Supplementary Material for
“Doubly Robust Semiparametric Inference Using Regularized Calibrated Estimation with High-dimensional Data”
Satyajit Ghosh & Zhiqiang Tan

I  Technical details

I.1 Probability lemmas

Denote by \( \Omega_0 \) the event that \((\hat{\alpha}_1 - \bar{\alpha}_1)^T \tilde{\Sigma}_0 (\hat{\alpha}_1 - \bar{\alpha}_1) \leq M_0 \lambda_0^2, \|\hat{\alpha}_1 - \bar{\alpha}_1\|_1 \leq M_0 \lambda_0, (\hat{\gamma}_1 - \bar{\gamma}_1)^T \tilde{\Sigma}_0 (\hat{\gamma}_1 - \bar{\gamma}_1) \leq M_0 \lambda_0^2, \|\hat{\gamma}_1 - \bar{\gamma}_1\|_1 \leq M_0 \lambda_0, \) and \(|\hat{\theta}_1 - \theta^*| \leq M_1/2 \lambda_0 \). Then Assumption 2(iv) says that \( P(\Omega_0) \geq 1 - \epsilon \).

Lemma S1. Denote by \( \Omega_1 \) the event that
\[
\sup_{j=1,\ldots,p} \left| \tilde{E} \left\{ \xi_j(X) \frac{\partial \tau}{\partial \eta} (U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2) \right\} \right| \leq B_0 \lambda_0,
\]
where \( B_0 = C_0(B_{02} + \sqrt{2}B_{01}) \). Under Assumption 2(i)–(ii), if \( \lambda_0 \leq 1 \), then \( P(\Omega_1) \geq 1 - 2\epsilon \).

Proof. The variable \( \frac{\partial \tau}{\partial \eta} (U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2) \) has mean 0 (because \( \xi \) includes a constant) and is sub-exponential with parameters \((B_{01}, B_{02})\). For \( j = 1, \ldots, p \), the variable \( \xi_j(X) \frac{\partial \tau}{\partial \eta} (U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2) \) has mean 0 and is sub-exponential with parameter \((C_0B_{01}, C_0B_{02})\). By Bernstein’s inequality (Bühlmann & Van de Geer 2011, Lemma 14.9; Tan 2020a, Lemma 16),
\[
P \left\{ \left| \tilde{E}(Y_j) \right| \geq C_0B_{02}t + C_0B_{01}\sqrt{2t} \right\} \leq 2\epsilon \frac{t}{p}
\]
where \( t = \log(p/\epsilon)/n = \lambda_0^2 \). The result then follows from the union bound. \( \square \)

Lemma S2. Denote by \( \Omega_{21}, \Omega_{22}, \Omega_{23} \) respectively the events that
\[
\sup_{j,k=1,\ldots,p} \left| (\tilde{E} - E) \left\{ \xi_j \xi_k T_{n_0}^{(1)} (U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2) \right\} \right| \leq B_{15} \lambda_0,
\]
\[
\sup_{j,k=1,\ldots,p} \left| (\tilde{E} - E) \left\{ \xi_j \xi_k T_{n_0}^{(1)} (U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2) \right\} \right| \leq B_{15} \lambda_0,
\]
\[
\sup_{j,k=1,\ldots,p} \left| (\tilde{E} - E) \left\{ \xi_j \xi_k \frac{\partial^2 \tau}{\partial \eta \partial f} (U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2) \right\} \right| \leq B_{15} \lambda_0,
\]
where $B_{15} = 2(B_{14} + B_{13})$, $B_{13} = (2C_0^2B_{11}^2 + 8C_0^4C_1^2)^{1/2}$, and $B_{14} = 2B_{12} + 2C_0^2C_1$. Under Assumptions 2(i) and 3(i)–(ii), if $\lambda_0 \leq 1$, then $P(\Omega_{21}) \geq 1 - 2\epsilon^2$, $P(\Omega_{22}) \geq 1 - 2\epsilon^2$, and $P(\Omega_{23}) \geq 1 - 2\epsilon^2$.

**Proof.** First, we show that for $j, k = 1, \ldots, p$, $\xi_j \xi_k T_{(1)_{\eta,\theta}}(U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2)$ is sub-exponential with parameter $(B_{13}, B_{14})$. Denote $T_{(1)_{\eta,\theta}} = T_{(1)_{\eta,\theta}}(U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2)$. Then for $k \geq 2$, 

$$
\begin{align*}
E \left| \xi_j \xi_k T_{(1)_{\eta,\theta}} - E \left\{ \xi_j \xi_k T_{(1)_{\eta,\theta}} \right\} \right|^k \\
\leq 2^{k-1} \left[ E \left| \xi_j \xi_k \left( T_{(1)_{\eta,\theta}} - E(T_{(1)_{\eta,\theta}}) \right) \right|^k + E \left| \xi_j \xi_k E(T_{(1)_{\eta,\theta}}) - E(\xi_j \xi_k T_{(1)_{\eta,\theta}}) \right|^k \right] \\
\leq 2^{k-1} \left\{ \left( C_0^2 \frac{k!}{2} B_{11}^2 B_{12}^{k-2} + (2C_0^2C_1)k \right) \right\} \leq \frac{k!}{2}(2C_0^2B_{11}^2 + 8C_0^4C_1^2)(2B_{12} + 2C_0^2C_1)^{k-2}.
\end{align*}
$$

Applying Bernstein’s inequality to $\xi_j \xi_k T_{(1)_{\eta,\theta}}(U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2)$ yields

$$
P \left\{ \left| (\hat{E} - E) \left\{ \xi_j \xi_k T_{(1)_{\eta,\theta}}(U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2) \right\} \right| \geq B_{14}t + B_{13}\sqrt{2t} \right\} \leq \frac{2\epsilon^2}{p^2}
$$

where $t = \log(p^2/\epsilon^2)/n = 2\lambda_0^2$. Then $P(\Omega_{21}) \geq 1 - 2\epsilon^2$ by the union bound. Similarly, $P(\Omega_{22}) \geq 1 - 2\epsilon^2$ and $P(\Omega_{23}) \geq 1 - 2\epsilon^2$.

**Lemma S3.** Denote by $\Omega_{24}$ the event that

$$
\sup_{j, k = 1, \ldots, p} \left| (\hat{E} - E)(\xi_j \xi_k) \right| \leq 4C_0^2\lambda_0.
$$

Under Assumption 2(i), $P(\Omega_{24}) \geq 1 - 2\epsilon^2$.

**Proof.** For $j, k = 1, \ldots, p$, by Hoeffding’s inequality (Bühlmann & Van de Geer 2011, Lemma 14.11; Tan 2020a, Lemma 14),

$$
P \left\{ \left| (\hat{E} - E)(\xi_j \xi_k) \right| \geq 2C_0^2(\sqrt{2}t) \right\} \leq \frac{2\epsilon^2}{p^2}
$$

where $|\xi_j \xi_k - E(\xi_j \xi_k)| \leq 2C_0^2$ and $t = \{\log(p^2/\epsilon^2)/n\}^{1/2} = \sqrt{2}\lambda_0$. The result then follows from the union bound.

**Lemma S4.** Denote $\Omega_2 = \Omega_{21} \cap \Omega_{22} \cap \Omega_{23} \cap \Omega_{24}$. Under Assumptions 2(i) and 3(i)–(ii), if $\lambda_0 \leq 1$, then $P(\Omega_2) \geq 1 - 8\epsilon^2$. Moreover, in the event $\Omega_2$, we have for any vector $b \in \mathbb{R}^p$,

$$
\begin{align*}
\hat{E} \left\{ T_{(1)_{\eta,\theta}}(U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2)(b^T\xi)^2 \right\} &\leq C_1 \hat{E} \left\{ (b^T\xi)^2 \right\} + (1 + C_1)|b|_1 \lambda_0, \\
\hat{E} \left\{ T_{(1)_{\eta,\theta}}(U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2)(b^T\xi)^2 \right\} &\leq C_1 \hat{E} \left\{ (b^T\xi)^2 \right\} + (1 + C_1)|b|_1 \lambda_0, \\
\hat{E} \left\{ \frac{\partial^2 \tau}{\partial\eta_0 \partial\eta_f}(U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2)(b^T\xi)^2 \right\} &\geq c_2 \hat{E} \left\{ (b^T\xi)^2 \right\} - (1 + c_2)|b|_1 \lambda_0.
\end{align*}
$$

2
where \( B_1 = \max(4C_0^2, B_{15}) \).

**Proof.** Combining Lemmas S2–S3 shows that \( P(\Omega_2) \geq 1 - 8\epsilon^2 \). In the event \( \Omega_2 \), simple manipulation yields

\[
\left| (\hat{E} - E) \left\{ T_{\eta_0}^{(1)}(U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2)(b^T \xi)^2 \right\} \right| \leq B_{15}\lambda_0 \|b\|_1^2,
\]

\[
\left| (\hat{E} - E) \left\{ (b^T \xi)^2 \right\} \right| \leq 4C_0^2\lambda_0 \|b\|_1.
\]

By the law of iterated expectations and Assumption 3(i),

\[
E \left\{ T_{\eta_0}^{(1)}(U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2)(b^T \xi)^2 \right\}
= E \left\{ T_{\eta_0}^{(1)}(U; \theta^*, \bar{\alpha}_1, \bar{\gamma}_2)(b^T \xi)^2 \right\} \leq C_1 E \left\{ (b^T \xi)^2 \right\}
\]

Combining the preceding three inequalities yields the result on \( T_{\eta_0}^{(1)} \). Similarly, the results on \( T_{\eta_0 \theta}^{(1)} \) and \( \frac{\partial^2 x}{\partial \eta_0 \partial \eta_f} \) can be shown. \( \square \)

### I.2 Proofs of Theorem 1, Corollary 3, and Theorem 2

We split the proof of Theorem 1 into a series of lemmas. The first one is usually called a basic inequality for \( \hat{\gamma}_2 \), but depending on the first-step estimators \( (\hat{\theta}_1, \hat{\alpha}_1) \).

**Lemma S5.** For any vector \( \gamma \in \mathbb{R}^p \), we have

\[
D_2^1(\hat{\gamma}_2, \gamma; \hat{\theta}_1, \hat{\alpha}_1) + A_1\lambda_0\|\hat{\gamma}_2\|_1 \leq (\hat{\gamma}_2 - \gamma)^T \hat{E} \left\{ \xi \frac{\partial \tau}{\partial \eta_g}(U; \hat{\theta}_1, \hat{\alpha}_1, \gamma) \right\} + A_1\lambda_0\|\gamma\|_1. \tag{S1}
\]

**Proof.** For any \( u \in (0, 1] \), the definition of \( \hat{\gamma}_2 \) implies

\[
L_2(\hat{\gamma}_2; \hat{\theta}_1, \hat{\alpha}_1) + A_1\lambda_0\|\hat{\gamma}_2\|_1
\leq L_2\{(1 - u)\hat{\gamma}_2 + u\gamma; \hat{\theta}_1, \hat{\alpha}_1\} + A_1\lambda_0\|(1 - u)\hat{\gamma}_2 + u\gamma\|_1,
\]

which, by the convexity of \( \| \cdot \|_1 \), gives

\[
L_2(\hat{\gamma}_2; \hat{\theta}_1, \hat{\alpha}_1) - L_2\{(1 - u)\hat{\gamma}_2 + u\gamma; \hat{\theta}_1, \hat{\alpha}_1\} + A_1\lambda_0u\|\hat{\gamma}_2\|_1 \leq A_1\lambda_0u\|\gamma\|_1.
\]

Dividing both sides of the preceding inequality by \( u \) and letting \( u \to 0^+ \) yields

\[
-(\hat{\gamma}_2 - \gamma)^T \hat{E} \left\{ \xi \frac{\partial \tau}{\partial \eta_g}(U; \hat{\theta}_1, \hat{\alpha}_1, \hat{\gamma}_2) \right\} + A_1\lambda_0\|\hat{\gamma}_2\|_1 \leq A_1\lambda_0\|\gamma\|_1.
\]
which leads to (S1) after a simple rearrangement using (38).

