =$

$\mathbb{C}_n - E[\mathbb{C}_n|X^{(n)}, A^{(n)}]$. It follows from the central limit theorem and $E[f^2(X, a, R(a))] < \infty$ that

$$\mathbb{C}_{1,n} \xrightarrow{d} N(0, V_{1,f}) .$$

Next, it follows from Assumption 4.2 that

$$\rho(\mathcal{L}(\mathbb{C}_{2,n}|X^{(n)}), N(0, V_{2,f})) \xrightarrow{P} 0 .$$

For $\mathbb{C}_{3,n}$, it follows from Assumption 4.1 that

$$\mathrm{Var}[\mathbb{C}_{3,n}|X^{(n)}, A^{(n)}] \xrightarrow{P} V_{3,f} .$$

As a result, one can verify the Lindeberg condition conditional on $X^{(n)}$ and $A^{(n)}$ as in the proof of Lemma S.1.4 of [3], and obtain

$$\rho(\mathcal{L}(\mathbb{C}_{3,n}|X^{(n)}, A^{(n)}), N(0, V_{3,f})) \xrightarrow{P} 0 .$$

The proof can then be completed by the subsequencing argument at the end of the proof of Lemma S.1.4 of [3]. ■

PROOF OF THEOREM 4.2. Recall from Lemma A.3 that $L_{t,n}(g)$ is also asymptotically linear. Because the treatment assignment mechanism satisfies Assumptions 2.1, 4.1, and 4.2, the path satisfies Condition A.1, and hence ψ and s_g jointly satisfy the conditions in Lemma A.6 (in particular, note that $E[s_g(X, 1, R(1)) - s_g(X, 0, R(0))|X] = 0$),

$$\begin{pmatrix} \sqrt{n}(\tilde{\theta}_n - \theta(P)) \\ L_{t,n}(g) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} \psi(X_i, A_i, R_i, \theta(P)) \\ \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} t' s_g(X_i, A_i, R_i) \end{pmatrix} - \begin{pmatrix} 0 \\ -\frac{1}{2} t' I t \end{pmatrix} + o_P(1)$$

are jointly asymptotically normal. Because $E[s_g(X, 1, R(1)) - s_g(X, 0, R(0))|X] = 0$, the covariance in the second term in the joint variance in Lemma A.6 vanishes, and thus the overall covariance is given by

$$\begin{aligned} &E[E[\psi|X] g^X(X)']t \\ &\qquad + \eta E[(\psi(X, 1, R(1), \theta(P)) - E[\psi(X, 1, R(1), \theta(P))|X]) g^{R(1)|X}(R(1)|X)']t \\ &\qquad + (1 - \eta) E[(\psi(X, 0, R(0), \theta(P)) - E[\psi(X, 0, R(0), \theta(P))|X]) g^{R(0)|X}(R(0)|X)']t \\ &= E[\psi s_g']t . \end{aligned}$$

It then follows from Le Cam's third lemma that under $P_{t/\sqrt{n}, g}$, $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ converges in distribution to a normal distribution with mean $E[\psi s_g']t$ and the same variance as in the limit under P. At the same time,

$$\sqrt{n}(\tilde{\theta}_n - \theta(P_{t/\sqrt{n}, g})) = \sqrt{n}(\tilde{\theta}_n - \theta(P)) - \sqrt{n}(\theta(P_{t/\sqrt{n}, g} - \theta(P)) .$$

