Supplementary Material for
“Model-assisted inference for treatment effects using regularized calibrated estimation with high-dimensional data”

Zhiqiang Tan

The Supplementary Material contains Appendices I–III.

I Additional results for simulation study

I.1 Results for simulation setup

Denote by $\phi()$ the probability density function and $\Phi()$ the cumulative distribution function for $N(0, 1)$. Let $X \sim N(0, 1)$ and $W = X + (X + 1)^2$. Then $X^\dagger = \{W - E(W)\}/\sqrt{\text{var}(W)}$ by standardization. We need to compute $E(W) = E\{(X + 1)^2\}$ and $E(W^2) = 1 + E\{2X(X + 1)^2 + (X + 1)^4\}$, where

- $E\{(x + 1)^2\} = \int_{-\infty}^{\infty} (x^2 + 2x + 1)\phi(x)dx$,
- $E\{x(x + 1)^2\} = \int_{-1}^{\infty} x(x^2 + 2x + 1)\phi(x)dx$,
- $E\{(x + 1)^4\} = \int_{-1}^{\infty} (x^4 + 4x^3 + 6x^2 + 4x + 1)\phi(x)dx$.

The expectations involved above can be computed using the following results.

- $\int x\phi(x)dx = -\phi(x)$,
- $\int x^2\phi(x)dx = \frac{1}{2}\{2\Phi(x) - 1\} - x\phi(x)$,
- $\int x^3\phi(x)dx = -\phi(x)(x^2 + 2)$,
- $\int x^4\phi(x)dx = \frac{3}{2}\{2\Phi(x) - 1\} - x(x^2 + 3)\phi(x)$.

I.2 Additional simulation results

Figures S1 and S2 show the boxplots of $(X_1, \ldots, X_4)$ by $T$ and the scatterplots of $Y$ against $(X_1, \ldots, X_4)$ within $\{T = 1\}$. The logistic propensity score model and linear outcome model with the main effects of $(X_1, \ldots, X_4)$ appear reasonable from these plots.

Table S1 summarizes the results from 1000 repeated simulations, including the oracle estimators (non-penalized, using sub-models with only the covariates $X_1, \ldots, X_4$) and also the
results in Table 1 with \( n = 800 \) and \( p = 200 \) or 1000 for completeness. Figures S3–S6 present the QQ plots of the point estimates and their associated t-statistics.

There are interesting features from these results. The oracle estimators based on ML or CAL lead to confidence intervals with coverage proportions roughly close to the nominal probabilities in all three cases (C1)–(C3), even though asymptotic exact coverage does not in general hold for the ML-based method unless both PS and OR models are correctly specified (Kim & Haziza 2014; Vermeulen & Vansteelandt 2015). On one hand, confidence intervals from the ML-based method using asymptotic expansion like (8) are known to be conservative (with over-coverage) if the PS model is correctly specified, but the OR model is misspecified. This is manifested in the relatively high coverage proportions based on ML in the case (C2) in Table S1. On the other hand, the impact of model misspecification on the coverage accuracy of ML-based confidence intervals may be limited under the data configurations here, especially when the OR model is correctly specified but the PS model is misspecified.

The RML and RCAL estimators achieve similar variances to each other and to those of the oracle estimators while their absolute biases tend to increase, as \( p \) ranges from 200 to 1000. As mentioned in Section 4, the RCAL estimator has noticeably smaller absolute biases than RML in the case (C2), correct PS and misspecified OR models. The post-Lasso refitting method RML2 appears to achieve close coverage proportions to the nominal probabilities, but at the cost of increasingly large variances as \( p \) increases in all three cases (C1)–(C3).

We also experimented with RCAL with post-Lasso refitting, where \( \hat{\pi}_{\text{RCAL}}^1 \) is first replaced by the fitted value from post-Lasso refitting of the PS model, then the weighted least-squares loss \( \ell_{WL}(\alpha^1; \hat{\gamma}^1_{\text{RCAL}}) \) for fitting the OR model is modified with the weights \( (1 - \hat{\pi}^1_{\text{RCAL}})/\hat{\pi}^1_{\text{RCAL}} \) from PS refitting, the Lasso estimation is performed for the OR model with the modified weighted least-squares loss, and \( \hat{m}^1_{\text{RML}} \) is replaced by the fitted value from post-Lasso refitting of the OR model. The relationship between RCAL with and without post-Lasso refitting is not as straightforward as that in the case of RML, because of the weights involved, \( (1 - \hat{\pi}^1_{\text{RCAL}})/\hat{\pi}^1_{\text{RCAL}} \). From the simulation results (not shown), there does not seem to be any advantage of this post-Lasso version of RCAL, in bias, variance, or coverage accuracy.

Further research would be needed to study the effect of post-Lasso refitting when used with RML or RCAL and investigate possible remedies.
Figure S1: Boxplots of \((X_1, \ldots, X_4)\) within \(\{T = 0\}\) and \(\{T = 1\}\) and scatterplots of \(Y\) against \((X_1, \ldots, X_4)\) within \(\{T = 1\}\) from a sample of size \(n = 800\) in data configuration (C2)
Figure S2: Boxplots of \((X_1, \ldots, X_4)\) within \(\{T = 0\}\) and \(\{T = 1\}\) and scatterplots of \(Y\) against \((X_1, \ldots, X_4)\) within \(\{T = 1\}\) from a sample of size \(n = 800\) in data configuration (C3)
Table S1: Summary of results with linear outcome models

<table>
<thead>
<tr>
<th></th>
<th>cor PS, cor OR</th>
<th>cor PS, mis OR</th>
<th>mis PS, cor OR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RML</td>
<td>RML2</td>
<td>RCAL</td>
</tr>
<tr>
<td><strong>n = 800 and p = 4 (non-penalized)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>.004</td>
<td>—</td>
<td>.004</td>
</tr>
<tr>
<td>$\sqrt{\text{Var}}$</td>
<td>.070</td>
<td>—</td>
<td>.070</td>
</tr>
<tr>
<td>$\sqrt{\text{EVar}}$</td>
<td>.073</td>
<td>—</td>
<td>.071</td>
</tr>
<tr>
<td>Cov90</td>
<td>.918</td>
<td>—</td>
<td>.915</td>
</tr>
<tr>
<td>Cov95</td>
<td>.950</td>
<td>—</td>
<td>.948</td>
</tr>
<tr>
<td><strong>n = 800 and p = 200</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>.019</td>
<td>.002</td>
<td>.026</td>
</tr>
<tr>
<td>$\sqrt{\text{Var}}$</td>
<td>.070</td>
<td>.079</td>
<td>.070</td>
</tr>
<tr>
<td>$\sqrt{\text{EVar}}$</td>
<td>.068</td>
<td>.076</td>
<td>.068</td>
</tr>
<tr>
<td>Cov90</td>
<td>.864</td>
<td>.880</td>
<td>.854</td>
</tr>
<tr>
<td>Cov95</td>
<td>.922</td>
<td>.947</td>
<td>.914</td>
</tr>
<tr>
<td><strong>n = 800 and p = 400</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>.024</td>
<td>.000</td>
<td>.027</td>
</tr>
<tr>
<td>$\sqrt{\text{Var}}$</td>
<td>.072</td>
<td>.082</td>
<td>.073</td>
</tr>
<tr>
<td>$\sqrt{\text{EVar}}$</td>
<td>.067</td>
<td>.080</td>
<td>.068</td>
</tr>
<tr>
<td>Cov90</td>
<td>.861</td>
<td>.896</td>
<td>.845</td>
</tr>
<tr>
<td>Cov95</td>
<td>.924</td>
<td>.946</td>
<td>.910</td>
</tr>
<tr>
<td><strong>n = 800 and p = 1000</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>.031</td>
<td>—</td>
<td>.033</td>
</tr>
<tr>
<td>$\sqrt{\text{Var}}$</td>
<td>.068</td>
<td>.109</td>
<td>.070</td>
</tr>
<tr>
<td>$\sqrt{\text{EVar}}$</td>
<td>.067</td>
<td>.112</td>
<td>.067</td>
</tr>
<tr>
<td>Cov90</td>
<td>.851</td>
<td>.911</td>
<td>.836</td>
</tr>
<tr>
<td>Cov95</td>
<td>.922</td>
<td>.949</td>
<td>.915</td>
</tr>
</tbody>
</table>