The second lemma deals with the dependency on \((\hat{\theta}_1, \hat{\alpha}_1)\) in the upper bound from the basic inequality (S1). Denote

\[
Q_2(\gamma_2, \bar{\gamma}_2; \theta^*, \bar{\alpha}_1) = \bar{E}\left\{ \frac{\partial^2 \tau}{\partial \eta_2 \partial \eta_f}(U; \theta^*, \bar{\alpha}_1, \gamma_2)(\hat{\gamma}_2 \xi - \bar{\gamma}_2 \xi)^2 \right\}
\]

\[= (\hat{\gamma}_2 - \bar{\gamma}_2)^T \Sigma_\gamma(\hat{\gamma}_2 - \bar{\gamma}_2).\]

where \(\Sigma_\gamma = \bar{E}\left\{ \frac{\partial^2 \tau}{\partial \eta_2 \partial \eta_f}(U; \theta^*, \bar{\alpha}_1, \gamma_2)\xi \xi^T \right\}.\)

**Lemma S6.** Suppose that Assumptions 3(i) and 3(iv) hold. In the event \(\Omega_0 \cap \Omega_2\), we have

\[
(\hat{\gamma}_2 - \bar{\gamma}_2)^T \bar{E}\left\{ \xi \frac{\partial \tau}{\partial \eta_2}(U; \hat{\theta}_1, \hat{\alpha}_1, \gamma) \right\}
\]

\[
\leq (\hat{\gamma}_2 - \bar{\gamma}_2)^T \bar{E}\left\{ \xi \frac{\partial \tau}{\partial \eta_2}(U; \theta^*, \bar{\alpha}_1, \gamma_2) \right\}
\]

\[
+ (C_{12} M_0 \gamma_0^2)^{1/2} \{Q_2(\hat{\gamma}_2, \bar{\gamma}_2; \theta^*, \bar{\alpha}_1)\}^{1/2} + C_{13} \lambda_0 \|\hat{\gamma}_2 - \gamma_2\|_1,
\]

where \(C_{12} = 4c_1^{-1} C_1 (C_1 + C_{11} \gamma_0)\), \(C_{13} = \{C_1^{1/2} + (1 + c_2^{-1})^{1/2}B_1^{1/2}C_1^{1/2}\} \{4(C_1 + C_{11} \gamma_0) \gamma_0\}^{1/2}\), and \(C_{11} = (1 + C_1) B_1\) with \(B_1\) from Lemma S4.

**Proof.** Consider the following decomposition

\[
(\hat{\gamma}_2 - \bar{\gamma}_2)^T \bar{E}\left\{ \xi \frac{\partial \tau}{\partial \eta_2}(U; \hat{\theta}_1, \hat{\alpha}_1, \gamma_2) \xi \right\}
\]

\[= (\hat{\gamma}_2 - \bar{\gamma}_2)^T \bar{E}\left\{ \xi \frac{\partial \tau}{\partial \eta_2}(U; \theta^*, \bar{\alpha}_1, \gamma_2) \right\} + \Delta_1 + \Delta_2,
\]

where

\[
\Delta_1 = (\hat{\gamma}_2 - \bar{\gamma}_2)^T \bar{E}\left\{ \xi \left\{ \frac{\partial \tau}{\partial \eta_2}(U; \hat{\theta}_1, \hat{\alpha}_1, \gamma_2) - \frac{\partial \tau}{\partial \eta_2}(U; \theta^*, \bar{\alpha}_1, \gamma_2) \right\} \right\},
\]

\[
\Delta_2 = (\hat{\gamma}_2 - \bar{\gamma}_2)^T \bar{E}\left\{ \xi \left\{ \frac{\partial \tau}{\partial \eta_2}(U; \theta^*, \bar{\alpha}_1, \gamma_2) - \frac{\partial \tau}{\partial \eta_2}(U; \theta^*, \bar{\alpha}_1, \gamma_2) \right\} \right\}.
\]

In the event \(\Omega_0, (\hat{\theta}_1, \hat{\alpha}_1) \in \mathcal{N}_1\) by Assumption 3(iv). By the mean value theorem and the Cauchy–Schwarz inequality, and Assumption 3(i),

\[|\Delta_2| = \left| \bar{E}\left\{ (\hat{\gamma}_2 - \bar{\gamma}_2)^T \xi Q_{y_2}^2(U; \theta^*, \bar{\alpha}_1, \gamma_2)(\hat{\alpha}_1 - \bar{\alpha}_1)^T \xi \right\} \right|
\]

\[\leq \bar{E}^{1/2}\left\{ T_{y_2}^1(U; \theta^*, \bar{\alpha}_1, \gamma_2)(\hat{\gamma}_2 \xi - \bar{\gamma}_2 \xi)^2 \right\} \bar{E}^{1/2}\left\{ T_{y_2}^1(U; \theta^*, \bar{\alpha}_1, \gamma_2)(\hat{\alpha}_1 \xi - \bar{\alpha}_1 \xi)^2 \right\},\]
where \( \hat{\alpha} \) lies between \( \hat{\alpha}_1 \) and \( \bar{\alpha}_1 \). Hence in the event \( \Omega_0 \cap \Omega_2 \) by Lemma S4,
\[
|\Delta_2| \leq \left| C_1 \tilde{E} \left\{ (\hat{\gamma}_2^T \xi - \bar{\gamma}_2^T \xi)^2 \right\} + C_{11} \lambda_0 \| \hat{\gamma}_2 - \bar{\gamma}_2 \|_1^{1/2} \left| C_1 \tilde{E} \left\{ (\hat{\alpha}_1^T \xi - \bar{\alpha}_1^T \xi)^2 \right\} + C_{11} \lambda_0 \| \hat{\alpha}_1 - \bar{\alpha}_1 \|_1^{1/2} \right|,
\]
where \( C_{11} = (1 + C_1) B_1 \). Similarly, in the event \( \Omega_0 \cap \Omega_2 \) by Lemma S4,
\[
|\Delta_1| = \left| \tilde{E} \left\{ (\hat{\gamma}_2 - \bar{\gamma}_2)^T \xi \frac{\partial^2 \tau}{\partial \eta_\theta \partial \theta_\theta} (U; \hat{\theta}, \hat{\alpha}_1, \bar{\gamma}_2)(\hat{\theta}_1 - \theta^*) \right\} \right| \\
\leq \tilde{E}^{1/2} \left\{ T_{\eta_\theta}(U; \theta^*, \hat{\alpha}_1, \bar{\gamma}_2)(\hat{\gamma}_2^T \xi - \bar{\gamma}_2^T \xi)^2 \right\} \tilde{E}^{1/2} \left\{ T_{\eta_\theta}(U; \theta^*, \hat{\alpha}_1, \bar{\gamma}_2)(\hat{\theta}_1 - \theta^*)^2 \right\} \\
\leq \left| C_1 \tilde{E} \left\{ (\hat{\gamma}_2^T \xi - \bar{\gamma}_2^T \xi)^2 \right\} + C_{11} \lambda_0 \| \hat{\gamma}_2 - \bar{\gamma}_2 \|_1^{1/2} \right| \left[ (C_1 + B_{15} \lambda_0)(\hat{\theta}_1 - \theta^*)^2 \right]^{1/2},
\]
where \( \hat{\theta} \) lies between \( \hat{\theta}_1 \) and \( \theta^* \). Hence in the event \( \Omega_0 \cap \Omega_2 \),
\[
|\Delta_1| + |\Delta_2| \leq 2 \left\{ (C_1 + C_{11} M_0 \lambda_0) M_0 \lambda_0 \right\} \left\{ (C_1 + C_{11} M_0 \lambda_0) M_0 \lambda_0 \right\}^{1/2} \left\{ (\hat{\gamma}_2^T \xi - \bar{\gamma}_2^T \xi)^2 \right\} + C_{11} \lambda_0 \| \hat{\gamma}_2 - \bar{\gamma}_2 \|_1^{1/2} \\
= (M_{01} \lambda_0^2)^{1/2} \tilde{E}^{1/2} \left\{ (\hat{\gamma}_2^T \xi - \bar{\gamma}_2^T \xi)^2 \right\} + M_{02} \lambda_0 \| \hat{\gamma}_2 - \bar{\gamma}_2 \|_1,
\]
where \( M_{01} = 4 C_1 (C_1 + C_{11} M_0 \lambda_0) M_0 \) and \( M_{02} = \{4 C_{11} (C_1 + C_{11} M_0 \lambda_0) M_0 \lambda_0 \}^{1/2} \). Furthermore, in the event \( \Omega_2 \) by Lemma S4,
\[
\tilde{E} \left\{ (\hat{\gamma}_2^T \xi - \bar{\gamma}_2^T \xi)^2 \right\} \\
\leq c_2^{-1} \tilde{E} \left\{ \frac{\partial^2 \tau}{\partial \eta_\theta \partial \eta_f} (U; \theta^*, \hat{\alpha}_1, \bar{\gamma}_2)(\hat{\gamma}_2^T \xi - \bar{\gamma}_2^T \xi)^2 \right\} + (1 + c_2^{-1}) B_1 \lambda_0 \| \hat{\gamma}_2 - \bar{\gamma}_2 \|_1^{2}.
\]
Combining the preceding inequalities shows that in event \( \Omega_0 \cap \Omega_2 \),
\[
|\Delta_1| + |\Delta_2| \leq (M_{03} \lambda_0^2)^{1/2} \tilde{E}^{1/2} \left\{ \frac{\partial^2 \tau}{\partial \eta_\theta \partial \eta_f} (U; \theta^*, \hat{\alpha}_1, \bar{\gamma}_2)(\hat{\gamma}_2^T \xi - \bar{\gamma}_2^T \xi)^2 \right\} + M_{04} \lambda_0 \| \hat{\gamma}_2 - \bar{\gamma}_2 \|_1,
\]
where \( M_{03} = c_2^{-1} M_{01} \) and \( M_{04} = M_{02} + \{ (1 + c_2^{-1}) B_1 \hat{C}_{12} M_0 \lambda_0 \}^{1/2} = \{ C_{11}^{1/2} + (1 + c_2^{-1}) B_1^{1/2} C_1^{1/2} \} \times \{4 (C_1 + C_{11} M_0 \lambda_0) M_0 \lambda_0 \}^{1/2} \). Using \( M_0 \lambda_0 \leq \varrho_0 \) by Assumption 3(iv) yields the desired result.

The third lemma derives an implication of the basic inequality (S1) using the triangle inequality for the \( L_1 \) norm, while incorporating the bound from Lemma S6.