Therefore, (S.16) holds if and only if

$$E[\psi s_g']t = E[\psi^* s_g']t .$$

Furthermore, $\tilde{\theta}_n$ is regular if the equality holds for all t and all g satisfying Condition A.1, if and only if

$$E[(\psi - \psi^*) s_g'] = 0$$

for all g satisfying Condition A.1. Note that

$$\begin{aligned}
&E[(\psi-\psi^*)s_g']\\
&=E[E[\psi(X,A,R,\theta(P))-\psi^*(X,A,R,\theta(P))|X]g^X(X)']\\
&\qquad+E[A_i(\psi(X,1,R(1),\theta(P))-\psi^*(X,1,R(1),\theta(P))\\
&\qquad\qquad-E[\psi(X,1,R(1),\theta(P))-\psi^*(X,1,R(1),\theta(P))|X])g^{R(1)|X}(R(1)|X)']\\
&\qquad+E[(1-A_i)(\psi(X,0,R(0),\theta(P))-\psi^*(X,0,R(0),\theta(P))\\
&\qquad\qquad-E[\psi(X,0,R(0),\theta(P))-\psi^*(X,0,R(0),\theta(P))|X])g^{R(0)|X}(R(0)|X)']\ .
\end{aligned}$$

By setting $g^{R(1)|X}=0$, $g^{R(0)|X}=0$, and $g^X(X)=E[\psi(X,A,R,\theta(P))-\psi^*(X,A,R,\theta(P))|X]$, we get that

$$E[\psi(X,A,R,\theta(P))|X]=E[\psi^*(X,A,R,\theta(P))|X]\ . \tag{S.21}$$

Setting $g^X(X)=0=g^{R(0)|X}$ and $g^{R(1)|X}=\psi(X,1,R(1),\theta(P))-\psi^*(X,1,R(1),\theta(P))-E[\psi(X,1,R(1),\theta(P))-\psi^*(X,1,R(1),\theta(P))|X]$, we get

$$\begin{aligned}
&\psi(X,1,R(1),\theta(P))-\psi^*(X,1,R(1),\theta(P))\\
&\qquad-E[\psi(X,1,R(1),\theta(P))-\psi^*(X,1,R(1),\theta(P))|X]=0\ ,
\end{aligned}$$

which implies that $\psi(X,1,R(1),\theta(P))-\psi^*(X,1,R(1),\theta(P))$ can only be a function of X. Denote it by $\psi^\perp(X,1)$. Similarly, $\psi(X,0,R(0),\theta(P))-\psi^*(X,0,R(0),\theta(P))$ can only be a function of X. Denote it by $\psi^\perp(X,0)$. We have

$$\begin{aligned}
\psi(X,A,R,\theta(P))&=A\psi(X,1,R(1),\theta(P))+(1-A)\psi(X,0,R(0),\theta(P))\\
&=A(\psi^*(X,1,R(1),\theta(P))+\psi^\perp(X,1,\theta(P)))\\
&\qquad+(1-A)(\psi^*(X,0,R(0),\theta(P))+\psi^\perp(X,0,\theta(P)))\\
&=\psi^*(X,A,R,\theta(P))+\psi^\perp(X,A,\theta(P))\ ,
\end{aligned}$$

and it follows from (S.21) that $E[\psi^\perp(X,A,\theta(P))|X]=0$. Finally, to show that $\tilde\theta_n$ is efficient if and only if (23) holds, we apply Lemma A.6 with $f=\psi$. If (23) holds, then $V_\psi=V_{1,\psi}+V_{3,\psi}=V_{\psi^*}$ because $V_{1,\psi}=V_{1,\psi^*}$, $V_{3,\psi}=V_{3,\psi^*}$ and $V_{2,\psi^*}=0$. On the other hand, if $\tilde\theta_n$ is efficient, then $V_{2,\psi}=0$, so (23) holds by Markov's inequality combined with the fact that probabilities are bounded and thus uniformly integrable. ■

A.4. Auxiliary Lemmas.

LEMMA A.7. *Suppose* (2) *holds and* $\Pr\{A_i=1|X_i=x\}$ *as a function is identical across* $1\le i\le n$. *Then,*

$$(R_i(1),R_i(0))\perp\!\!\!\perp A_i|X_i\ . \tag{S.22}$$

Moreover, (X_i,A_i,R_i) *is identically distributed across* $1\le i\le n$.