Note: RML denotes $\hat{\mu}^{1}(\hat{m}_{RML}^{1}, \hat{\pi}_{RML}^{1})$, and RML2 denotes the variant with $\hat{m}_{RML}^{1}$ and $\hat{\pi}_{RML}^{1}$ replaced by the fitted values obtained by refitting OR and PS models only including the variables selected from the corresponding Lasso estimation. RCAL denotes $\hat{\mu}^{1}(\hat{m}_{RCAL}^{1}, \hat{\pi}_{RCAL}^{1})$. Bias and Var are the Monte Carlo bias and variance of the points estimates. EVar is the mean of the variance estimates, and hence $\sqrt{\text{EVar}}$ also measures the $L_{2}$-average of lengths of confidence intervals. Cov90 or Cov95 is the coverage proportion of the 90% or 95% confidence intervals.
Figure S3: QQ plots of the estimates (first row) and t-statistics (second row) against standard normal ($n = 800, p = 4$), based on the non-penalized versions of $\hat{\mu}_{1}(\hat{m}_{1}^{RML}, \hat{\pi}_{1}^{RML})$ ($\circ$) and $\hat{\mu}_{1}(\hat{m}_{1}^{RWL}, \hat{\pi}_{1}^{RCAL})$ ($\times$). For readability, only a subset of 101 order statistics are shown as points on the QQ lines.
Figure S4: QQ plots of the estimates (first row) and t-statistics (second row) against standard normal (n = 800, p = 200), based on \(\hat{\mu}(\hat{m}_{\text{RML}}, \hat{\pi}_{\text{RML}})\) (○), the post-Lasso variant (△) and \(\hat{\mu}(\hat{m}_{\text{RWL}}, \hat{\pi}_{\text{RCAL}})\) (×). For readability, only a subset of 101 order statistics are shown as points on the QQ lines.
Figure S5: QQ plots of the estimates (first row) and t-statistics (second row) against standard normal $(n = 800, p = 400)$, based on $\hat{\mu}(\hat{m}_{RML}, \hat{\pi}_{RML})$, the post-Lasso variant $(\triangle)$, and $\hat{\mu}(\hat{m}_{RWL}, \hat{\pi}_{RCAL})$ $(\times)$. For readability, only a subset of 101 order statistics are shown as points on the QQ lines.
Figure S6: QQ plots of the estimates (first row) and \( t \)-statistics (second row) against standard normal (\( n = 800, p = 1000 \)), based on \( \hat{\mu}'(\hat{m}_{\text{Lasso}}, \hat{\pi}_{\text{Lasso}}) \), the post-Lasso variant (\( \triangle \)), and \( \hat{\mu}'(\hat{m}_{\text{RML}}, \hat{\pi}_{\text{RCAL}}) \) (\( \times \)). For readability, only a subset of 101 order statistics are shown as points on the QQ lines.
II Additional results for empirical application

Figure S7: Boxplots of inverse probability weights within the treated (left) and untreated (middle) groups, each normalized to sum to the sample size \( n \), and QQ plots with a 45-degree line of the standardized sample influence functions based on \( \varphi(Y, T, X; \cdot) \) in (7) for ATE (right).

Figure S8: Boxplots of inverse probability weights and QQ plots of the standardized sample influence functions as in Figure S7, but for comparison between RML with post-Lasso refitting (RML post) and RCAL. For RML post, there are 5 of the weights within the treated which are larger than 40 (the largest being 296.3) and hence not shown, and 5 of the weights within the untreated which are larger than 15 (the largest being 20.4) and hence not shown. For RML post, there are 7 values of the influence function which fall outside \((-10, 10)\), ranging from \(-22.8\) to \(44.0\).
III Technical details

III.1 Inside Theorem 1

The following result (ii) is taken from Tan (2017), Lemma 1(ii), and result (i) can be shown similarly using Lemma 14 in Section III.10 and the union bound.

Lemma 1. (i) Denoted by Ω₀ the event that
\[ \sup_{j=0,1,...,p} \left| \tilde{E} \left[ \left\{ -Te^{-h_{\text{CAL}}^1(X)} + (1 - T) \right\} f_j(X) \right] \right| \leq \lambda_0. \]
Under Assumption 1(i)–(ii), if \( \lambda_0 \geq \sqrt{2}(e^{-B_0} + 1)C_0\sqrt{\log\{(1 + p)/\epsilon\}/n} \), then \( P(\Omega_0) \geq 1 - 2\epsilon \).

(ii) Denote by Ω₁ the event that
\[ \sup_{j,k=0,1,...,p} |(\tilde{\Sigma}_\gamma)_{jk} - (\Sigma_\gamma)_{jk}| \leq \lambda_0, \quad (S1) \]
Under Assumption 1(i)–(ii), if \( \lambda_0 \geq (4e^{-B_0}C^2_0)\sqrt{\log\{(1 + p)/\epsilon\}/n} \), then \( P(\Omega_1) \geq 1 - 2\epsilon^2 \).

Take \( \lambda_0 = C_{01}\sqrt{\log\{(1 + p)/\epsilon\}/n} \) with
\[ C_{01} = \max \left\{ \sqrt{2}(e^{-B_0} + 1)C_0, 4e^{-B_0}C^2_0 \right\}. \]
Then under the conditions of Theorem 1, inequality (34) holds in the event \( \Omega_0 \cap \Omega_1 \), with probability at least \( 1 - 4\epsilon \), by the proof of Tan (2017, Corollary 2).

III.2 Probability lemmas

Lemma 2. Denote by Ω₂ the event that
\[ \sup_{j=0,1,...,p} \left| \tilde{E} \left[ Tw(X; \tilde{\gamma}_{\text{CAL}}^1)\{Y - \tilde{m}_{\text{WL}}^1(X)\}f_j(X) \right] \right| \leq \lambda_1. \quad (S2) \]
Under Assumptions 1(i)–(ii) and 2(i), if \( \lambda_1 \geq (e^{-B_0}C_0)\sqrt{8(D^4_0 + D^2_3)}\sqrt{\log\{(1 + p)/\epsilon\}/n} \), then \( P(\Omega_2) \geq 1 - 2\epsilon \).