**Lemma S7.** Denote \( b = \hat{\gamma}_2 - \bar{\gamma}_2 \). In the event \( \Omega_0 \cap \Omega_1 \cap \Omega_2 \), (S1) implies that
\[
D^1_1(\hat{\gamma}_2, \bar{\gamma}_2; \hat{\theta}_1, \hat{\alpha}_1) + A_{11} \lambda_0 \| b \|_1 \\
\leq 2 A_{11} \lambda_0 \sum_{j \in \mathcal{S}_{\hat{\gamma}_2}} |b_j| + (C_{12} M_0 \lambda_0^2)^{1/2} \left\{ Q_2(\hat{\gamma}_2, \bar{\gamma}_2; \theta^*, \hat{\alpha}_1) \right\}^{1/2}, \tag{S2}
\]
where \( A_{11} = A_1 - B_0 - C_{13} \), with \( B_0 \) from Lemma S1.
Proof. In the event $\Omega_1$ from Lemma S1, we have
\[
b^\top \tilde{E} \left\{ \xi \frac{\partial \tau}{\partial \eta_g} (U; \theta^*, \tilde{\alpha}_1, \tilde{\gamma}_2) \right\} \leq B_0 \lambda_0 \|b\|_1.
\]
From (S1), the preceding bound, and Lemma S6, we have in the event $\Omega_0 \cap \Omega_1 \cap \Omega_2$,
\[
D_2^1(\tilde{\gamma}_2, \tilde{\gamma}_2; \tilde{\theta}_1, \tilde{\alpha}_1) + A_1 \lambda_0 \|\tilde{\gamma}_2\|_1 \\
\leq B_0 \lambda_0 \|b\|_1 + A_1 \lambda_0 \|\tilde{\gamma}_2\|_1 + (C_{12} M_0 \lambda_0^2)^{1/2} \{Q_2(\tilde{\gamma}_2; \theta^*, \tilde{\alpha}_1)\}^{1/2} + C_{13} \lambda_0 \|b\|_1.
\]
Using the identity $\|\tilde{\gamma}_2\| = \|\tilde{\gamma}_2 - \check{\gamma}_2\|_{S_\gamma}$ and the triangle inequality $\|\tilde{\gamma}_2\| \geq \|\tilde{\gamma}_2\| - \|\tilde{\gamma}_2 - \check{\gamma}_2\|$ for $j \in S_{\tilde{\gamma}_2}$ and rearranging the result yields (S2).

The following lemma provides a desired bound relating the Bregman divergence $D_2^1(\gamma, \tilde{\gamma}_2; \tilde{\theta}_1, \tilde{\alpha}_1)$ with the quadratic function $(\gamma - \tilde{\gamma}_2)^\top \bar{\Sigma}_\gamma (\gamma - \tilde{\gamma}_2)$.

Lemma S8. Suppose that Assumptions 2(i) and 3(iii)–(iv) hold. In the event $\Omega_0$, we have for any $\gamma \in \mathbb{R}^p$, 
\[
D_2^1(\gamma, \tilde{\gamma}_2; \tilde{\theta}_1, \tilde{\alpha}_1) \geq e^{-\Delta} \frac{1 - e^{-c_0 C_2 \|\gamma - \tilde{\gamma}_2\|_1}}{C_0 C_2 \|\gamma - \tilde{\gamma}_2\|_1} (b^\top \bar{\Sigma}_\gamma b),
\]
where $b = \gamma - \tilde{\gamma}_2$ and $\Delta = C_2(\|\tilde{\theta}_1 - \theta^*\| + C_0 \|\tilde{\alpha}_1 - \tilde{\alpha}_1\|_1)$. Throughout, set $(1 - e^{-c})/c = 1$ for $c = 0$.

Proof. By direct calculation, we have
\[
D_2^1(\gamma, \tilde{\gamma}_2; \tilde{\theta}_1, \tilde{\alpha}_1) = (\gamma - \tilde{\gamma}_2)^\top \tilde{E} \left\{ \xi \frac{\partial \tau}{\partial \eta_g} (U; \tilde{\theta}_1, \tilde{\alpha}_1, \gamma) - \frac{\partial \tau}{\partial \eta_g} (U; \tilde{\theta}_1, \tilde{\alpha}_1, \tilde{\gamma}_2) \right\} \]
\[
= \tilde{E} \left\{ \int_0^1 \frac{\partial^2 \tau}{\partial \eta_g \partial \eta_f} (U; \tilde{\theta}_1, \tilde{\alpha}_1, \gamma_u) du \right\} (\gamma^\top \xi - \tilde{\gamma}_2^\top \xi)^2,
\]
where $\gamma_u = \tilde{\gamma}_2 + u(\gamma - \tilde{\gamma}_2)$. In the event $\Omega_0$, $(\tilde{\theta}_1, \tilde{\alpha}_1) \in \mathcal{N}_1$ by Assumption 3(iv). Then by Assumption 2(i) and 3(iii), we have
\[
D_2^1(\gamma, \tilde{\gamma}_2; \tilde{\theta}_1, \tilde{\alpha}_1) \\
\geq \tilde{E} \left\{ \int_0^1 e^{-c_2 (|\tilde{\theta}_1 - \theta^*| + |(\tilde{\alpha}_1 - \tilde{\alpha}_1)^\top \widetilde{\xi} + u(\gamma - \tilde{\gamma}_2)^\top \xi|)} du \right\} \frac{\partial^2 \tau}{\partial \eta_g \partial \eta_f} (U; \theta^*, \tilde{\alpha}_1, \tilde{\gamma}_2)(\gamma^\top \xi - \tilde{\gamma}_2^\top \xi)^2 \]
\[
\geq \left\{ \int_0^1 e^{-c_2 (|\tilde{\theta}_1 - \theta^*| + C_0 \|\tilde{\alpha}_1 - \tilde{\alpha}_1\|_1 + u C_0 \|\gamma - \tilde{\gamma}_2\|_1)} du \right\} \tilde{E} \left\{ \frac{\partial^2 \tau}{\partial \eta_g \partial \eta_f} (U; \theta^*, \tilde{\alpha}_1, \tilde{\gamma}_2)(\gamma^\top \xi - \tilde{\gamma}_2^\top \xi)^2 \right\}.
\]
The desired result follows because \( f_0^1 e^{-cu} \, du = (1 - e^{-c})/c \) for \( c \geq 0 \).

The following lemma shows that Assumption 2(iii), a theoretical compatibility condition for \( \Sigma_\gamma \), implies an empirical compatibility condition for \( \tilde{\Sigma}_\gamma \).

**Lemma S9.** Suppose that Assumption 3(iv) holds. In the event \( \Omega_2 \), Assumption 2(iii) implies that for any vector \( b \in \mathbb{R}^p \) such that \( \sum_{j \notin S_{\gamma_2}} |b_j| \leq \mu_1 \sum_{j \in S_{\gamma_2}} |b_j| \), we have

\[
\nu_{11} \left( \sum_{j \in S_{\gamma_2}} |b_j| \right)^2 \leq |S_{\gamma_2}| \left( b^T \tilde{\Sigma}_\gamma b \right),
\]

where \( \nu_{11} = \nu_1 \{ 1 - \nu_1^{-2} (1 + \mu_1)^2 \varrho_1 B_{15} \}^{1/2} = \nu_1 (1 - \varrho_2)^{1/2} \).

**Proof.** In the event \( \Omega_2 \), we have \( |b^T (\tilde{\Sigma}_\gamma - \Sigma_\gamma) b| \leq B_{15} \lambda_0 \|b\|_1^2 \) from Lemma S2. Then Assumption 2(iii) implies that for any \( b = (b_1, \ldots, b_p)^T \) satisfying \( \sum_{j \notin S_{\gamma_2}} |b_j| \leq \mu_1 \sum_{j \in S_{\gamma_2}} |b_j| \),

\[
\nu_2^2 \| b_{S_{\gamma_2}} \|_1^2 \leq |S_{\gamma_2}| (b^T \Sigma_\gamma b) \leq |S_{\gamma_2}| \left( b^T \tilde{\Sigma}_\gamma b + B_{15} \lambda_0 \|b\|_1^2 \right) \leq |S_{\gamma_2}| (b^T \tilde{\Sigma}_\gamma b) + B_{15} |S_{\gamma_2}| \lambda_0 (1 + \mu_1)^2 \| b_{S_{\gamma_2}} \|_1^2,
\]

where \( \| b_{S_{\gamma_2}} \|_1 = \sum_{j \in S_{\gamma_2}} |b_j| \). The last inequality uses \( \|b\|_1 \leq (1 + \mu_1) \|b_{S_{\gamma_2}}\|_1 \). The desired result follows because \( |S_{\gamma_2}| \lambda_0 \leq \varrho_1 \) and \( \varrho_2 = \nu_1^2 B_{15} (1 + \mu_1)^2 \varrho_1 < 1 \) by Assumption 3(iv). \( \square \)

The final lemma completes the proof of Theorem 1, because \( P(\Omega_0 \cap \Omega_1 \cap \Omega_2) \geq 1 - (c_0 + 10) \varepsilon \) by Assumption 2(iv) and Lemmas S1 and S4.

**Lemma S10.** Suppose that Assumptions 2–3 hold and \( \lambda_0 \leq 1 \). Then for \( A_1 > (B_0 + C_{13})(\mu_1 + 1)/(\mu_1 - 1) \), inequality (39) holds in the event \( \Omega_0 \cap \Omega_1 \cap \Omega_2 \):

\[
D_1^2(\hat{\gamma}_2, \tilde{\gamma}_2; \hat{\theta}_1, \hat{\alpha}_1) + A_{11} \lambda_0 |\hat{\gamma}_2 - \tilde{\gamma}_2|_1 \leq \left\{ e^{\rho_2} (1 - \varrho_3)^{-1} \mu_{12} \nu_{11}^{-2} (|S_{\gamma_2}| \lambda_0^2) \right\} \lor \left\{ e^{\rho_2} (1 - \varrho_4)^{-1} \mu_{11}^{-2} C_{12} (M_0 \lambda_0^2) \right\}.
\]

**Proof.** Denote \( b = \hat{\gamma}_2 - \tilde{\gamma}_2 \), \( D_2^2 = D_1^2(\hat{\gamma}_2, \tilde{\gamma}_2; \hat{\theta}_1, \hat{\alpha}_1) \), \( D_2^4 = D_1^4(\hat{\gamma}_2, \tilde{\gamma}_2; \hat{\theta}_1, \hat{\alpha}_1) + A_{11} \lambda_0 \|b\|_1 \), \( Q_2 = Q_2(\hat{\gamma}_2, \tilde{\gamma}_2; \hat{\theta}_1, \hat{\alpha}_1) = b^T \tilde{\Sigma}_\gamma b \). In the event \( \Omega_0 \cap \Omega_1 \cap \Omega_2 \), inequality (S2) from Lemma S7 leads to two possible cases:

\[
\mu_{11} D_2^4 \leq (C_{12} M_0 \lambda_0^2)^{1/2} Q_2^{1/2}, \quad (S3)
\]
or \((1 - \mu_{11})D_2^\dagger \leq 2A_1\lambda_0 \sum_{j \in S_{\gamma_2}} |b_j|\), that is,

\[
D_2^\dagger \leq (\mu_1 + 1)A_{11}\lambda_0 \sum_{j \in S_{\gamma_2}} |b_j| = \mu_{12}\lambda_0 \sum_{j \in S_{\gamma_2}} |b_j|,
\]

(S4)

where \(\mu_{11} = 1 - 2A_1/\{(\mu_1 + 1)A_{11}\} \in (0,1]\) because \(A_1 > (B_0 + C_{13})(\mu_1 + 1)/(\mu_1 - 1)\), and \(\mu_{12} = (\mu_1 + 1)A_{11}\). We deal with the two cases separately as follows.