PROOF. Fix $a\in\{0,1\}$ and any Borel sets $B\in\mathbf{R}^{d_r}\times\mathbf{R}^{d_r}$ and $C\in\mathbf{R}^{d_x}$.

$$\begin{aligned}
&E[\Pr\{(R_i(1),R_i(0))\in B,A_i=a|X_i\}I\{X_i\in C\}]\\
&=E[E[\Pr\{(R_i(1),R_i(0))\in B,A_i=a|X^{(n)}\}|X_i]I\{X_i\in C\}]\\
&=E[E[\Pr\{(R_i(1),R_i(0))\in B|X^{(n)}\}\Pr\{A_i=a|X^{(n)}\}|X_i]I\{X_i\in C\}]\\
&=E[\Pr\{(R_i(1),R_i(0))\in B|X_i\}\Pr\{A_i=a|X_i\}I\{X_i\in C\}]\ ,
\end{aligned}$$

where the first equality follows from the law of iterated expectations, the second equality follows from (2), the third equality follows from the law of iterated expectations as well as the fact that $Q_n = Q^n$. The first statement of the lemma then follows from the definition of a conditional expectation.

To prove the second statement, fix units i and i'. Clearly X_i and $X_{i'}$ are identically distributed. Conditional on X_i, for any Borel set $C \in \mathbf{R}^{d_r}$ and $a \in \{0,1\}$, it follows (a) that

$$\Pr\{R_i \in C, A_i = a | X_i\} = \Pr\{A_i = a|X_i\}\Pr\{R_i(a) \in C|X_i\} .$$

The conclusion then follows because $\Pr\{A_i = 1|X_i = x\}$ is identical across $1 \le i \le n$ and $Q_n = Q^n$. ■

LEMMA A.8. *Suppose the treatment assignment mechanism satisfies Assumptions 3.1–3.2 and the moment functions satisfy Assumption 3.3. Then,* $\hat{\theta}_n \xrightarrow{P} \theta_0$.

PROOF. It follows from Assumption 3.3(a) and Theorem 5.9 in [14] that we only need to establish for each $1 \le s \le d_\theta$,

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{1 \le i \le n} (m_s(X_i, A_i, R_i, \theta) - E_P[m_s(X_i, A_i, R_i, \theta)]) \right| \xrightarrow{P} 0 . \tag{S.23}$$

To begin, note it follows from Assumption 3.3(d) and the dominated convergence theorem that if $m_s(x,a,r,\theta_m) \to m_s(x,a,r,\theta)$ as $m \to \infty$ for $\{\theta_m\} \subset \Theta^*$, then $E_P[m_s(X_i, A_i, R_i, \theta_m)] \to E_P[m_s(X_i, A_i, R_i, \theta)]$. Here, the dominating function exists by Problem 2.4.1 in [15]. Assumption 3.3(c) then implies

$$\begin{aligned} \sup_{\theta \in \Theta} & \left| \frac{1}{n} \sum_{1 \le i \le n} (m_s(X_i, A_i, R_i, \theta) - E_P[m_s(X_i, A_i, R_i, \theta)]) \right| \\ &= \sup_{\theta \in \Theta^*} \left| \frac{1}{n} \sum_{1 \le i \le n} (m_s(X_i, A_i, R_i, \theta) - E_P[m_s(X_i, A_i, R_i, \theta)]) \right| , \end{aligned} \tag{S.24}$$

which is measurable. Next, note that

$$m(X_i, A_i, R_i, \theta) = A_i m(X_i, 1, R_i(1), \theta) + (1 - A_i) m(X_i, 0, R_i(0), \theta) . \tag{S.25}$$

and it follows from Lemma A.7 that

$$E_P[m(X_i, A_i, R_i, \theta)] = \eta E_Q[m(X_i, 1, R_i(1), \theta)] + (1 - \eta) E_Q[m(X_i, 0, R_i(0), \theta)] , \tag{S.26}$$