Proof. Let \( Z_j = Tw(X; \tilde{\gamma}_{\text{CAL}}^1)\{Y - \tilde{m}_{\text{WL}}^1(X)\}f_j(X) \) for \( j = 0,1,...,p \). Then \( E(Z_j) = 0 \) by the definition of \( \tilde{\alpha}_{\text{WL}}^1 \). Under Assumption 1(i)–(ii), \( |Z_j| \leq e^{-B_0}C_0|T\{Y - \tilde{m}_{\text{WL}}^1(X)\}| \). By Assumption 2(i), the variables \( (Z_0,Z_1,\ldots,Z_p) \) are uniformly sub-gaussian: \( \max_{j=0,1,...,p} D^2_2 E\{\exp(Z_j^2/D^2_2) - 1\} \leq D^2_3 \), with \( D_2 = e^{-B_0}C_0D_0 \) and \( D_3 = e^{-B_0}C_0D_1 \). Therefore, Lemma 2(i) holds by Lemma 15 in Section III.10 and the union bound. \( \Box \)

Denote \( \Sigma_{\alpha_2} = E[Tw(X; \tilde{\gamma}_{\text{CAL}}^1)\{Y - \tilde{m}_{\text{WL}}^1(X)\}^2f(X)f^T(X)] \), and \( \tilde{\Sigma}_{\alpha_2} = \tilde{E}[Tw(X; \tilde{\gamma}_{\text{CAL}}^1)\{Y - \tilde{m}_{\text{WL}}^1(X)\}^2f(X)f^T(X)] \), the sample version of \( \tilde{\Sigma}_{\alpha_2} \).
**Lemma 3.** Denote by $\Omega_3$ the event that

$$\sup_{j,k=0,1,...,p} |(\tilde{\Sigma}_{a2})_{jk} - (\Sigma_{a2})_{jk}| \leq (D_0^2 + D_0D_1)\lambda_0, \quad (S3)$$

Under Assumptions 1(i)–(ii) and 2(i), if

$$(D_0^2 + D_0D_1)\lambda_0 \geq 4e^{-B_0C_0^2} \left[ D_0^2 \log \{1 + p\}/\epsilon \right]/n + D_0D_1\sqrt{\log \{1 + p\}/\epsilon}/n,$$

then $P(\Omega_3) \geq 1 - 2e^2$.

**Proof.** For any $j, k = 0, 1, \ldots, p$, the variable $Tw(X; \tilde{\gamma}_{\text{CAL}}^1)\{Y - \tilde{m}_{\text{WL}}(X)\}^2 f_j(X)f_k(X)$ is the product of $w(X; \tilde{\gamma}_{\text{CAL}}^1)f_j(X)f_k(X)$ and $T\{Y - \tilde{m}_{\text{WL}}(X)\}^2$, where $|w(X; \tilde{\gamma}_{\text{CAL}}^1)f_j(X)f_k(X)| \leq e^{-B_0C_0^2}$ by Assumptions 1(i)–(ii) and $T\{Y - \tilde{m}_{\text{WL}}(X)\}$ is sub-gaussian by Assumption 2(i). Applying Lemmas 16 and 18 in Section III.10 yields

$$P \{ |(\tilde{\Sigma}_{a2})_{jk} - (\Sigma_{a2})_{jk}| > 2e^{-B_0C_0^2}D_0^2t + 2e^{-B_0C_0^2}D_0D_1t\sqrt{2t} \} \leq 2\frac{e^2}{(1 + p)^2},$$

for $j, k = 0, 1, \ldots, p$, where $t = \log \{(1 + p)^2/\epsilon^2\}/n$. The result then follows from the union bound. \qed

Denote $\Sigma_{a1} = E[Tw(X; \tilde{\gamma}_{\text{CAL}}^1)|Y - \tilde{m}_{\text{WL}}(X)|f(X)f^T(X)]$, and $\tilde{\Sigma}_{a1} = \tilde{E}[Tw(X; \tilde{\gamma}_{\text{CAL}}^1)|Y - \tilde{m}_{\text{WL}}(X)|f(X)f^T(X)]$, the sample version of $\Sigma_{a1}$.

**Lemma 4.** Denote by $\Omega_4$ the event that

$$\sup_{j,k=0,1,...,p} |(\tilde{\Sigma}_{a1})_{jk} - (\Sigma_{a1})_{jk}| \leq \sqrt{D_0^2 + D_1^2}\lambda_0, \quad (S4)$$

Under Assumptions 1(i)–(ii) and 2(i), if $\lambda_0 \geq 4e^{-B_0C_0^2}\sqrt{\log \{(1 + p)/\epsilon\}/n}$, then $P(\Omega_3) \geq 1 - 2e^2$.

**Proof.** The variables $Tw(X; \tilde{\gamma}_{\text{CAL}}^1)|Y - \tilde{m}_{\text{WL}}(X)|f_j(X)f_k(X)$ for $j, k = 0, 1, \ldots, p$ are uniformly sub-gaussian, because $|w(X; \tilde{\gamma}_{\text{CAL}}^1)f_j(X)f_k(X)| \leq e^{-B_0C_0^2}$ by Assumptions 1(i)–(ii) and $T\{Y - \tilde{m}_{\text{WL}}(X)\}$ is sub-gaussian by Assumption 2(i). Applying Lemma 15 yields

$$P \{ |(\tilde{\Sigma}_{a1})_{jk} - (\Sigma_{a1})_{jk}| > t \} \leq 2\frac{e^2}{(1 + p)^2},$$

for $j, k = 0, 1, \ldots, p$, where $t = e^{-B_0C_0^2}\sqrt{8(D_0^2 + D_1^2)}\sqrt{\log \{(1 + p)^2/\epsilon^2\}/n}$. The result then follows from the union bound. \qed

Denote $\Sigma_0 = E[f(X)f^T(X)]$ and $\tilde{\Sigma}_0 = \tilde{E}[f(X)f^T(X)]$, the sample version of $\Sigma_0$. 

12
Lemma 5. Denote by $\Omega_5$ the event that
\[
\sup_{j,k=0,1,\ldots,p} |(\tilde{\Sigma}_0)_{jk} - (\Sigma_0)_{jk}| \leq e^{B_0} \lambda_0,
\]
(S5)

Under Assumption 1(i), if $\lambda_0 \geq 4e^{-B_0}C_0^2 \sqrt{\log\left\{(1 + p)/\epsilon\right\}/n}$, then $P(\Omega_5) \geq 1 - 2\epsilon^2$.

Proof. This result follows directly from Lemma 14 and the union bound, with $|f_j(X)f_k(X)| \leq C_2^0$ and hence $|f_j(X)f_k(X) - (\Sigma_0)_{jk}| \leq 2C_0^2$.

III.3 Proof of Theorems 2 and 5

Thoughout this section, suppose that Assumption 1 holds. The proof of Theorem 5 is completed by combining Lemmas 2–3 and 6–12. Theorem 2 is a special case of Theorem 5, where Assumptions 3(ii)–(iv) are satisfied with $C_2 = 1$ and $C_3 = \eta_2 = \eta_3 = 0$.