In the case where (S4) holds, \(\sum_{j \notin S_{\gamma_2}} |b_j| \leq \mu_1 \sum_{j \in S_{\gamma_2}} |b_j|\). Then by Lemma S9,

\[
\sum_{j \in S_{\gamma_2}} |b_j| \leq \nu_{11}^{-1}|S_{\gamma_2}|^{1/2} \left(b^T \tilde{\Sigma}_\gamma b\right)^{1/2}.
\]

(S5)

By Lemma S8, we have

\[
D_2^\dagger \geq e^{-\Delta \frac{1}{C_0C_2}} \left(\frac{1}{b^T \tilde{\Sigma}_\gamma b}\right),
\]

(S6)

where \(\Delta = C_2(|\hat{\theta}_1 - \theta^*| + C_0||\hat{\alpha}_1 - \bar{\alpha}_1||_1)\). Combining (S3), (S5), and (S6) yields

\[
D_2^\dagger \leq \mu_{12}^2\nu_{11}^{-2}|S_{\gamma_2}|\lambda_0^2 \frac{\Delta}{1 - e^{-C_0C_2||b||_1}}.
\]

(S7)

But \(A_{11}\lambda_0||b||_1 \leq D_2^\dagger\). Then (S7) along with \(|\Delta| \leq C_2(1+C_0)\varrho_0\) implies that \(1 - e^{-C_0C_2||b||_1} \leq C_0C_2A_{11}^{-1}\mu_{12}^2\nu_{11}^{-2}|S_{\gamma_2}|\lambda_0e^{C_2(1+C_0)\varrho_0} \leq \varrho_3 \leq 1\) by Assumption 3(iv). As a result, \(C_0C_2||b||_1 \leq -\log(1 - \varrho_3)\) and hence

\[
\frac{1}{C_0C_2||b||_1} = \int_0^1 e^{-C_0C_2||b||_1u} du \geq e^{-C_0C_2||b||_1} \geq 1 - \varrho_3.
\]

(S8)

From this bound, (S7) leads to \(D_2^\dagger \leq e^{C_2(1+C_0)\varrho_0}(1 - \varrho_3)^{-1}\mu_{12}^2\nu_{11}^{-2}|S_{\gamma_2}|\lambda_0^2\).

In the first case where (S3) holds, simple manipulation using (S6) yields

\[
D_2^\dagger \leq \mu_{11}^{-2}(C_{12}M_0^2\lambda_0^2)e^\Delta \frac{C_0C_2||b||_1}{1 - e^{-C_0C_2||b||_1}}.
\]

(S9)

Similarly as above, using \(A_{11}\lambda_0||b||_1 \leq D_2^\dagger\) and (S9) along with \(|\Delta| \leq C_2(1+C_0)\varrho_0\), we find \(1 - e^{-C_0C_2||b||_1} \leq C_0C_2A_{11}^{-1}\mu_{11}^2C_{12}M_0^2\lambda_0e^{C_2(1+C_0)\varrho_0} \leq \varrho_4 \leq 1\) by Assumption 3(iv). As a result, \(C_0C_2||b||_1 \leq -\log(1 - \varrho_4)\) and hence

\[
\frac{1}{C_0C_2||b||_1} = \int_0^1 e^{-C_0C_2||b||_1u} du \geq e^{-C_0C_2||b||_1} \geq 1 - \varrho_4.
\]

(S10)

From this bound, (S9) leads to \(D_2^\dagger \leq e^{C_2(1+C_0)\varrho_0}(1 - \varrho_4)^{-1}\mu_{11}^{-2}C_{12}M_0^2\lambda_0^2\). Therefore, (39) holds through (S3) and (S4) in the event \(\Omega_0 \cap \Omega_1 \cap \Omega_2\).
Proof of Corollary 3. Return to the proof of Lemma S10, where (S3) or (S4) holds in the event \( \Omega_0 \cap \Omega_1 \cap \Omega_2 \). If (S4) holds, then we have, by (S6) and (S10), \( b^T \Sigma_n b \leq e^{o_2}(1 - \varrho_3)^{-1} D_2^2 \). If (S3) holds, then we have, by (S6) and (S10), \( b^T \Sigma_n b \leq e^{o_2}(1 - \varrho_4)^{-1} D_2^2 \). Hence

\[
b^T \Sigma_n b \leq e^{o_2}(1 - \varrho_3 \lor \varrho_4)^{-1} D_2^2 \leq e^{o_2}(1 - \varrho_4)^{-1} C_3(|S_{\gamma_0} \lor M_0) \lambda_0^2 .\]

Moreover, by \((39)\), we have \( \|b\|_1 \leq A_1^{-1} C_3(|S_{\gamma_2} \lor M_0) \lambda_0 \) and hence \( \lambda_0 \|b\|_1^2 \leq A_1^{-2} C_3^2 (\varrho_0 \lor \varrho_1)(|S_{\gamma_2} \lor M_0) \lambda_0^2 \), because \( |S_{\gamma_2} \lor M_0) \lambda_0 \leq \varrho_0 \lor \varrho_1 \) by Assumption 3(iv). Then \((40)\) follows from the third inequality in Lemma S4. \( \square \)

Proof of Theorem 2. The proof follows from similar steps as in that of Theorem 1. The probability decreases from 1 - \((c_0 + 10)\epsilon\) to 1 - \((c_0 + 18)\epsilon\), due to additional restriction to the events similar to \( \Omega_1 \), \( \Omega_2 \), \( \Omega_2 \), and \( \Omega_2 \), while \( \Omega_2 \) is unchanged. \( \square \)

I.3 Proof of Theorem 3

We split the proof into three lemmas. The first one shows the consistency of \( \hat{\theta}_2 \) for \( \theta^* \).

Lemma S11. In the setting of Theorem 2, suppose that Assumption 6 holds and \( M_2 r_0 = o(1) \). Then \( \hat{\theta}_2 \) is consistent for \( \theta^* \), i.e., \( |\hat{\theta}_2 - \theta^*| = o_p(1) \).

Proof. By Theorem 2 and \( M_2 r_0 = o(1) \), we have \( \|\hat{\alpha}_2 - \tilde{\alpha}_2\|_1 = o_p(1) \) and \( \|\hat{\gamma}_2 - \tilde{\gamma}_2\|_1 = o_p(1) \). Hence for any small \( \epsilon > 0 \), \((\hat{\alpha}_2, \hat{\gamma}_2) \in \mathcal{N}_2 \) with probability at least 1 - \( \epsilon \) for all sufficiently large \( n \). In the following, we restrict analysis within this event.

To show \( |\hat{\theta}_2 - \theta^*| = o_p(1) \), by standard consistency arguments (e.g., Van der Vaart 2000) using Assumption 6(i)–(ii), it suffices to show that \( \tilde{E}\{\tau(U; \hat{\theta}_2, \tilde{\alpha}_2, \tilde{\gamma}_2)\} = o_p(1) \). Because \( \tilde{E}\{\tau(U; \hat{\theta}_2, \tilde{\alpha}_2, \tilde{\gamma}_2)\} = 0 \) by definition of \( \hat{\theta}_2 \), consider the decomposition

\[
\tilde{E}\{\tau(U; \hat{\theta}_2, \tilde{\alpha}_2, \tilde{\gamma}_2)\} - \tilde{E}\{\tau(U; \hat{\theta}_2, \tilde{\alpha}_2, \tilde{\gamma}_2)\} = \Delta_1 + \Delta_2 ,
\]

where

\[
\Delta_1 = \tilde{E}\{\tau(U; \hat{\theta}_2, \tilde{\alpha}_2, \tilde{\gamma}_2)\} - \tilde{E}\{\tau(U; \hat{\theta}_2, \tilde{\alpha}_2, \tilde{\gamma}_2)\},
\]

\[
\Delta_2 = \tilde{E}\{\tau(U; \hat{\theta}_2, \tilde{\alpha}_2, \tilde{\gamma}_2)\} - \tilde{E}\{\tau(U; \hat{\theta}_2, \tilde{\alpha}_2, \tilde{\gamma}_2)\}.
\]
By the mean value theorem, the Cauchy–Schwartz inequality and Assumption 6 (iii),

\[
\Delta_1 = \left\| (\hat{\alpha}_2 - \bar{\alpha}_2)^T \hat{E} \left\{ \xi \frac{\partial \tau}{\partial \alpha} (U; \hat{\theta}_2, \hat{\alpha}, \hat{\gamma}_2) \right\} \right\| \\
\leq \hat{E}^{1/2} \left\{ (\hat{\alpha}_2^T \xi - \bar{\alpha}_2^T \xi)^2 \right\} \hat{E}^{1/2} \left\{ T^{(2)}_{\eta_2}(U; \bar{\alpha}_2, \bar{\gamma}_2) \right\} = O_p(M_2^{1/2}r_0),
\]

\[
\Delta_2 = \left\| (\hat{\gamma}_2 - \bar{\gamma}_2)^T \hat{E} \left\{ \xi \frac{\partial \tau}{\partial \gamma} (U; \hat{\theta}_2, \hat{\alpha}_2, \hat{\gamma}) \right\} \right\| \\
\leq \hat{E}^{1/2} \left\{ (\hat{\gamma}_2^T \xi - \bar{\gamma}_2^T \xi)^2 \right\} \hat{E}^{1/2} \left\{ T^{(2)}_{\eta_2}(U; \bar{\alpha}_2, \bar{\gamma}_2) \right\} = O_p(M_2^{1/2}r_0),
\]

where \( \bar{\alpha} \) lies between \( \hat{\alpha}_2 \) and \( \bar{\alpha}_2 \), and \( \bar{\gamma} \) lies between \( \hat{\gamma}_2 \) and \( \bar{\gamma}_2 \). Hence \( |\Delta_1| + |\Delta_2| = o_p(1) \) because \( M_2^{1/2}r_0 \leq M_2r_0 = o(1) \) with \( M_2 \geq M_0 \geq 1 \).

The following lemma establishes the asymptotic expansion (43) for \( \hat{\theta}_2 \).

**Lemma S12.** In the setting of Theorem 2, suppose that Assumption 6 and 7(ii)–(iv) hold and \( M_2r_0 = o(1) \). Then \( \hat{\theta}_2 \) admits the asymptotic expansion (43),

\[
\hat{\theta}_2 - \theta^* = -H^{-1} \hat{E} \left\{ \tau (U; \theta^*, \bar{\alpha}_2, \bar{\gamma}_2) \right\} + O_p(M_2r_0^2),
\]

where \( H = E \left\{ \frac{\partial \tau}{\partial \theta} (U; \theta^*, \bar{\alpha}_2, \bar{\gamma}_2) \right\} \).