from which we obtain that

$$\begin{aligned} & E\left[\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{1 \le i \le n} (m_s(X_i, A_i, R_i, \theta) - E_P[m_s(X_i, A_i, R_i, \theta)]) \right| \right] \\ &= E\left[\sup_{\theta \in \Theta^*} \left| \frac{1}{n} \sum_{1 \le i \le n} (m_s(X_i, A_i, R_i, \theta) - E_P[m_s(X_i, A_i, R_i, \theta)]) \right| \right] \\ &\le E\left[\sup_{\theta \in \Theta^*} \left| \frac{1}{n} \sum_{1 \le i \le n} A_i (m_s(X_i, 1, R_i(1), \theta) - \eta E[m_s(X_i, 1, R_i(1), \theta)]) \right| \right] \\ &\quad + E\left[\sup_{\theta \in \Theta^*} \left| \frac{1}{n} \sum_{1 \le i \le n} (1 - A_i)(m_s(X_i, 0, R_i(0), \theta) - (1 - \eta) E[m_s(X_i, 0, R_i(0), \theta)]) \right| \right] \end{aligned}$$

$$= E\Bigg[\sup_{\theta\in\Theta^*}\Bigg|\frac{1}{n}\sum_{1\le i\le n}A_i(m_s(X_i,1,R_i(1),\theta)-E[m_s(X_i,1,R_i(1),\theta)])\Bigg|\Bigg]$$

$$+ E\Bigg[\sup_{\theta\in\Theta^*}\Bigg|\frac{1}{n}\sum_{1\le i\le n}(1-A_i)(m_s(X_i,0,R_i(0),\theta)-E[m_s(X_i,0,R_i(0),\theta)])\Bigg|\Bigg]$$

$$\lesssim E\Bigg[\sup_{\theta\in\Theta^*}\Bigg|\frac{1}{n}\sum_{1\le i\le n}(m_s(X_i,1,R_i(1),\theta)-E[m_s(X_i,1,R_i(1),\theta)])\Bigg|\Bigg]$$

$$+ E\Bigg[\sup_{\theta\in\Theta^*}\Bigg|\frac{1}{n}\sum_{1\le i\le n}(m_s(X_i,0,R_i(0),\theta)-E[m_s(X_i,0,R_i(0),\theta)])\Bigg|\Bigg] \to 0\,.$$

where the first equality follows from (S.24), the first inequality follows from the triangle inequality, the second equality follows because $\sum_{1\le i\le n}A_i = n\eta$ and $\sum_{1\le i\le n}(1-A_i) = n(1-\eta)$, and the last inequality follows from Proposition C.1 in [7]. The convergence follows from Assumption 3.3(e) and an application of the backward submartingale convergence theorem [see, for instance, Theorem 12.30 in 11], as detailed in the proof of Theorem 3.1 in [7]. The desired result in (S.23) then follows by Markov's inequality. ■

LEMMA A.9. *Suppose the treatment assignment mechanism satisfies Assumptions 3.1, 3.2, and 3.4, and the moment functions satisfy Assumptions 3.3 and 3.5. Then, for each deterministic sequence $\theta_n \to \theta_0$, $\hat\mu^{(s)}_{1,n}(\theta_n) \xrightarrow{P} \mu^{(s)}_1$ for $\mu^{(s)}_1$ in (S.5) and $\hat\mu^{(s)}_{1,n}(\theta_n)$ in (S.6).*

PROOF. The conclusion could follow from similar arguments to those in the proof of Lemma A.8 but we provide a different proof because it will apply generally to other components of the variance estimator. It suffices to prove that

$$\hat\mu^{(s)}_{1,n}(\theta_n) - E[\hat\mu^{(s)}_{1,n}(\theta_n)|X^{(n)},A^{(n)}] \xrightarrow{P} 0 \tag{S.27}$$

$$E[\hat\mu^{(s)}_{1,n}(\theta_n)|X^{(n)},A^{(n)}] \xrightarrow{P} \mu^{(s)}_1\,. \tag{S.28}$$