Lemma 6. For any coefficient vector $\alpha^1$ and $h(X) = \alpha^1 f(X)$, we have
\[
\begin{aligned}
D_{\text{RL}}^\dagger(\hat{h}_{\text{RL}}^1,h;\hat{\gamma}_{\text{RCAL}}) &+ \lambda\|\hat{\alpha}^1_{\text{RL},1:p}\|_1 \\
&\leq (\hat{\alpha}^1_{\text{RL}} - \alpha^1)^T \tilde{E} \left[Tw(X;\hat{\gamma}_{\text{RCAL}})\{Y - m_1(X;\alpha^1)\}f(X)\right] + \lambda\|\alpha^1_{1:p}\|_1.
\end{aligned}
\]
(S6)

Proof. For any $u \in (0,1]$, the definition of $\hat{\alpha}^1_{\text{RL}}$ implies
\[
\begin{aligned}
\ell_{\text{RL}}(\hat{\alpha}^1_{\text{RL}};\hat{\gamma}_{\text{RCAL}}) &+ \lambda\|\hat{\alpha}^1_{\text{RL},1:p}\|_1 \\
&\leq \ell_{\text{RL}}\{(1 - u)\hat{\alpha}^1_{\text{RL}} + u\alpha^1;\hat{\gamma}_{\text{RCAL}}\} + \lambda\|(1 - u)\hat{\alpha}^1_{\text{RL},1:p} + u\alpha^1_{1:p}\|_1,
\end{aligned}
\]
which, by the convexity of $\| \cdot \|_1$, gives
\[
\ell_{\text{RL}}(\hat{\alpha}^1_{\text{RL}};\hat{\gamma}_{\text{RCAL}}) - \ell_{\text{RL}}\{(1 - u)\hat{\alpha}^1_{\text{RL}} + u\alpha^1;\hat{\gamma}_{\text{RCAL}}\} + \lambda u\|\hat{\alpha}^1_{\text{RL},1:p}\|_1 \leq \lambda u\|\alpha^1_{1:p}\|_1.
\]
Dividing both sides of the preceding inequality by $u$ and letting $u \to 0+$ yields
\[
-\tilde{E} \left[Tw(X;\hat{\gamma}_{\text{RCAL}})\{Y - \hat{m}_{\text{RL}}^1(X)\}\{\hat{h}_{\text{RL}}^1(X) - h(X)\}\right] + \lambda\|\hat{\alpha}^1_{\text{RL},1:p}\|_1 \leq \lambda\|\alpha^1_{1:p}\|_1,
\]
which leads to (S6) after simple rearrangement using (54).

Lemma 7. In the event $\Omega_0 \cap \Omega_1$, we have
\[
\tilde{E} \left[Tw(X;\hat{\gamma}_{\text{CAL}})\{\hat{h}_{\text{CAL}}^1(X) - \hat{h}_{\text{CAL}}^1(X)\}\right]^2 \leq e^{\eta_0} M_0|S_\gamma|\lambda_0^2,
\]
(S7)
and for any function \(h(X)\),
\[
D_{WL}^1(\hat{h}_{RLW}^1, h; \hat{\gamma}_{CAL}^1) \geq e^{-\eta_0} D_{WL}^1(\hat{h}_{RLW}^1, h; \bar{\gamma}_{CAL}^1),
\]
(S8)
where \(\eta_0 = (A_0 - 1)^{-1}M_0\eta_0 C_0\).

Proof. By direct calculation from the definition of \(D_{CAL}()\), we find
\[
D_{CAL}^1(\bar{h}_{RLC}^1, \bar{h}_{CAL}^1) = -\hat{E}\left[T\left\{e^{-h_{RLC}^1(X) - h_{CAL}^1(X)}\{\hat{h}_{RLC}^1(X) - \bar{h}_{CAL}^1(X)\}\right\}\right]
= \hat{E}\left[T e^{-u(\tilde{\gamma}_{RLC}^1 - \tilde{\gamma}_{CAL}^1)^T f(X)} {w(X; \tilde{\gamma}_{CAL}^1)(\hat{h}_{RLC}^1(X) - \bar{h}_{CAL}^1(X))}^2\right]
\]
for some \(u \in (0, 1)\), where the second step uses the mean value theorem,
\[
e^{-h_{RLC}^1(X)} - e^{-h_{CAL}^1(X) - (1-u)h_{RLC}^1(X)}(\tilde{\gamma}_{RLC}^1 - \tilde{\gamma}_{CAL}^1)^T f(X).
\]
(S9)
In the event \(\Omega_0 \cap \Omega_1\) that (34) holds, we have
\[
\|\tilde{\gamma}_{RLC} - \tilde{\gamma}_{CAL}^1\| \leq (A_0 - 1)^{-1}M_0|S_\gamma|\lambda_0 \leq (A_0 - 1)^{-1}M_0\eta_0,
\]
(S10)
by Assumption 1(iv), \(|S_\gamma|\lambda_0 \leq \eta_0\), and hence
\[
M_0|S_\gamma|\lambda_0^2 \geq D_{CAL}(\bar{h}_{RLC}^1, \bar{h}_{CAL}^1) \geq e^{-\eta_0} \hat{E}\left[T w(X; \tilde{\gamma}_{CAL}^1)\{\hat{h}_{RLC}^1(X) - \bar{h}_{CAL}^1(X)\}^2\right],
\]
which gives the desired inequality (S7). In addition, we write
\[
D_{WL}^1(\hat{h}_{RLW}^1, h; \hat{\gamma}_{CAL}^1)
= \hat{E}\left[T w(X; \tilde{\gamma}_{CAL}^1)\left\{\psi(\hat{h}_{RLW}^1(X)) - \psi(h(X))\right\}\{\hat{h}_{RLW}^1(X) - h(X)\}\right]
= \hat{E}\left[T e^{-\hat{\gamma}_{RLW}^1(X)} f(X)\{\psi(\hat{h}_{RLW}^1(X)) - \psi(h(X))\}\{\hat{m}_{RLW}^1(X) - h(X)\}\right],
\]
which, in the event \(\Omega_0 \cap \Omega_1\), yields inequality (S8) by (S10) and Assumption 1(i).