**Proof.** By Theorem 2, Lemma S11, and \( M_2r_0 = o(1) \), we have \( \| \hat{\alpha}_2 - \bar{\alpha}_2 \| = o_p(1), \| \hat{\gamma}_2 - \bar{\gamma}_2 \| = o_p(1), \) and \( |\hat{\theta}_2 - \theta^*| = o_p(1) \). Hence for any small \( \epsilon > 0, (\hat{\theta}_2, \hat{\alpha}_2, \hat{\gamma}_2) \in \mathcal{N}_3 \) with probability at least \( 1 - \epsilon \) for all sufficiently large \( n \). In the following, we restrict analysis within this event. Consider the decomposition

\[
\hat{E} \left\{ \tau (U; \hat{\theta}_2, \hat{\alpha}_2, \hat{\gamma}_2) \right\} - \hat{E} \left\{ \tau (U; \theta^*, \bar{\alpha}_2, \bar{\gamma}_2) \right\} = \Delta_3 + \Delta_4,
\]

where

\[
\Delta_3 = \hat{E} \left\{ \tau (U; \hat{\theta}_2, \hat{\alpha}_2, \hat{\gamma}_2) \right\} - \hat{E} \left\{ \tau (U; \theta^*, \hat{\alpha}_2, \hat{\gamma}_2) \right\},
\]

\[
\Delta_4 = \hat{E} \left\{ \tau (U; \theta^*, \hat{\alpha}_2, \hat{\gamma}_2) \right\} - \hat{E} \left\{ \tau (U; \theta^*, \bar{\alpha}_2, \bar{\gamma}_2) \right\}.
\]

We deal with the two terms \( \Delta_3 \) and \( \Delta_4 \) respectively.
By a Taylor expansion, $\Delta_4 = \Delta_{41} + \Delta_{42}$ with

$$\Delta_{41} = (\hat{\alpha}_2 - \bar{\alpha}_2)^T \hat{E} \left\{ \xi \frac{\partial^2 \tau}{\partial \eta_f^2} (U; \theta^*, \bar{\alpha}_2, \bar{\gamma}_2) \right\} + (\hat{\gamma}_2 - \bar{\gamma}_2)^T \hat{E} \left\{ \xi \frac{\partial \tau}{\partial \eta_f} (U; \theta^*, \bar{\alpha}_2, \bar{\gamma}_2) \right\}$$

$$\Delta_{42} = \frac{1}{2} (\hat{\alpha}_2 - \bar{\alpha}_2)^T \hat{E} \left\{ \xi \frac{\partial^2 \tau}{\partial \eta_f^2} (U; \theta^*, \bar{\alpha}, \bar{\gamma}) \xi^T \right\} (\hat{\alpha}_2 - \bar{\alpha}_2)$$

$$+ \frac{1}{2} (\hat{\gamma}_2 - \bar{\gamma}_2)^T \hat{E} \left\{ \xi \frac{\partial^2 \tau}{\partial \eta_f^2} (U; \theta^*, \bar{\alpha}, \bar{\gamma}) \xi^T \right\} (\hat{\gamma}_2 - \bar{\gamma}_2)$$

$$+ (\hat{\alpha}_2 - \bar{\alpha}_2)^T \hat{E} \left\{ \xi \frac{\partial^2 \tau}{\partial \eta_f \partial \eta_f} (U; \theta^*, \bar{\alpha}, \bar{\gamma}) \xi^T \right\} (\hat{\gamma}_2 - \bar{\gamma}_2)$$

where $(\bar{\alpha}, \bar{\gamma})$ lie between $(\hat{\alpha}_2, \bar{\gamma}_2)$ and $(\hat{\alpha}_2, \bar{\gamma}_2)$. As model (11) or (12) is correctly specified, Proposition 2 implies that calibration equations (22)–(23) are satisfied by $(\alpha, \gamma) = (\hat{\alpha}_2, \bar{\gamma}_2)$, that is, the variables $\xi_j \frac{\partial \tau}{\partial \eta_y} (U; \theta^*, \bar{\alpha}_2, \bar{\gamma}_2)$ and $\xi_j \frac{\partial \tau}{\partial \eta_y} (U; \theta^*, \hat{\alpha}_2, \bar{\gamma}_2)$ have mean 0 for $j = 1, \ldots, p$.

By Assumption 7(iii) and similar reasoning as in Lemma S1, we have

$$\sup_j |\hat{E} \{ \xi_j \frac{\partial \tau}{\partial \eta_y} (U; \theta^*, \bar{\alpha}_2, \bar{\gamma}_2) \}| = O_p(r_0), \quad \sup_j |\hat{E} \{ \xi_j \frac{\partial \tau}{\partial \eta_f} (U; \theta^*, \bar{\alpha}_2, \bar{\gamma}_2) \}| = O_p(r_0).$$

By Theorem 2, $\|\hat{\alpha}_2 - \bar{\alpha}_2\|_1 = O_p(M_2 r_0)$ and $\|\hat{\gamma}_2 - \bar{\gamma}_2\|_1 = O_p(M_2 r_0)$. Hence

$$\left| (\hat{\alpha}_2 - \bar{\alpha}_2)^T \hat{E} \left\{ \xi \frac{\partial \tau}{\partial \eta_y} (U; \theta^*, \bar{\alpha}_2, \bar{\gamma}_2) \right\} \right| = O_p(r_0) \|\hat{\alpha}_2 - \bar{\alpha}_2\|_1 = O_p(M_2 r_0^2),$$

$$\left| (\hat{\gamma}_2 - \bar{\gamma}_2)^T \hat{E} \left\{ \xi \frac{\partial \tau}{\partial \eta_f} (U; \theta^*, \bar{\alpha}_2, \bar{\gamma}_2) \right\} \right| = O_p(r_0) \|\hat{\gamma}_2 - \bar{\gamma}_2\|_1 = O_p(M_2 r_0^2),$$

and $|\Delta_{41}| = O_p(M_2 r_0^2)$. Moreover, by Assumption 7(iv) and similar reasoning as in Lemma S4, we have

$$\left| (\hat{\alpha}_2 - \bar{\alpha}_2)^T \hat{E} \left\{ \xi \frac{\partial^2 \tau}{\partial \eta_y^2} (U; \theta^*, \bar{\alpha}_2, \bar{\gamma}_2) \xi^T \right\} (\hat{\alpha}_2 - \bar{\alpha}_2) \right| \leq (\hat{\alpha}_2 - \bar{\alpha}_2)^T \hat{E} \left\{ \xi \frac{\partial^2 \tau}{\partial \eta_y^2} (U; \theta^*, \bar{\alpha}_2, \bar{\gamma}_2) \xi^T \right\} (\hat{\alpha}_2 - \bar{\alpha}_2)$$

$$\leq C_4 \hat{E} \{ (\hat{\alpha}_2^T \xi - \bar{\alpha}_2^T \xi)^2 \} + (1 + C_4) O_p(r_0) \|\hat{\alpha}_2 - \bar{\alpha}_2\|_1^2 = O_p(M_2 r_0^2),$$

where $\hat{E} \{ (\hat{\alpha}_2^T \xi - \bar{\alpha}_2^T \xi)^2 \} = O_p(M_2 r_0^2)$ and $O_p(r_0) O_p(M_2 r_0^2) = o_p(M_2 r_0^2)$ because $M_2 r_0 = o(1)$.

Similarly, we have

$$\left| (\hat{\gamma}_2 - \bar{\gamma}_2)^T \hat{E} \left\{ \xi \frac{\partial^2 \tau}{\partial \eta_f^2} (U; \theta^*, \bar{\alpha}, \bar{\gamma}) \xi^T \right\} (\hat{\gamma}_2 - \bar{\gamma}_2) \right| = O_p(M_2 r_0^2).$$
and by the Cauchy–Schwartz inequality,

\[ \left| (\hat{\alpha}_2 - \bar{\alpha})^T \bar{E} \left\{ \xi \frac{\partial^2 \tau}{\partial \eta_0 \partial \eta_f} (U; \theta^*, \bar{\alpha}, \bar{\gamma}) \right\} (\hat{\gamma}_2 - \bar{\gamma}_2) \right| \]

\[ \leq \left| (\hat{\alpha}_2 - \bar{\alpha})^T \bar{E} \left\{ \xi T^{(2)}_{\eta_0 \eta_f} (U; \theta^*, \bar{\alpha}_2, \bar{\gamma}_2) \right\} (\hat{\alpha}_2 - \bar{\alpha}_2) \right|^{1/2} \]

\[ \times \left[ (\hat{\gamma}_2 - \bar{\gamma}_2)^T \bar{E} \left\{ \xi T^{(2)}_{\eta_0 \eta_f} (U; \theta^*, \bar{\alpha}_2, \bar{\gamma}_2) \right\} (\hat{\gamma}_2 - \bar{\gamma}_2) \right]^{1/2} = O_p(M_2 r_0^2). \]

Hence \( |\Delta_2| = O_p(M_2 r_0^2) \) and \( |\Delta_4| = O_p(M_2 r_0^2) \).

Next, by the mean value theorem, we have

\[ \Delta_3 = (\hat{\theta}_2 - \theta^*) \bar{E} \left\{ \frac{\partial \tau}{\partial \theta} (U; \hat{\theta}, \hat{\alpha}_2, \hat{\gamma}_2) \right\}, \]

where \( \hat{\theta} \) lies between \( \hat{\theta}_2 \) and \( \theta^* \). Consider the decomposition

\[ \bar{E} \left\{ \frac{\partial \tau}{\partial \theta} (U; \hat{\theta}, \hat{\alpha}_2, \hat{\gamma}_2) \right\} = E \left\{ \frac{\partial \tau}{\partial \theta} (U; \theta^*, \hat{\alpha}_2, \hat{\gamma}_2) \right\} + \Delta_{31} + \Delta_{32}, \]  

where

\[ \Delta_{31} = \bar{E} \left\{ \frac{\partial \tau}{\partial \theta} (U; \hat{\theta}, \hat{\alpha}_2, \hat{\gamma}_2) \right\} - E \left\{ \frac{\partial \tau}{\partial \theta} (U; \hat{\theta}, \hat{\alpha}_2, \hat{\gamma}_2) \right\}, \]

\[ \Delta_{32} = E \left\{ \frac{\partial \tau}{\partial \theta} (U; \hat{\theta}, \hat{\alpha}_2, \hat{\gamma}_2) \right\} - E \left\{ \frac{\partial \tau}{\partial \theta} (U; \theta^*, \hat{\alpha}_2, \hat{\gamma}_2) \right\}. \]

By Assumption 7(ii) and the uniform law of large numbers (Ferguson 1996, Theorem 16),

\[ |\Delta_{31}| \leq \sup_{(\theta, \alpha, \gamma) \in \mathcal{N}_3} |(\bar{E} - E) \left\{ \frac{\partial \tau}{\partial \theta} (U; \theta, \alpha, \gamma) \right\} | = o_p(1). \]

Moreover, by \( |\hat{\theta} - \theta^*| \leq |\hat{\theta}_2 - \theta^*| = o_p(1) \) and the continuous mapping theorem, \( |\Delta_{32}| = o_p(1) \). Hence \( \bar{E} \left\{ \frac{\partial \tau}{\partial \theta} (U; \hat{\theta}, \hat{\alpha}_2, \hat{\gamma}_2) \right\} = H + o_p(1) \) and \( \Delta_3 = (\hat{\theta}_2 - \theta^*) \{ H + o_p(1) \} \).

Finally, from the preceding analysis, (S11) yields

\[ -\bar{E} \{ \tau (U; \theta^*, \hat{\alpha}_2, \hat{\gamma}_2) \} = (\hat{\theta}_2 - \theta^*) \{ H + o_p(1) \} + O_p(M_2 r_0^2). \]

The desired result then follows because \( H \neq 0 \).

The following lemma establishes the consistency of \( \hat{V} \) for \( V \).