The desired result in (S.27) follows from similar arguments to those in the proof of equation (S.27) in [3], where the uniform integrability condition is replaced by Assumption 3.5(a). To show (S.28), note

$$E[\hat\mu^{(s)}_{1,n}(\theta_n)|X^{(n)},A^{(n)}] - \mu^{(s)}_1$$

$$= \frac{1}{\eta n}\sum_{1\le i\le n}A_iE[m_s(X_i,1,R_i(1),\theta_n)|X_i] - E[m_s(X_i,1,R_i(1),\theta_0)]$$

$$= \frac{1}{\eta n}\sum_{1\le i\le n}A_iE[m_s(X_i,1,R_i(1),\theta_n)|X_i] - \frac{1}{n}\sum_{1\le i\le n}E[m_s(X_i,1,R_i(1),\theta_n)|X_i] \tag{S.29}$$

$$+ \frac{1}{n}\sum_{1\le i\le n}E[m_s(X_i,1,R_i(1),\theta_n)|X_i] - E[m_s(X_i,1,R_i(1),\theta_0)]\,. \tag{S.30}$$

The difference in (S.30) converges in probability to zero by Assumption 3.5(b) and because $E[m(X_i,1,R_i(1),\theta_n)] \to E[m(X_i,1,R_i(1),\theta_0)]$ from $\theta_n \to \theta_0$ and Assumption 3.3(c). Next, note the absolute value of (S.29) can be written as

$$\Bigg|\frac{k}{n}\sum_{1\le j\le n/k}\frac{1}{\eta k}\sum_{i\in\lambda_j}A_i\Big(E[m_s(X_i,1,R_i(1),\theta_n)|X_i]$$

$$- \frac{1}{k} \sum_{i' \in \lambda_j} E[m_s(X_{i'}, 1, R_{i'}(1), \theta_n)|X_{i'}]\Bigg)\Bigg|$$

$$\leq \frac{k}{n} \sum_{1 \leq j \leq n/k} \frac{1}{\eta k} \sum_{i \in \lambda_j} A_i \Big| E[m_s(X_i, 1, R_i(1), \theta_n)|X_i] - E[m_s(X_{i'}, 1, R_{i'}(1), \theta_n)|X_{i'}] \Big|$$

$$\lesssim \frac{1}{n} \sum_{1 \leq j \leq n/k} \max_{i, i' \in \lambda_j} \|X_i - X_{i'}\|^2 \xrightarrow{P} 0 \,,$$

where the inequality follows from the uniform Lipschitzness condition in Assumption 3.5(c) and the convergencef follows from Assumption 3.2. Therefore, (S.27) holds. ■

APPENDIX B: DETAILS FOR SIMULATIONS

B.1. Local Average Treatment Effect. In this section, we present the model specifications and estimators for estimating the LATE as in Example 2.3. Recall that in this case the moment condition we consider is given by

$$m(X_i, A_i, R_i, \theta) = \frac{Y_i A_i}{\eta} - \frac{Y_i(1 - A_i)}{1 - \eta} - \theta \left(\frac{D_i A_i}{\eta} - \frac{D_i(1 - A_i)}{1 - \eta} \right) ,$$

with $R_i = (Y_i, D_i)$. The outcome is determined by the relationship $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$, where $Y_i(d)$ is again given by (24). In addition, the take-up decision is determined as $D_i = A_i D_i(1) + (1 - A_i) D_i(0)$, where

$$D_i(0) = I\left\{\alpha_0 + \alpha\left(X_i\right) > \varepsilon_{1,i}\right\} \,,$$

$$D_i(1) = \begin{cases} I\left\{\alpha_1 + \alpha\left(X_i\right) > \varepsilon_{2,i}\right\} & \text{if } D_i(0) = 0 \\ 1 & \text{otherwise} \end{cases} ,$$