For two functions \(h(x)\) and \(h'(x)\), denote
\[
Q_{WL}(h', h; \gamma) = \hat{E}\left[T w(X; \gamma)\{h'(X) - h(X)\}^2\right].
\]

Lemma 8. Take \(\alpha^1 = \tilde{\alpha}_{WL}^1\) and \(h(X) = \tilde{\alpha}_{WL}^{1T} f(X)\). Suppose that Assumption 2(i) holds. Then in the event \(\Omega_0 \cap \Omega_1 \cap \Omega_3\), (S6) implies
\[
e^{-\eta_0} D_{WL}^1(\hat{h}_{RLW}^1, h; \hat{\gamma}_{CAL}^1) + \lambda\|\tilde{\alpha}_{RLW,1,p}\|_1
\leq (\tilde{\alpha}_{RLW} - \alpha_1^1)^T \hat{E}\left[T w(X; \tilde{\gamma}_{CAL}^1)\{Y - m_1(X; \alpha_1^1)\} f(X)\right] + \lambda\|\alpha_1^1\|_1
\]
\[
+ e^{\eta_0} (M_0|S_\gamma|\lambda_0^2)^{1/2} \{Q_{WL}(\hat{h}_{RLW}^1, h; \hat{\gamma}_{CAL}^1)^{1/2},
\]
where \(M_0 = (D_0^2 + D_1^2)(e^{\eta_0} M_0 + \eta_0) + (D_0^2 + D_0 D_1)\eta_0\), and \(\eta_0 = (A_0 - 1)^{-2}M_0^2\eta_0\).
Proof. Consider the following decomposition,

\[
(\hat{a}_{\text{RWL}} - \alpha^1)^T \tilde{E} \left[ T \omega(X; \gamma^1_{\text{CAL}}) \{ Y - m_1(X; \alpha^1) \} f(X) \right] \\
= (\hat{a}_{\text{RWL}} - \alpha^1)^T \tilde{E} \left[ T \omega(X; \gamma^1_{\text{CAL}}) \{ Y - m_1(X; \alpha^1) \} f(X) \right] \\
+ \tilde{E} \left\{ e^{-\hat{h}^1_{\text{CAL}}(X)} - e^{-h^1_{\text{CAL}}(X)} \right\} \left\{ Y - m_1(X; \alpha^1) \right\} \{ \hat{h}^1_{\text{RWL}}(X) - h(X) \},
\]

(S11)
denoted as \( \Delta_1 + \Delta_2 \). By the mean value equation (S9) and the Cauchy–Schwartz inequality, the second term \( \Delta_2 \) can be bounded from above as

\[
\Delta_2 \leq e^{C_0 \| \hat{\gamma}_{\text{CAL}} - \gamma^1_{\text{CAL}} \|^2_1} \times \tilde{E}^{1/2} \left[ T \omega(X; \gamma^1_{\text{CAL}}) \{ \hat{h}^1_{\text{RWL}}(X) - h(X) \}^2 \right] \\
\times \tilde{E}^{1/2} \left[ T \omega(X; \gamma^1_{\text{CAL}}) \{ Y - m_1(X; \alpha^1) \}^2 \{ \hat{h}^1_{\text{CAL}}(X) - h^1_{\text{CAL}}(X) \}^2 \right].
\]

(S12)
We upper-bound the third term on the right hand side in several steps. First, in the event \( \Omega_3 \), we have by inequality (S3),

\[
(\tilde{E} - E) \left[ T \omega(X; \gamma^1_{\text{CAL}}) \{ Y - m_1(X; \alpha^1) \}^2 \{ \hat{h}^1_{\text{CAL}}(X) - h^1_{\text{CAL}}(X) \}^2 \right] \\
\leq (D_0^2 + D_0 D_1) \lambda_0 \| \gamma^1_{\text{CAL}} - \hat{\gamma}_{\text{CAL}} \|^2_1,
\]
where, by some abuse of notation, \( (\tilde{E} - E)(Z) \) denotes \( n^{-1} \sum_{i=1}^n (Z_i - E(Z)) \) for a variable \( Z \) that is a function of \( (T, Y, X) \). Second, by Assumption 2(i) and Lemma 17, \( E[\| Y^1 - m_1(X; \alpha^1) \|^2 | X] \leq D_0^2 + D_1^2 \) and hence

\[
E \left[ T \omega(X; \gamma^1_{\text{CAL}}) \{ Y - m_1(X; \alpha^1) \}^2 \{ \hat{h}^1_{\text{CAL}}(X) - h^1_{\text{CAL}}(X) \}^2 \right] \\
\leq (D_0^2 + D_1^2) E \left[ T \omega(X; \gamma^1_{\text{CAL}}) \{ \hat{h}^1_{\text{CAL}}(X) - h^1_{\text{CAL}}(X) \}^2 \right].
\]

Third, in the event \( \Omega_1 \), we have by inequality (S1),

\[
(E - \tilde{E}) \left[ T \omega(X; \gamma^1_{\text{CAL}}) \{ \hat{h}^1_{\text{CAL}}(X) - \hat{h}^1_{\text{CAL}}(X) \}^2 \right] \leq \lambda_0 \| \gamma^1_{\text{CAL}} - \hat{\gamma}_{\text{CAL}} \|^2_1.
\]
Combining the preceding inequalities, we have in the event \( \Omega_1 \cap \Omega_3 \),

\[
\tilde{E} \left[ T \omega(X; \gamma^1_{\text{CAL}}) \{ Y - m_1(X; \alpha^1) \}^2 \{ \hat{h}^1_{\text{CAL}}(X) - h^1_{\text{CAL}}(X) \}^2 \right] \\
\leq (D_0^2 + D_0 D_1) \lambda_0 \| \gamma^1_{\text{CAL}} - \hat{\gamma}_{\text{CAL}} \|^2_1 \\
+ (D_0^2 + D_1^2) \left\{ \lambda_0 \| \gamma^1_{\text{CAL}} - \hat{\gamma}_{\text{CAL}} \|^2_1 + \tilde{E} \left[ T \omega(X; \gamma^1_{\text{CAL}}) \{ \hat{h}^1_{\text{CAL}}(X) - \hat{h}^1_{\text{CAL}}(X) \}^2 \right] \right\}.
\]

(S13)
The desired result follows by collecting inequalities (S11)–(S13) and applying (S7), (S8) and (S10) in the event \( \Omega_0 \cap \Omega_1 \).

\( \Box \)
Lemma 9. Denote $b = \hat{\alpha}_{\text{RWL}}^1 - \hat{\alpha}_{\text{WL}}^1$. Suppose that Assumption 2(i) holds. In the event $\Omega_0 \cap \Omega_1 \cap \Omega_2 \cap \Omega_3$, we have
\[
e^{-\eta_0} D_{\text{WL}}^\dagger (\hat{h}_{\text{RWL}}^1, \hat{h}_{\text{WL}}^1; \bar{\gamma}_{\text{CAL}}^1) + (A_1 - 1) \lambda_1 \|b\|_1
\leq e^{\eta_0} \left( M_01|S_\gamma|\lambda_0^2 \right)^{1/2} \{ Q_{\text{WL}}(\hat{h}_{\text{RWL}}^1, \hat{h}_{\text{WL}}^1; \bar{\gamma}_{\text{CAL}}^1) \}^{1/2} + 2A_1 \lambda_1 \sum_{j \in S_\alpha} |b_j|. \quad (S14)\]