**Lemma S13.** In the setting of Theorem 2, suppose that Assumption 6 and 7 hold and \( M_2 r_0 = o(1) \). Then a consistent estimator of \( V = \text{var} \{ \tau(U; \theta^*, \hat{\alpha}_2, \hat{\gamma}_2) \} / H^2 \) is

\[ \hat{V} = \bar{E} \{ \tau^2(U; \hat{\theta}_2, \hat{\alpha}_2, \hat{\gamma}_2) \} / \hat{H}^2, \]

where \( \hat{H} = \bar{E} \{ \frac{\partial \tau}{\partial \theta} (U; \hat{\theta}_2, \hat{\alpha}_2, \hat{\gamma}_2) \} \).
Proof. First, $\hat{H} = H + o_p(1)$ can be shown similarly as $\hat{E}\{\frac{\partial^2}{\partial \theta^2}(U; \hat{\theta}, \hat{\alpha_2}, \hat{\gamma}_2)\} = H + o_p(1)$ in the proof of Lemma S12. Next, we show that $\hat{G} = G + o_p(1)$, where $G = E\{\tau^2(U; \theta^*, \hat{\alpha_2}, \hat{\gamma}_2)\}$ and $\hat{G} = \hat{E}\{\tau^2(U; \hat{\theta}_2, \hat{\alpha_2}, \hat{\gamma}_2)\}$. Similarly as (S12), consider the decomposition

$$\hat{E}\{\tau^2(U; \hat{\theta}_2, \hat{\alpha_2}, \hat{\gamma}_2)\} = E\{\tau^2(U; \theta^*, \hat{\alpha_2}, \hat{\gamma}_2)\} + \Delta_{51} + \Delta_{52},$$

where

$$\Delta_{51} = \hat{E}\{\tau^2(U; \hat{\theta}_2, \hat{\alpha_2}, \hat{\gamma}_2)\} - E\{\tau^2(U; \hat{\theta}_2, \hat{\alpha_2}, \hat{\gamma}_2)\},$$

$$\Delta_{52} = E\{\tau^2(U; \hat{\theta}_2, \hat{\alpha_2}, \hat{\gamma}_2)\} - E\{\tau^2(U; \theta^*, \hat{\alpha_2}, \hat{\gamma}_2)\}.$$

By Assumption 7(i) and the uniform law of large numbers (Ferguson 1996, Theorem 16), $|\Delta_{51}| \leq \sup_{(\theta, \alpha, \gamma) \in \mathcal{N}_4} |(\hat{E} - E)\{\tau^2(U; \theta, \alpha, \gamma)\}| = o_p(1)$. Moreover, by $|\hat{\theta}_2 - \theta^*| = o_p(1)$ and the continuous mapping theorem, $|\Delta_{52}| = o_p(1)$. Hence $\hat{G} = G + o_p(1)$. □

I.4 Proof of Corollary 4

Assume that $\psi_f \equiv 1$ in model (45) and $\hat{\gamma}_2 = \hat{\gamma}_1$. First, we show that option (i) or (ii) in the discussion preceding Corollary 4 yields $(\hat{\theta}_1, \hat{\alpha}_2) = (\hat{\theta}_0, \hat{\alpha}_1)$ and hence $\hat{\theta}_2 = \hat{\theta}(\hat{\alpha}_1, \hat{\gamma}_1)$, provided that the same Lasso tuning parameter is used in computing $\hat{\alpha}_2$ as in computing $(\hat{\theta}_0, \hat{\alpha}_1)$. Suppose that $(\hat{\theta}_0, \hat{\alpha}_1)$ are Lasso least square estimators as

$$(\hat{\theta}_0, \hat{\alpha}_1) = \arg\min_{(\theta, \alpha)} \left[ \hat{E}\{(Y - \theta Z - \alpha^T \xi)^2\} + \lambda|\theta| + \lambda||\alpha||_1 \right].$$

(S13)

For option (ii), if $(\hat{\theta}_1, \hat{\alpha}_2)$ are redefined as Lasso least square estimators with the same tuning parameter $\lambda$, then $(\hat{\theta}_1, \hat{\alpha}_2) = (\hat{\theta}_0, \hat{\alpha}_1)$ by definition. For option (i), $\hat{\theta}_1$ is replaced by $\hat{\theta}_0$ in (47). If $\hat{\alpha}_2$ is redefined as follows, with the same tuning parameter $\lambda$ as in (S13),

$$\hat{\alpha}_2 = \arg\min_{\alpha} \left[ \hat{E}\{(Y - \hat{\theta}_0 Z - \alpha^T \xi)^2\} + \lambda||\alpha||_1 \right],$$

(S14)

then $\hat{\alpha}_2 = \hat{\alpha}_1$, because for any $\alpha$,

$$\hat{E}\{(Y - \hat{\theta}_0 Z - \alpha^T \xi)^2\} + \lambda|\hat{\theta}_0| + \lambda||\alpha||_1 \geq \hat{E}\{(Y - \hat{\theta}_0 Z - \hat{\alpha}_1^T \xi)^2\} + \lambda|\hat{\theta}_0| + \lambda||\hat{\alpha}_1||_1$$

$$\iff \hat{E}\{(Y - \hat{\theta}_0 Z - \alpha^T \xi)^2\} + \lambda||\alpha||_1 \geq \hat{E}\{(Y - \hat{\theta}_0 Z - \hat{\alpha}_1^T \xi)^2\} + \lambda||\hat{\alpha}_1||_1,$$
and hence $\hat{\alpha}_1$ is also a minimizer to the objective in (S14).

Now suppose that option (ii) is used, i.e., $\hat{\theta}_1$ is redefined as $\hat{\theta}_0$, in our two-step algorithm. Then $\hat{\alpha}_2 = \hat{\alpha}_1$ as shown above. Proposition 3 can be applied with $(\hat{\theta}_1, \hat{\alpha}_2, \hat{\gamma}_2)$ replaced by $(\hat{\theta}_0, \hat{\alpha}_1, \hat{\gamma}_1)$ and $\hat{\theta}_{\text{CAL}} = \hat{\theta}_2$ by $\hat{\theta}(\hat{\alpha}_1, \hat{\gamma}_1)$, because $\hat{\theta}_0$ can be shown to be pointwise doubly robust and hence Assumption 2(iv) is satisfied under the stated regularity conditions. In fact, the target value (i.e., probability limit) $\tilde{\theta}_0$ for $\hat{\theta}_0$, by definition, satisfies $E\{(Y - \tilde{\theta}_0 Z - \tilde{\alpha}_1^T \xi)(Z, \xi^T)\} = 0$, which implies the population doubly robust estimating equation $E\{(Y - \tilde{\theta}_0 Z - \tilde{\alpha}_1^T \xi)(Z - \tilde{\gamma}_1^T \xi)\} = 0$ or equivalently

$$\tilde{\theta}_0 = \frac{E\{(Y - \tilde{\alpha}_1^T \xi)(Z - \tilde{\gamma}_1^T \xi)\}}{E\{Z(Z - \tilde{\gamma}_1^T \xi)\}}.$$ 

Hence $\tilde{\theta}_0$ coincides with $\theta^*$ if model (44) or model (45) with $\psi_f \equiv 1$ is correctly specified. This reasoning is a sample analogue of that in Example 5.
II Additional material for simulation studies

We provide implementation details and additional simulation results.

II.1 Partially linear modeling

We describe the data-generating configurations used for \((Z, X)\), related to Fisher’s discrimination analysis. For setting (C1), we first generate \(Z\) such that \(P(Z = 1) = q\). Next we generate \(X|Z = 1 \sim N(\mu_1, \Sigma)\) and \(X|Z = 0 \sim N(\mu_0, \Sigma)\). Then

\[
P(Z = 1|X) = \frac{1}{1 + \exp(-\beta_0 - \beta_1^T X)},
\]

(S15)

where \(\beta_0 = -\frac{1}{2}\mu_1^T \Sigma^{-1} \mu_1 + \frac{1}{2} \mu_0^T \Sigma^{-1} \mu_0 + \log \left( \frac{q}{1-q} \right) \) and \(\beta_1 = \Sigma^{-1}(\mu_1 - \mu_0)\). In our experiments, we choose \(q = 0.5\), \(\Sigma = I\) (identity matrix), \(\mu_0 = 0\) and \(\mu_1\) a sparse \(p \times 1\) vector with first 5 components being \((-0.25, 0.5, 0.75, 1, 1.25)\), which leads to \(\beta_0 = -\frac{1}{2} \mu_1^T \mu_1 = -0.4297\) and \(\beta_1 = \mu_1 = (-0.25, 0.5, 0.75, 1, 1.25, 0, \ldots, 0)^T\) in (S15). This gives the stated expression of \(P(Z = 1|X)\) in setting (C1).

For setting (C2), we first generate \(Z\) such that \(P(Z = 1) = q\). Next we generate \(X|Z = 1 \sim N(\mu_1, \Sigma_1)\) and \(X|Z = 0 \sim N(\mu_0, \Sigma_0)\). Then

\[
P(Z|X) = \frac{1}{1 + \exp(-\beta_0 - \beta_1^T X - X^T \Omega X)}
\]

(S16)

where \(\beta_0 = -\frac{1}{2} \mu_1^T \Sigma_1^{-1} \mu_1 + \frac{1}{2} \mu_0^T \Sigma_0^{-1} \mu_0 + \log \left( \frac{q}{1-q} \right) \), \(\beta_1 = \Sigma_1^{-1} \mu_1 - \Sigma_0^{-1} \mu_0\), \(\Omega = \frac{1}{2} (\Sigma_0^{-1} - \Sigma_1^{-1})\).

In our experiments, we choose \(q = 0.5\), \(\Sigma_0 = I\), \(\Sigma_1^{-1} = 2I\), \(\mu_0 = 0\), and \(\mu_1\) a sparse \(p \times 1\) vector with first 4 components being \((-0.25, 0.5, 0.75, 1)/2\), which leads to \(\beta_0 = -\mu_1^T \mu_1 + \log(2^{p/2}) = -0.4687 + \frac{p}{2} \log 2\) and \(\beta_1 = 2 \mu_1 = (-0.25, 0.5, 0.75, 1, 0, \ldots, 0)^T\) in (S16). This gives the stated expression of \(P(Z = 1|X)\) in setting (C2).

Our two-step Algorithm 1, specialized to partially linear modeling, is presented in Algorithm S1, including associated commands from R package glmnet (Friedman et al. 2010). Here \(Y\) and \(Z\) are \(n \times 1\) vectors of observations \(\{Y_i : i = 1, \ldots, n\}\) and \(\{Z_i : i = 1, \ldots, n\}\), and \(X\) and \(ZX\) are the design matrix of dimension \(n \times p\) and \(n \times (p + 1)\) with \(i\)th row being \(X_i^T\) and \((Z_i, X_i^T)\) respectively. In Step 7 of Algorithm S1, offset is a vector with components \(\hat{\beta}_1^T Z_i\), and weights is a vector with components \(\psi_f(\hat{\gamma}_1^T X_i) = \expit(\hat{\gamma}_1^T X_i)(1 - \expit(\hat{\gamma}_1^T X_i))\) for \(\psi_f = \expit(\cdot)\). The argument alpha=1 stands for the \(\ell_1\) penalty.
The tuning parameters $\lambda_1, \lambda_2, \lambda_3$ are sequentially selected from 5-fold cross validation, using \texttt{cv.glmnet()} in the R package \texttt{glmnet}. For linear regression we set \texttt{type.measure="MSE"} and for logistic or log-linear regression, we set \texttt{type.measure="deviance"}. By default, there are 100 values of $\lambda$ in the grid search over $\lambda$ (Friedman et al. 2010).

The debiased Lasso method used in our experiments is shown in Algorithm S2, where robust variance estimation is employed (Zhang & Zhang 2014; Van de Geer et al. 2014;
Bühlmann & Van de Geer 2015). Step 3 in Algorithm S2 involves fitting a linear model of Z given X, instead of a logistic model in Algorithm S1.

Table S1 presents simulation results and Figures S1–S3 show QQ plots of estimates and t-statistics for \( n = 400 \) and \( p = 100, 200 \) as well as \( p = 800 \) (for completeness). Comparison between the three methods is similar as discussed in the main paper.

II.2 Partially log-linear modeling

Our two-step Algorithm 1, specialized to partially log-linear modeling, is presented in Algorithm S3, including associated commands from R package \texttt{glmnet}. Because \( Z_i \)'s are binary, a closed-form solution can be obtained from the doubly robust estimating equation:

\[
 e^{-\theta} = \frac{\sum Z_i e^{\alpha^T X_i} (1 - \expit(\gamma^T X_i)) + \sum Z_i (Y_i - e^{\alpha^T X_i}) \expit(\gamma^T X_i)}{\sum Y_i (1 - \expit(\gamma^T X_i))}. 
\]  

(S17)

Steps 7 and 8 are implemented as regularized weighted maximum likelihood estimation, by specifying weights in \texttt{glmnet} as follows: \texttt{weights1} is a vector with components \( e^{\hat{\alpha}^T X_i} \) and \texttt{weights2} is a vector with components \( e^{-\hat{\theta} Z_i \expit(\hat{\gamma}^T X_i)(1 - \expit(\hat{\gamma}^T X_i))} \).