where $\alpha_0 = 0.2$, $\alpha_1 = 4$, $\alpha(X_i) = \sum_{1 \leq l \leq 8} \left(X_{i,l} + \frac{1}{3}(X_{i,l}^2 - 1) \right)$, $\varepsilon_{1,i}, \varepsilon_{2,i} \sim N(0, 4)$, $(X_i, \epsilon_i, \varepsilon_{1,i}, \varepsilon_{2,i})$, $1 \leq i \leq n$ are i.i.d., and for each $1 \leq i \leq n$, $(X_i, \epsilon_i, \varepsilon_{1,i}, \varepsilon_{2,i})$ are independent.

We consider the following three estimators for the LATE:

Unadjusted Estimator:

$$\hat{\theta}_n^{\text{unadj}} = \frac{\sum_{1 \leq i \leq n} (Y_i A_i - Y_i(1 - A_i))}{\sum_{1 \leq i \leq n} (D_i A_i - D_i(1 - A_i))} \,.$$

Adjusted Estimator 1:

$$\hat{\theta}_n^{\text{adj,1}} = \frac{\sum_{1 \leq i \leq n} \left(2A_i(Y_i - \hat{\mu}_1^Y(X_i)) - 2(1 - A_i)(Y_i - \hat{\mu}_0^Y(X_i)) + \hat{\mu}_1^Y(X_i) - \hat{\mu}_0^Y(X_i) \right)}{\sum_{1 \leq i \leq n} \left(2A_i(D_i - \hat{\mu}_1^Y(X_i)) - 2(1 - A_i)(D_i - \hat{\mu}_0^Y(X_i)) + \hat{\mu}_1^D(X_i) - \hat{\mu}_0^D(X_i) \right)} \,,$$

where $\hat{\mu}_a^Y(X_i)$ is the linear projection of Y_i on $(1, (X_{i,l}, X_{i,l}^2 : 1 \leq l \leq T))$ in the subsample with $A_i = a$, and $\hat{\mu}_a^D(X_i)$ is estimated using from a logistic regression model with the same set of regressors in the subsample with $A_i = a$.

Adjusted Estimator 2: As in Adjusted Estimator 1, but in $\hat{\mu}_a^Y(X_i)$ and $\hat{\mu}_a^D(X_i)$ the regressors are instead $(1, (X_{i,l}, X_{i,l}^2, X_{i,l} 1\{X_{i,l} > \hat{t}_l\} : 1 \leq l \leq T))$, where $\hat{t}_l$ is the sample median of $X_{i,l}$, $1 \leq i \leq n$.

Similarly to Section 5.1, $\hat{\theta}_n^{\mathrm{unadj}}$ solves (6) for the moment condition given in (5). The second and third estimators are covariate adjusted estimators which can be obtained as two-step method of moments estimators from solving an "augmented" version of the moment condition (5) [see, for instance, 6, 8].

B.2. Additional Simulation Tables. Table S.1 displays the absolute value of the bias for each design and estimator pair computed across 4000 Monte Carlo replications. We find that all estimators have fairly low and comparable bias.

Table S.2 reports the coverage and average length of confidence intervals constructed using the estimator described in Section 3.1, again computed across 4000 Monte Carlo replications. We find that the confidence intervals have appropriate coverage at all sample sizes and for any number of covariates. The average length of the intervals follows a pattern similar to what was observed for the MSE, again because the first 4 covariates are much stronger predictors of the control outcome than the last 4 covariates, which are almost uninformative. Although we found in Table 1 that the MSE of the matched tuples design was worse than that of matched pairs with 8 covariates, this does not seem to translate to longer confidence intervals for matched tuples: this may be due to the fact that the matched tuples variance estimator is a "within block" estimator, as described in Section 3.1. Similar findings have been reported by [2] for related problems in factorial designs.