Proof. In the event $\Omega_2$, we have
\[
b^T \hat{E} \left[ Tw(X; \bar{\gamma}_{\text{CAL}}^1) \{ Y - \bar{m}_{\text{WL}}^1(X) \} f(X) \right] \leq \lambda_1 \|b\|_1.
\]
From this bound and Lemma 8 with $\alpha^1 = \hat{\alpha}_{\text{WL}}^1$, we have in the event $\Omega_0 \cap \Omega_1 \cap \Omega_2 \cap \Omega_3$,
\[
e^{-\eta_0} D_{\text{WL}}^\dagger (\hat{h}_{\text{RWL}}^1, \hat{h}_{\text{WL}}^1; \bar{\gamma}_{\text{CAL}}^1) + A_1 \lambda_1 \|\hat{\alpha}_{\text{RWL},1;P}\|_1
\leq \lambda_1 \|b\|_1 + A_1 \lambda_1 \|\hat{\alpha}_{\text{WL},1;P}\|_1 + e^{\eta_0} \left( M_01|S_\gamma|\lambda_0^2 \right)^{1/2} \{ Q_{\text{WL}}(\hat{h}_{\text{RWL}}^1, \hat{h}_{\text{WL}}^1; \bar{\gamma}_{\text{CAL}}^1) \}^{1/2}.
\]
Applying to the preceding inequality the identity $|\hat{\alpha}_{\text{RWL},j}^1| = |\hat{\alpha}_{\text{WL},j}^1 - \hat{\alpha}_{\text{WL},j}^1|$ for $j \notin S_\alpha$ and the triangle inequality
\[
|\hat{\alpha}_{\text{RWL},j}^1| \geq |\hat{\alpha}_{\text{WL},j}^1| - |\hat{\alpha}_{\text{WL},j}^1 - \hat{\alpha}_{\text{WL},j}^1|, \quad j \in S_\alpha \setminus \{0\},
\]
and rearranging the result gives
\[
e^{-\eta_0} D_{\text{WL}}^\dagger (\hat{h}_{\text{RWL}}^1, \hat{h}_{\text{WL}}^1; \bar{\gamma}_{\text{CAL}}^1) + (A_1 - 1) \lambda_1 \|b_{\text{WL},1;P}\|_1
\leq \lambda_1 \|b_0\| + 2A_1 \lambda_1 \sum_{j \in S_\alpha \setminus \{0\}} |b_j| + e^{\eta_0} \left( M_01|S_\gamma|\lambda_0^2 \right)^{1/2} \{ Q_{\text{WL}}(\hat{h}_{\text{RWL}}^1, \hat{h}_{\text{WL}}^1; \bar{\gamma}_{\text{CAL}}^1) \}^{1/2}.
\]
The conclusion follows by adding $(A_0 - 1) \lambda_0 \|b_0\|$ to both sides above. \qed

Denote $\hat{\Sigma}_\alpha = \hat{E}[Tw(X; \bar{\gamma}_{\text{CAL}}^1) \psi_2(\hat{h}_{\text{WL}}^1(X)) f(X) f^T(X)]$.

Lemma 10. Suppose that Assumption 3(iii) holds. Then for any $h = \alpha^{\text{T}} f$ and $h' = \alpha'^{\text{T}} f$,
\[
D_{\text{WL}}^\dagger (h, h'; \bar{\gamma}_{\text{CAL}}^1) \geq \frac{1 - e^{-C_4 \|b\|_1}}{C_4 \|b\|_1} \left( b^T \hat{\Sigma}_\alpha b \right),
\]
where $b = \alpha'' - \alpha^1$ and $C_4 = C_0 C_3$. Throughout, set $(1 - e^{-c})/c = 1$ for $c = 0$.

Proof. Set $\gamma = \bar{\gamma}_{\text{CAL}}^1$. By direct calculation, we have
\[
D_{\text{WL}}^\dagger (h, h'; \gamma) = \hat{E} \left[ Tw(X; \gamma) \left[ \psi(h'(X)) - \psi(h(X)) \right] \{ h'(X) - h(X) \} \right]
= \hat{E} \left[ Tw(X; \gamma) \left( \int_0^1 \psi_2 \left[ h(X) + u \{ h'(X) - h(X) \} \right] du \right) \{ h'(X) - h(X) \}^2 \right].
\]
By Assumption 3(iii) and the fact that \(|h'(X) - h(X)| \leq \{\sup_{j=0,1,...,p} |f_j(X)|\} \|\alpha'' - \alpha^1\|_1 \leq C_0 \|\alpha'' - \alpha^1\|_1\) by Assumption 1(i), it follows that
\[
D_{WL}^\dagger(h, h'; \gamma) \geq \hat{E} \left[ T w(X; \gamma) \left( \int_0^1 \psi_2 \{h(X)\} e^{-C_3 u|h'(X) - h(X)|} du \right) \{h'(X) - h(X)\}^2 \right] \\
\geq \hat{E} \left[ T w(X; \gamma) \psi_2 \{h(X)\} \{h'(X) - h(X)\}^2 \right] \left( \int_0^1 e^{-C_4 u\|\alpha'' - \alpha^1\|_1} du \right),
\]
which gives the desired result because \(\int_0^1 e^{-cu} du = (1 - e^{-c})/c\) for \(c \geq 0\). \(\Box\)

**Lemma 11.** Suppose that Assumption 2(iii) holds. In the event \(\Omega_1\), Assumption 2(ii) implies a compatibility condition for \(\tilde{\Sigma}_\gamma\): for any vector \(b = (b_0, b_1, \ldots, b_p)^T \in \mathbb{R}^{1+p}\) such that \(\sum_{j \notin S_\alpha} |b_j| \leq \xi_1 \sum_{j \in S_\alpha} |b_j|\), we have
\[
(1 - \eta_1) \nu_1^2 \left( \sum_{j \in S_\alpha} |b_j| \right)^2 \leq |S_\alpha| \left( b^T \tilde{\Sigma}_\gamma b \right).
\]  \(\text{(S15)}\)

**Proof.** In the event \(\Omega_1\), we have \(|b^T (\Sigma_\gamma - \Sigma_\gamma) b| \leq \lambda_1 \|b\|^2_1\) by (S1). Then Assumption 2(ii) implies that for any vector \(b = (b_0, b_1, \ldots, b_p)^T\) satisfying \(\sum_{j \notin S_\alpha} |b_j| \leq \xi_1 \sum_{j \in S_\alpha} |b_j|\),
\[
\nu_1^2 \|b_{S_\alpha}\|^2_1 \leq |S_\alpha| (b^T \Sigma_\gamma b) \leq |S_\alpha| \left( b^T \tilde{\Sigma}_\gamma b + \lambda_0 \|b\|^2_1 \right) \\
\leq |S_\alpha| (b^T \tilde{\Sigma}_\gamma b) + |S_\alpha| \lambda_1 (1 + \xi_1)^2 \|b_{S_\alpha}\|^2_1,
\]
where \(\|b_{S_\alpha}\|_1 = \sum_{j \in S_\alpha} |b_j|\). The last inequality uses \(|b_1| \leq (1 + \xi_1)\|b_{S_\alpha}\|_1\). Then (S15) follows because \((1 + \xi_1)^2 \nu_1^2 |S_\alpha| \lambda_1 \leq \eta_1 (< 1)\) by Assumption 2(iii). \(\Box\)

**Lemma 12.** Suppose that Assumptions 2 and 3 hold, and \(A_1 > (\xi_1 + 1)/(\xi_1 - 1)\). In the event \(\Omega_0 \cap \Omega_1 \cap \Omega_2 \cap \Omega_3\), inequality (37) holds as in Theorem 2.