The debiased Lasso method used in our experiments is shown in Algorithm S4, where robust variance estimation is employed. Step 3 is implemented as regularized least square estimation, where \texttt{weights3} is a vector with components \( e^{\theta_0 Z_i + \hat{\alpha}^T X_i} \).

Table S2 presents simulation results and Figure S4–S5 show QQ plots of estimates and t-statistics for \( n = 400 \) and \( p = 100, 200 \) as well as \( p = 800 \) (for completeness). Comparison between the three methods is similar as discussed in the main paper.
Algorithm S3 Two-step algorithm for partially log-linear modeling

1: procedure INITIAL ESTIMATION
2: Compute \((\hat{\theta}_0, \hat{\alpha}_1) = \arg\min_{\theta, \alpha} \left[ \tilde{E} \{-Y(\theta Z + \alpha^T X) + \theta Z + \alpha^T X \} + \lambda_1 (|\theta| + \|\alpha\|_1) \right] \)
   using glmnet(ZX, y=Y, alpha=1, family="poisson").
3: Compute \(\hat{\gamma}_1 = \arg\min_{\gamma} \left[ \tilde{E} \{-Z\gamma^T X + \log(1 + e^{\gamma^T X}) \} + \lambda_2 \|\gamma\|_1 \right] \)
   using glmnet(X, y=Z, alpha=1, family="binomial").
4: Compute \(\hat{\theta}_1\) from (S17) with \(\alpha = \hat{\alpha}_1\) and \(\gamma = \hat{\gamma}_1\).
5: end procedure

6: procedure CALIBRATED ESTIMATION
7: Compute \(\hat{\gamma}_2 = \arg\min_{\gamma} \tilde{E} e^{\hat{\theta}_0 Z + \hat{\alpha}_1 X} \{-Z\gamma^T X + \log(1 + e^{\gamma^T X}) \} + \lambda_3 \|\gamma\|_1 \)
   using glmnet(X, y=Z, alpha=1, weights3, family="gaussian").
8: Compute \(\hat{\alpha}_2 = \arg\min_{\alpha} \tilde{E} \psi'(\hat{\gamma}_2) e^{-\hat{\theta}_1 Z} \{-Y(\hat{\theta}_1 Z + \alpha^T X) + \hat{\theta}_1 Z + \alpha^T X \} + \lambda_4 \|\alpha\|_1 \)
   using glmnet(X, y=Y, alpha=1, offset=theta1*Z, weights2, family="poisson").
9: Compute \(\hat{\theta}_2\) from (S17) with \(\alpha = \hat{\alpha}_2\) and \(\gamma = \hat{\gamma}_2\)
   and \(\hat{V}(\hat{\theta}_2) = \tilde{E}(Y e^{-\hat{\theta}_2 Z - \hat{\alpha}_2^T X} (Z - \expit(\hat{\gamma}_2 X))^2) / \tilde{E}(e^{-\hat{\theta}_2 Z Y Z (Z - \expit(\hat{\gamma}_2 X)))}.\)
10: end procedure

Algorithm S4 Debiased Lasso for log-linear modeling

1: procedure LINEAR PROJECTION
2: Compute \((\hat{\theta}_0, \hat{\alpha}_1) = \arg\min_{\theta, \alpha} \left[ \tilde{E} \{-Y(\theta Z + \alpha^T X) + \theta Z + \alpha^T X \} + \lambda_1 (|\theta| + \|\alpha\|_1) \right] \)
   using glmnet(ZX, y=Y, alpha=1, family="poisson").
3: Compute \(\hat{\gamma}_1 = \arg\min_{\gamma} \tilde{E} \{e^{\hat{\theta}_0 Z + \hat{\alpha}_1 X} (Z - \gamma^T X)^2 \} + \lambda_2 \|\gamma\|_1 \)
   using glmnet(X, y=Z, alpha=1, weights3, family="gaussian").
4: Compute \(\hat{\theta}_{DB} = \hat{\theta}_0 + \tilde{E}( (Y - e^{\hat{\theta}_0 Z + \hat{\alpha}_1 X} (Z - \hat{\gamma}_1 X)) / \tilde{E}(e^{\hat{\theta}_0 Z + \hat{\alpha}_1 X} Z (Z - \hat{\gamma}_1 X)^2) \)
   and \(\hat{V}(\hat{\theta}_{DB}) = \tilde{E}( (Y - e^{\hat{\theta}_0 Z + \hat{\alpha}_1 X} (Z - \hat{\gamma}_1 X))^2) / \tilde{E}(e^{\hat{\theta}_0 Z + \hat{\alpha}_1 X} Z (Z - \hat{\gamma}_1 X)^2) \).
5: end procedure
II.3 Partially logistic modeling

We describe the data-generating configurations used for \((Z, Y, X)\), related to the odds ratio model in Chen (2007). We first generate \(X \sim N(0, \Sigma)\), where \(\Sigma = \text{Toeplitz}(\rho = 0.5)\). Given \(X\), we generate binary variables \((Z, Y)\) according to the probabilities proportional to the entries in the following \(2 \times 2\) table:

| \(Z = 0\) | \(Z = 1\) |
|----------|----------|
| \(Y = 0\) | 1        | \(e^{\beta_1+ h_1(X)}\) |
| \(Y = 1\) | \(e^{\beta_2+ h_2(X)}\) | \(e^{\theta^* + \beta_1 + \beta_2 + h_3(X)}\) |

Here \(\theta^*, \beta_1\) and \(\beta_2\) are the true parameter values and \(h_1(X), h_2(X)\) and \(h_3(X)\) are functions in \(X\) such that \(h_3(X) = h_1(X) + h_2(X)\). The implied conditional probabilities are

\[
P(Y = 1|X, Z) = \expit(\theta^* Z + h_2(X)), \quad \text{(S18)}
\]
\[
P(Z = 1|X, Y = 0) = \expit(\beta_1 + h_1(X)). \quad \text{(S19)}
\]

In our experiments, we set \(\theta^* = 2, \beta_1 = 0.25, \beta_2 = -0.25\), both \(\alpha\) and \(\gamma\) as a sparse vector with first four components being \((-0.25, 0.25, 0.5, 0.75)/2\). The functions \(h_1\) and \(h_2\) are chosen differently, depending on settings (C7)–(C9).

(i) Taking \(h_1(X) = \gamma^T X\) and \(h_2(X) = \alpha^T X\) in (S18)–(S19) leads to the stated expressions for \(P(Y = 1|X, Z)\) and \(P(Z = 1|X, Y = 0)\) in settings (C7).

(ii) Taking \(h_1(X) = \gamma^T X\) and \(h_2(X) = 0.25X_1 + 0.8X_2^2 + \expit(X_3)\) in (S18)–(S19) leads to the stated expressions for \(P(Y = 1|X, Z)\) and \(P(Z = 1|X, Y = 0)\) in settings (C8).

(iii) Taking \(h_2(X) = \alpha^T X\) and \(h_1(X) = 0.25X_1 + 0.8X_2^2 + \expit(X_3)\) in (S18)–(S19) leads to the stated expressions for \(P(Y = 1|X, Z)\) and \(P(Z = 1|X, Y = 0)\) in settings (C9).
**Algorithm S5** Two-step algorithm for partially logistic modeling

1: **procedure** INITIAL ESTIMATION

2: Computes \( \hat{\theta}_0, \hat{\alpha}_1 = \text{argmin}_{\theta, \alpha} \left[ \tilde{E}\{ -Y(\theta Z + \alpha^T X) + \log(1 + e^{\theta Z + \alpha^T X}) \} + \lambda_1 (|\theta| + ||\alpha||_1) \right] \)
   using glmnet(XZ, y=Y, alpha=1, family="binomial").

3: Computes \( \hat{\gamma}_1 = \text{argmin}_\gamma \left[ \tilde{E}_{Y=0}\{ -Z\gamma^T X + \log(1 + e^{\gamma^T X}) \} + \lambda_2 ||\gamma||_1 \right] \)
   using glmnet(X0, y=Z0, alpha=1, family="binomial").

4: Computes \( \hat{\theta}_1 \) from (S20) with \( \alpha = \hat{\alpha}_1 \) and \( \gamma = \hat{\gamma}_1 \).

5: **end procedure**

6: **procedure** CALIBRATED ESTIMATION

7: Computes \( \hat{\gamma}_2 = \text{argmin}_\gamma \left[ \tilde{E}_{e^{-\theta_1 Z Y} \expit(\hat{\alpha}_1^T X)}\{ -Z\gamma^T X + \log(1 + e^{\gamma^T X}) \} + \lambda_3 ||\gamma||_1 \right] \)
   using glmnet(X, y=Z, alpha=1, weights1, family="binomial").

8: Computes \( \hat{\alpha}_2 = \text{argmin}_\alpha \left[ \tilde{E}_{e^{-\theta_2 Z Y} \expit(\hat{\gamma}_2^T X)}\{ -Y\alpha^T X + \log(1 + e^{\alpha^T X}) \} + \lambda_4 ||\alpha||_1 \right] \)
   using glmnet(X, y=Y, alpha=1, weights2, family="binomial").

9: Computes \( \hat{\theta}_2 \) from (S20) with \( \alpha = \hat{\alpha}_2 \) and \( \gamma = \hat{\gamma}_2 \)
   and \( \hat{V}(\hat{\theta}_2) = \frac{\tilde{E}_{2\theta_2 Z Y}(Y - \expit(\hat{\alpha}_2^T X))^2(Z - \expit(\hat{\gamma}_2^T X))^2}{\tilde{E}^2(Z e^{-2\theta_2 Z Y} (Y - \expit(\hat{\alpha}_2^T X))(Z - \expit(\hat{\gamma}_2^T X)))} \).

10: **end procedure**

**Algorithm S6** Debiased Lasso for logistic modeling

1: **procedure** LINEAR PROJECTION

2: Computes \( \hat{\theta}_0, \hat{\alpha}_1 = \text{argmin}_{\theta, \alpha} \left[ \tilde{E}\{ -Y(\theta Z + \alpha^T X) + \log(1 + e^{\theta Z + \alpha^T X}) \} + \lambda_1 (|\theta| + ||\alpha||_1) \right] \)
   using glmnet(XZ, y=Y, alpha=1, family="binomial").

3: Computes \( \hat{\gamma}_1 = \text{argmin}_\gamma \left[ \tilde{E}\{ \expit(\hat{\theta}_0 Z + \hat{\alpha}_1 X)(Z - \gamma^T X)^2 \} + \lambda_2 ||\gamma||_1 \right] \)
   using glmnet(X, y=Z, alpha=1, weights3, family="gaussian").

4: Computes \( \hat{\theta}_\text{DB} = \theta_0 + \frac{\tilde{E}\{(Y - \expit(\hat{\theta}_0 Z + \hat{\alpha}_1 X))(Z - \hat{\gamma}_1^T X) \}}{\tilde{E}\{\expit(\hat{\theta}_0 Z + \hat{\alpha}_1 X) Z(Z - \hat{\gamma}_1^T X) \}} \)
   and \( \hat{V}(\hat{\theta}_\text{DB}) = \frac{\tilde{E}\{(Y - \expit(\hat{\theta}_0 Z + \hat{\alpha}_1 X))^2(Z - \hat{\gamma}_1^T X)^2 \}}{\tilde{E}^2\{\expit(\hat{\theta}_0 Z + \hat{\alpha}_1 X) Z(Z - \hat{\gamma}_1^T X) \}} \).