REFERENCES

[1] ARMSTRONG, T. B. (2022). Asymptotic Efficiency Bounds for a Class of Experimental Designs. arXiv:2205.02726 [stat]. https://doi.org/10.48550/arXiv.2205.02726

[2] BAI, Y., LIU, J. and TABORD-MEEHAN, M. (2024). Inference for Matched Tuples and Fully Blocked Factorial Designs. *Quantitative Economics* **15** 279–330.

[3] BAI, Y., ROMANO, J. P. and SHAIKH, A. M. (2022). Inference in Experiments With Matched Pairs. *Journal of the American Statistical Association* **117** 1726–1737. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2021.1883437. https://doi.org/10.1080/01621459.2021.1883437

[4] CHEN, X., HONG, H. and TAROZZI, A. (2008). Semiparametric Efficiency in GMM Models with Auxiliary Data. *The Annals of Statistics* **36** 808–843. Publisher: Institute of Mathematical Statistics.

[5] CHEN, X. and SANTOS, A. (2018). Overidentification in Regular Models. *Econometrica* **86** 1771–1817. Publisher: [Wiley, The Econometric Society].

[6] CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* **21** C1–C68. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ectj.12097. https://doi.org/10.1111/ectj.12097

[7] HAN, Q. and WELLNER, J. A. (2021). Complex sampling designs: Uniform limit theorems and applications. *The Annals of Statistics* **49** 459–485. Publisher: Institute of Mathematical Statistics. https://doi.org/10.1214/20-AOS1964

[8] JIANG, L., LINTON, O. B., TANG, H. and ZHANG, Y. (2022). Improving Estimation Efficiency via Regression-Adjustment in Covariate-Adaptive Randomizations with Imperfect Compliance. arXiv:2201.13004 [econ, stat]. https://doi.org/10.48550/arXiv.2201.13004

[9] KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. *Springer Series in Statistics*. Springer-Verlag, New York. https://doi.org/10.1007/978-0-387-74978-5

[10] KRANTZ, S. G. and PARKS, H. R. (2013). *The Implicit Function Theorem: History, Theory, and Applications*. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-5981-1

[11] LE GALL, J.-F. (2022). *Measure theory, probability, and stochastic processes*. Springer.

[12] LEHMANN, E. L. and ROMANO, J. P. (2022). *Testing Statistical Hypotheses*. *Springer Texts in Statistics*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-70578-7

[13] NEWEY, W. K. (1994). The Asymptotic Variance of Semiparametric Estimators. *Econometrica* **62** 1349–1382. Publisher: [Wiley, Econometric Society]. https://doi.org/10.2307/2951752

[14] VAN DER VAART, A. W. (1998). *Asymptotic statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511802256 MR1652247

[15] VAN DER VAART, A. W. and WELLNER, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. *Springer Series in Statistics*. Springer-Verlag, New York. https://doi.org/10.1007/978-1-4757-2545-2

TABLE S.1
Absolute value of bias of estimators for different designs and estimators