**Proof.** Denote \(b = \hat{\alpha}_{RWL}^1 - \tilde{\alpha}_{WL}^1\), \(D_{WL}^\dagger = D_{WL}^\dagger(\hat{h}_{RWL}^1, \tilde{h}_{WL}^1; \gamma_{CAL})\), \(Q_{WL} = Q_{WL}(\hat{h}_{RWL}^1, \tilde{h}_{WL}^1; \gamma_{CAL})\), and
\[
D_{WL}^\dagger = e^{-\eta_{01}} D_{WL}^\dagger + (A_1 - 1) \lambda_1 \|b\|_1.
\]
In the event \(\Omega_0 \cap \Omega_1 \cap \Omega_2 \cap \Omega_3\), inequality (S14) from Lemma 9 with Assumption 2(i) leads to two possible cases: either
\[
\xi_2 D_{WL}^\dagger \leq e^{\eta_{01}} \left( M_{01} |S_\gamma| \lambda_0^2 \right)^{1/2} \left( Q_{WL} \right)^{1/2},
\]  \(\text{(S16)}\)
or \(1 - \xi_2) D_{WL}^\dagger \leq 2 A_1 \lambda_1 \sum_{j \in S_\alpha} |b_j|\), that is,
\[
D_{WL}^\dagger \leq (\xi_1 + 1)(A_1 - 1) \lambda_1 \sum_{j \in S_\alpha} |b_j| = \xi_3 \lambda_1 \sum_{j \in S_\alpha} |b_j|,
\]  \(\text{(S17)}\)

17
where $\xi_2 = 1 - 2A_1/(\xi_1 + 1)(A_1 - 1) \in (0, 1]$ because $A_1 > (\xi_1 + 1)/(\xi_1 - 1)$ and $\xi_3 = (\xi_1 + 1)(A_1 - 1)$. We deal with the two cases separately as follows.

If (S17) holds, then $\sum_{j \not\in S_\alpha} |b_j| \leq \xi_1 \sum_{j \in S_\alpha} |b_j|$, which, by Lemma 11 and Assumption 2(ii)–(iii), implies (S15), that is,

$$
\sum_{j \in S_\alpha} |b_j| \leq (1 - \eta_1)^{-1/2} \nu_1^{-1} |S_\alpha|^{1/2} \left( b^T \tilde{\Sigma}_\gamma b \right)^{1/2}.
$$

(S18)

By Assumption 3(ii) and Lemma 10 with Assumption 3(iii), we have

$$
D_{WL}^\dagger \geq 1 - \frac{e^{-C_4 \|b\|_1}}{C_4 \|b\|_1} \left( b^T \tilde{\Sigma}_\alpha b \right) \geq 1 - \frac{e^{-C_4 \|b\|_1}}{C_4 \|b\|_1} C_2 \left( b^T \tilde{\Sigma}_\gamma b \right).
$$

(S19)

Combining (S17), (S18), and (S19) and using $D_{WL}^\dagger \leq e^{\eta_1} D_{WL}^\dagger$ yields

$$
D_{WL}^\dagger \leq e^{\eta_1} \xi_3^2 (1 - \eta_1)^{-1} \nu_1^{-2} C_2^{-1} |S_\alpha| \lambda_1^2 \frac{C_4 \|b\|_1}{1 - e^{-C_4 \|b\|_1}}.
$$

(S20)

But $(A_1 - 1) \lambda_1 \|b\|_1 \leq D_{WL}^\dagger$. Inequality (S20) along with Assumption 3(iv) implies that $1 - e^{-C_4 \|b\|_1} \leq C_4 (A_1 - 1)^{-1} \xi_3^2 (1 - \eta_1)^{-1} \nu_1^{-2} C_2^{-1} |S_\alpha| \lambda_1 \leq \eta_2 (< 1)$. As a result, $C_4 \|b\|_1 \leq -\log(1 - \eta_2)$ and hence

$$
\frac{1 - e^{-C_4 \|b\|_1}}{C_4 \|b\|_1} = \int_0^1 e^{-C_4 \|b\|_1} u \, du \geq e^{-C_4 \|b\|_1} \geq 1 - \eta_2.
$$

From this bound, inequality (S20) then leads to $D_{WL}^\dagger \leq e^{\eta_1} \xi_3^2 \nu_3^{-2} |S_\alpha| \lambda_1^2$.

If (S16) holds, then simple manipulation using $D_{WL}^\dagger \leq e^{\eta_1} D_{WL}^\dagger$ and (S19) together with $Q_{WL} = b^T \tilde{\Sigma}_\gamma b$ gives

$$
D_{WL}^\dagger \leq e^{3\eta_1} \xi_2^{-2} C_2^{-1} (M_{01} |S_\gamma| \lambda_0^2) \frac{C_4 \|b\|_1}{1 - e^{-C_4 \|b\|_1}}.
$$

(S21)

Similarly as above, using $(A_1 - 1) \lambda_1 \|b\|_1 \leq D_{WL}^\dagger$ and inequality (S21) along with Assumption 3(iv), we find $1 - e^{-C_4 \|b\|_1} \leq C_4 e^{3\eta_1} (A_1 - 1)^{-1} \xi_2^{-2} C_2^{-1} (M_{01} |S_\gamma| \lambda_0) \leq \eta_3 (< 1)$. As a result, $C_4 \|b\|_1 \leq -\log(1 - \eta_3)$ and hence

$$
\frac{1 - e^{-C_4 \|b\|_1}}{C_4 \|b\|_1} = \int_0^1 e^{-C_4 \|b\|_1} u \, du \geq e^{-C_4 \|b\|_1} \geq 1 - \eta_3.
$$

From this bound, inequality (S21) then leads to $D_{WL}^\dagger \leq e^{3\eta_1} \xi_4^{-2} (M_{01} |S_\gamma| \lambda_0^2)$. Therefore, (55) holds through (S16) and (S17) in the event $\Omega_0 \cap \Omega_1 \cap \Omega_2 \cap \Omega_3$. □
III.4 Proof of Theorem 3

Denote \( \hat{\varphi} = \varphi(T, Y, X; \hat{m}_{\text{RWL}}^1, \hat{\pi}_{\text{RCAL}}^1) \) and \( \varphi = \varphi(T, Y, X; m_{\text{RWL}}^1, \pi_{\text{RCAL}}^1) \). Then

\[
\hat{\mu}^1(\hat{m}_{\text{RWL}}^1, \hat{\pi}_{\text{RCAL}}^1) = \hat{\mu}^1(m_{\text{RWL}}^1, \pi_{\text{RCAL}}^1) + \hat{E}(\hat{\varphi} - \varphi).
\]

Consider the following decomposition,

\[
\hat{\varphi} - \varphi = \{m_{\text{RWL}}^1(X) - \hat{m}_{\text{RWL}}^1(X)\} \left\{ 1 - \frac{T}{\pi_{\text{CAL}}^1(X)} \right\} + T\{Y - \hat{m}_{\text{RWL}}^1(X)\} \left\{ \frac{1}{\pi_{\text{RCAL}}^1(X)} - \frac{1}{\pi_{\text{CAL}}^1(X)} \right\} + \{\hat{m}_{\text{RWL}}^1(X) - \hat{m}_{\text{RWL}}^1(X)\} \left\{ \frac{T}{\pi_{\text{CAL}}^1(X)} - \frac{T}{\pi_{\text{RCAL}}^1(X)} \right\}.
\]

denoted as \( \delta_1 + \delta_2 + \delta_3 \).