5: **end procedure**

Our two-step Algorithm 1, specialized to partially logistic modeling, is presented in Algorithm S5, including associated commands from R package glmnet. Because \( Z \)’s are binary,
a closed-form solution can be obtained from the doubly robust estimating equation:

\[
e^{-\theta} = \frac{-\sum_{Z_i=0 \text{ or } Y_i=0} (Y_i - \expit(\alpha^T X_i))(Z_i - \expit(\gamma^T X_i))}{\sum_{Z_i=1 \text{ and } Y_i=1} (1 - \expit(\gamma^T X_i))(1 - \expit(\alpha^T X_i))}.
\] (S20)

In Step 3, the sample average \(\bar{E}_{Y=0}()\) is computed on over the subsample with \(Y_i = 0\), i.e., \{\((Z_i, X_i) : Y_i = 0, i = 1, \ldots, n\}\}. Here \(Z_0\) denotes \(\{Z_i : Y_i = 0\}\) and the \(X_0\) is the design matrix with \(i\)th row being \(\{X_i^T : Y_i = 0, i = 1, \ldots, n\}\). Steps 7 and 8 are implemented as regularized weighted maximum likelihood estimation by specifying weights in \texttt{glmnet} as follows: \texttt{weights1} is a vector with components \(e^{-\hat{\theta}_i Z_i Y_i} \expit(\hat{\alpha}_1 X_i)(1 - \expit(\hat{\alpha}_1^T X_i))\) and \texttt{weights2} is a vector with components \(e^{-\hat{\theta}_i Z_i Y_i} \expit(\hat{\gamma}_2 X_i)(1 - \expit(\hat{\gamma}_2^T X_i))\).

The debiased Lasso method used in our experiments is shown in Algorithm S6, where robust variance estimation is employed. Step 3 is implemented as regularized least square estimation, where \texttt{weights3} is a vector with components \(\expit(\hat{\theta}_0 Z_i + \hat{\alpha}_1 X_i) = \expit(\hat{\theta}_0 Z_i + \hat{\alpha}_1 X_i)(1 - \expit(\hat{\theta}_0 Z_i + \hat{\alpha}_1 X_i))\).

Table S3 presents simulation results and Figures S7–S9 show QQ plots of estimates and \(t\)-statistics for \(n = 400\) and \(p = 100, 200\) as well as \(p = 800\) (for completeness). Comparison between the three methods is similar as discussed in the main paper.
Table S1: Summary of results for partially linear modeling

|                | (C1) Cor Cor | (C2) Cor Miss | (C3) Mis Cor |
|----------------|--------------|---------------|--------------|
|                | $\hat{\theta}_{DB}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_{DB}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ |
| $n = 400, p = 100$ | | | | | | |
| Bias           | 0.004 | 0.006 | 0.005 | -0.003 | -0.004 | -0.003 | 0.103 | 0.049 | 0.003 |
| $\sqrt{\text{Var}}$ | 0.056 | 0.057 | 0.056 | 0.057 | 0.056 | 0.056 | 0.290 | 0.289 | 0.288 |
| $\sqrt{\text{Evar}}$ | 0.054 | 0.054 | 0.051 | 0.054 | 0.054 | 0.053 | 0.326 | 0.322 | 0.322 |
| Cov95          | 0.948 | 0.946 | 0.945 | 0.946 | 0.944 | 0.942 | 0.933 | 0.945 | 0.947 |
| $n = 400, p = 200$ | | | | | | |
| Bias           | 0.006 | 0.004 | 0.004 | 0.009 | 0.008 | 0.009 | 0.156 | 0.061 | 0.004 |
| $\sqrt{\text{Var}}$ | 0.057 | 0.056 | 0.056 | 0.053 | 0.053 | 0.055 | 0.321 | 0.299 | 0.298 |
| $\sqrt{\text{Evar}}$ | 0.054 | 0.054 | 0.055 | 0.054 | 0.055 | 0.054 | 0.335 | 0.325 | 0.325 |
| Cov95          | 0.936 | 0.939 | 0.940 | 0.945 | 0.943 | 0.944 | 0.923 | 0.939 | 0.944 |
| $n = 400, p = 800$ | | | | | | |
| Bias           | 0.006 | 0.006 | 0.007 | 0.013 | 0.012 | 0.012 | 0.283 | 0.082 | 0.004 |
| $\sqrt{\text{Var}}$ | 0.058 | 0.057 | 0.057 | 0.057 | 0.057 | 0.058 | 0.322 | 0.301 | 0.299 |
| $\sqrt{\text{Evar}}$ | 0.056 | 0.058 | 0.058 | 0.062 | 0.062 | 0.063 | 0.338 | 0.332 | 0.329 |
| Cov95          | 0.932 | 0.931 | 0.931 | 0.933 | 0.938 | 0.940 | 0.861 | 0.917 | 0.940 |
Table S2: Summary of results for partially log-linear modeling

|                | (C4) Cor Cor | (C5) Cor Miss | (C6) Miss Cor |
|----------------|--------------|---------------|---------------|
|                | $\hat{\theta}_{DB}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_{DB}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ |
| $n = 600, p = 100$ | | | | | | |
| Bias           | 0.003 | 0.001 | 0.005 | -0.013 | 0.004 | 0.003 | -0.024 | -0.005 | 0.002 |
| $\sqrt{\text{Var}}$ | 0.041 | 0.046 | 0.047 | 0.039 | 0.042 | 0.045 | 0.072 | 0.079 | 0.082 |
| $\sqrt{\text{Evar}}$ | 0.037 | 0.043 | 0.044 | 0.035 | 0.041 | 0.041 | 0.071 | 0.078 | 0.077 |
| Cov95          | 0.948 | 0.948 | 0.943 | 0.917 | 0.944 | 0.944 | 0.916 | 0.943 | 0.945 |
| $n = 600, p = 200$ | | | | | | |
| Bias           | 0.007 | 0.003 | 0.005 | -0.018 | -0.005 | 0.005 | -0.064 | -0.011 | 0.008 |
| $\sqrt{\text{Var}}$ | 0.042 | 0.046 | 0.046 | 0.041 | 0.043 | 0.047 | 0.074 | 0.078 | 0.083 |
| $\sqrt{\text{Evar}}$ | 0.039 | 0.043 | 0.044 | 0.037 | 0.041 | 0.043 | 0.073 | 0.080 | 0.081 |
| Cov95          | 0.946 | 0.939 | 0.942 | 0.892 | 0.927 | 0.935 | 0.866 | 0.942 | 0.944 |
| $n = 600, p = 800$ | | | | | | |
| Bias           | 0.008 | 0.009 | 0.005 | -0.023 | -0.015 | 0.005 | -0.093 | -0.021 | 0.010 |
| $\sqrt{\text{Var}}$ | 0.043 | 0.046 | 0.048 | 0.045 | 0.045 | 0.048 | 0.075 | 0.078 | 0.081 |
| $\sqrt{\text{Evar}}$ | 0.044 | 0.049 | 0.049 | 0.047 | 0.047 | 0.048 | 0.088 | 0.087 | 0.090 |
| Cov95          | 0.941 | 0.939 | 0.941 | 0.851 | 0.922 | 0.927 | 0.700 | 0.943 | 0.942 |
Table S3: Summary of results for partially logistic modeling

|                | (C7) Cor Cor | (C8) Cor Miss | (C9) Miss Cor |
|----------------|--------------|---------------|---------------|
|                | $\hat{\theta}_{DB}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_{DB}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_{DB}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ |
| $n = 600, p = 100$ | 0.035     | 0.029 | 0.025 | 0.011 | 0.012 | 0.011 | 0.134 | 0.072 | 0.015 |
| $\sqrt{\text{Var}}$ | 0.232 | 0.244 | 0.238 | 0.264 | 0.295 | 0.288 | 0.379 | 0.319 | 0.317 |
| $\sqrt{\text{Evar}}$ | 0.225 | 0.233 | 0.233 | 0.295 | 0.287 | 0.287 | 0.369 | 0.315 | 0.315 |
| Cov95 | 0.945 | 0.945 | 0.946 | 0.938 | 0.937 | 0.943 | 0.915 | 0.939 | 0.941 |

|                | (C7) Cor Cor | (C8) Cor Miss | (C9) Miss Cor |
|----------------|--------------|---------------|---------------|
|                | $\hat{\theta}_{DB}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_{DB}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_{DB}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ |
| $n = 600, p = 200$ | 0.047 | 0.053 | 0.037 | 0.042 | 0.063 | 0.043 | 0.179 | 0.061 | 0.019 |
| $\sqrt{\text{Var}}$ | 0.233 | 0.241 | 0.239 | 0.266 | 0.300 | 0.291 | 0.382 | 0.328 | 0.325 |
| $\sqrt{\text{Evar}}$ | 0.227 | 0.238 | 0.235 | 0.299 | 0.278 | 0.288 | 0.369 | 0.316 | 0.317 |
| Cov95 | 0.942 | 0.936 | 0.944 | 0.950 | 0.933 | 0.947 | 0.892 | 0.927 | 0.939 |

|                | (C7) Cor Cor | (C8) Cor Miss | (C9) Miss Cor |
|----------------|--------------|---------------|---------------|
|                | $\hat{\theta}_{DB}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_{DB}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_{DB}$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ |
| $n = 600, p = 800$ | 0.059 | 0.065 | 0.049 | 0.045 | 0.068 | 0.046 | 0.234 | 0.077 | 0.027 |
| $\sqrt{\text{Var}}$ | 0.232 | 0.245 | 0.239 | 0.273 | 0.315 | 0.298 | 0.375 | 0.326 | 0.331 |
| $\sqrt{\text{Evar}}$ | 0.226 | 0.241 | 0.238 | 0.296 | 0.287 | 0.289 | 0.369 | 0.315 | 0.316 |
| Cov95 | 0.941 | 0.936 | 0.940 | 0.946 | 0.938 | 0.949 | 0.882 | 0.920 | 0.938 |
Figure S1: QQ plots of the estimates (first column) and t-statistics (second column) against standard normal ($n = 400, p = 100$) with partially linear modeling.
Figure S2: QQ plots of the estimates (first column) and $t$-statistics (second column) against standard normal ($n = 400, p = 200$) for partially linear modeling.

![QQ plots](image-url)
Figure S3: QQ plots of the estimates (first column) and t-statistics (second column) against standard normal ($n = 400, p = 800$) for partially linear modeling.

Mis. Y|Z,X, Cor. Z|X

Cor. Y|Z,X, Mis. Z|X

Cor. Y|Z,X, Cor. Z|X

Cor. Y|Z,X, Cor. Z|X
Figure S4: QQ plots of the estimates (first column) and t-statistics (second column) against standard normal ($n = 600$, $p = 100$) for partially log-linear modeling.

![QQ plots](image-url)
Figure S5: QQ plots of the estimates (first column) and t-statistics (second column) against standard normal ($n = 600$, $p = 200$) for partially log-linear modeling.
Figure S6: QQ plots of the estimates (first column) and t-statistics (second column) against standard normal ($n = 600$, $p = 800$) for partially log-linear modeling.
Figure S7: QQ plots of the estimates (first column) and t-statistics (second column) against standard normal ($n = 600, p = 100$) for partially logistic modeling
Figure S8: QQ plots of the estimates (first column) and t-statistics (second column) against standard normal ($n = 600$, $p = 200$) for partially logistic modeling.
Figure S9: QQ plots of the estimates (first column) and t-statistics (second column) against standard normal ($n = 600$, $p = 800$) for partially logistic modeling.
Appendix References

Chen, H. Y. (2007). A semiparametric odds ratio model for measuring association. *Biometrics*, 63:413–421.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33:1–22.