		i.i.d. assignment			**Matched pairs**	**Matched tuples**
	# of covariates (T)	Unadjusted	Adjusted 1	Adjusted 2	Unadjusted	Unadjusted
			$n = 100$			
ATE	2	0.0083	0.0065	0.0051	0.0212	0.0068
	4	0.0083	0.0090	0.0090	0.0073	0.0124
	8	0.0083	0.0134	0.0131	0.0010	0.0209
LATE	2	0.0905	0.0436	0.0396	0.0341	0.0076
	4	0.0905	0.0276	0.0264	0.0301	0.0427
	8	0.0905	0.0401	0.0376	0.0264	0.0297
			$n = 200$			
ATE	2	0.0075	0.0068	0.0074	0.0084	0.0081
	4	0.0075	0.0052	0.0067	0.0040	0.0040
	8	0.0075	0.0045	0.0055	0.0145	0.0051
LATE	2	0.0545	0.0326	0.0339	0.0361	0.0328
	4	0.0545	0.0182	0.0217	0.0123	0.0144
	8	0.0545	0.0159	0.0180	0.0447	0.0216
			$n = 400$			
ATE	2	0.0001	0.0011	0.0009	0.0041	0.0055
	4	0.0001	0.0001	0.0001	0.0001	0.0016
	8	0.0001	0.0006	0.0006	0.0056	0.0041
LATE	2	0.0142	0.0064	0.0059	0.0179	0.0198
	4	0.0142	0.0012	0.0015	0.0004	0.0020
	8	0.0142	0.0021	0.0019	0.0129	0.0064
			$n = 1000$			
ATE	2	0.0021	0.0037	0.0039	0.0014	0.0025
	4	0.0021	0.0021	0.0022	0.0020	0.0006
	8	0.0021	0.0021	0.0023	0.0023	0.0027
LATE	2	0.0118	0.0120	0.0123	0.0019	0.0083
	4	0.0118	0.0055	0.0057	0.0051	0.0007
	8	0.0118	0.0054	0.0058	0.0057	0.0059
			$n = 2000$			
ATE	2	0.0013	0.0010	0.0010	0.0011	0.0025
	4	0.0013	0.0015	0.0015	0.0003	0.0021
	8	0.0013	0.0015	0.0015	0.0001	0.0014
LATE	2	0.0062	0.0035	0.0035	0.0018	0.0050
	4	0.0062	0.0037	0.0038	0.0011	0.0055
	8	0.0062	0.0038	0.0037	0.0008	0.0032

Note: Entries report the absolute bias of each estimator under different sample sizes, designs, and covariate adjustment strategies. Results are averaged across 2000 replications.

TABLE S.2
Coverage rates and lengths of 95% confidence intervals based on unadjusted estimator for matched pairs and matched tuples

	# of covariates (T)	**Matched pairs**		**Matched tuples**	
		Coverage	CI length	Coverage	CI length
		$n = 100$			
ATE	2	0.9403	2.3191	0.9490	2.3224
	4	0.9465	2.2744	0.9485	2.2830
	8	0.9487	2.3532	0.9395	2.3672
LATE	2	0.9590	6.4075	0.9590	6.3865
	4	0.9620	6.3184	0.9615	6.3176
	8	0.9610	6.5497	0.9527	6.5311
		$n = 200$			
ATE	2	0.9495	1.6401	0.9497	1.6418
	4	0.9485	1.5990	0.9440	1.6049
	8	0.9503	1.6529	0.9425	1.6617
LATE	2	0.9550	4.4046	0.9563	4.4264
	4	0.9557	4.3315	0.9530	4.3381
	8	0.9597	4.4833	0.9503	4.5038
		$n = 400$			
ATE	2	0.9487	1.1602	0.9483	1.1614
	4	0.9497	1.1255	0.9455	1.1285
	8	0.9495	1.1609	0.9445	1.1667
LATE	2	0.9530	3.0955	0.9493	3.0963
	4	0.9547	3.0122	0.9517	3.0110
	8	0.9555	3.1284	0.9505	3.1239
		$n = 1000$			
ATE	2	0.9485	0.7349	0.9550	0.7352
	4	0.9500	0.7086	0.9505	0.7102
	8	0.9417	0.7285	0.9430	0.7318
LATE	2	0.9507	1.9468	0.9560	1.9452
	4	0.9515	1.8836	0.9507	1.8857
	8	0.9467	1.9469	0.9463	1.9480
		$n = 2000$			
ATE	2	0.9490	0.5194	0.9530	0.5193
	4	0.9483	0.4998	0.9485	0.5004
	8	0.9505	0.5121	0.9483	0.5139
LATE	2	0.9495	1.3714	0.9530	1.3721
	4	0.9493	1.3250	0.9483	1.3258
	8	0.9533	1.3663	0.9505	1.3658