We show that in the event \( \Omega_0 \cap \Omega_1 \cap \Omega_2 \cap \Omega_3 \cap \Omega_4 \), inequality (40) holds as in Theorem 3. The estimator \( \hat{\mu}^1(\hat{m}_{\text{RWL}}^1, \hat{\pi}_{\text{RCAL}}^1) \) can be decomposed as

\[
\hat{\mu}^1(\hat{m}_{\text{RWL}}^1, \hat{\pi}_{\text{RCAL}}^1) = \hat{\mu}^1(m_{\text{RWL}}^1, \pi_{\text{RCAL}}^1) + \Delta_1 + \Delta_2,
\]

where

\[
\Delta_1 = \hat{E}(\delta_1 + \delta_3) = (\hat{\alpha}_{\text{RWL}}^1 - \bar{\alpha}_{\text{RWL}}^1)^T \hat{\pi}_{\text{RCAL}}^1 \hat{E} \left[ \left\{ 1 - \frac{T}{\pi_{\text{CAL}}^1(X)} \right\} f(X) \right],
\]

\[
\Delta_2 = \hat{E}(\delta_2) = \hat{E} \left[ T\{Y - \hat{m}_{\text{RWL}}^1(X)\} \left\{ \frac{1}{\pi_{\text{RCAL}}^1(X)} - \frac{1}{\pi_{\text{CAL}}^1(X)} \right\} \right].
\]

In the event \( \Omega_0 \cap \Omega_1 \cap \Omega_2 \cap \Omega_3 \), we have

\[
|\Delta_1| \leq (A_1 - 1)^{-1} M_1(|S_\gamma| \lambda_0 + |S_\alpha| \lambda_1) \times A_0 \lambda_0,
\]

by inequality (37) and the Karush–Kuhn–Tucker conditions (15)–(16). Moreover, a Taylor expansion for \( \Delta_2 \) yields for some \( u \in (0, 1) \),

\[
\Delta_2 = -(\bar{\gamma}_{\text{RCAL}}^1 - \hat{\gamma}_{\text{RCAL}}^1)^T \hat{\pi}_{\text{RCAL}}^1 \hat{E} \left[ T\{Y - \hat{m}_{\text{RWL}}^1(X)\} e^{-\hat{h}_{\text{RCAL}}^1(X)f(X)} \right] + (\bar{\gamma}_{\text{RCAL}}^1 - \hat{\gamma}_{\text{RCAL}}^1)^T \hat{\pi}_{\text{RCAL}}^1 \hat{E} \left[ T\{Y - \hat{m}_{\text{RWL}}^1(X)\} e^{-u\hat{h}_{\text{RCAL}}^1(X) - (1-u)\hat{h}_{\text{CAL}}^1(X)f(X)f^T(X)} \right] (\hat{\gamma}_{\text{RCAL}}^1 - \hat{\gamma}_{\text{RCAL}}^1)/2,
\]

denoted as \( \Delta_{21} + \Delta_{22} \). In the event \( (\Omega_0 \cap \Omega_1) \cap \Omega_2 \), we have

\[
|\Delta_{21}| \leq (A_0 - 1)^{-1} M_0 |S_\gamma| \lambda_0 \times \lambda_1,
\]

by inequalities (34) and (S2). The term \( \Delta_{22} \) can be bounded as

\[
|\Delta_{22}| \leq e^{\|\hat{h}_{\text{RCAL}}^1 - \hat{\gamma}_{\text{RCAL}}^1\|_1 C_0} \hat{E} \left[ T\{X; \hat{\gamma}_{\text{RCAL}}^1\} |Y - \hat{m}_{\text{RWL}}^1(X)| \{\hat{h}_{\text{RCAL}}^1(X) - \hat{h}_{\text{CAL}}^1(X)\}^2 \right] / 2.
\]
In the event $\Omega_1 \cap \Omega_4$, we have

$$\tilde{E} \left[ Tw(X; \tilde{\gamma}_1^{\text{cal}}) | Y - \tilde{m}_1^{\text{WL}}(X) | \{ \tilde{h}_1^{\text{RCAL}}(X) - \tilde{h}_1^{\text{CAL}}(X) \}^2 \right]$$

$$\leq \sqrt{D_0^2 + D_1^2 \lambda_0 \| \tilde{\gamma}_1^{\text{RCAL}} - \tilde{\gamma}_1^{\text{CAL}} \|^2}$$

$$+ \sqrt{D_0^2 + D_1^2} \left\{ \lambda_0 \| \tilde{\gamma}_1^{\text{RCAL}} - \tilde{\gamma}_1^{\text{CAL}} \|^2 + \tilde{E} \left[ Tw(X; \tilde{\gamma}_1^{\text{cal}}) | \tilde{h}_1^{\text{RCAL}}(X) - \tilde{h}_1^{\text{CAL}}(X) \}^2 \right] \right\}, \quad (S26)$$

by inequalities (S1) and (S4) and similar steps as in the proof of (S13). Then (40) follows by collecting inequalities (S23)–(S26) and applying (S7) and (S10) in the event $\Omega_0 \cap \Omega_1$.

### III.5 Proof of Theorem 4

Using $a^2 - b^2 = 2(a - b)b + (a - b)^2$ and the Cauchy–Schwartz inequality, we find

$$\left| \tilde{E} \left[ \tilde{\phi}_c^2 - \tilde{\phi}_e^2 \right] \right| \leq 2\tilde{E}^{1/2} \left( \tilde{\phi}_c^2 \right) \tilde{E}^{1/2} \left( \left( \tilde{\phi}_c - \tilde{\phi}_e \right)^2 \right) + \tilde{E} \left( \left( \tilde{\phi}_c - \tilde{\phi}_e \right)^2 \right). \quad (S27)$$

Using $\tilde{\phi}_c = \tilde{\phi} - \mu^{\text{rwl}}(\tilde{m}_1^{\text{rwl}}, \tilde{\pi}_1^{\text{RCAL}})$ and $\tilde{\phi}_e = \tilde{\phi} - \mu^{\text{rwl}}(\tilde{m}_1^{\text{WL}}, \tilde{\pi}_1^{\text{CAL}})$, we find

$$\tilde{E} \left( \left( \tilde{\phi}_c - \tilde{\phi}_e \right)^2 \right) \leq 2\tilde{E} \left( \left( \tilde{\phi} - \tilde{\phi}_c \right)^2 \right) + 2\tilde{E} \left( \left( \mu^{\text{rwl}}(\tilde{m}_1^{\text{rwl}}, \tilde{\pi}_1^{\text{RCAL}}) - \mu^{\text{rwl}}(\tilde{m}_1^{\text{WL}}, \tilde{\pi}_1^{\text{CAL}}) \right)^2 \right). \quad (S28)$$

To control $\tilde{E} \left( \left( \tilde{\phi} - \tilde{\phi}_c \right)^2 \right)$, we use the decomposition (S22), denoted as $\delta_1 + \delta_2 + \delta_3$.

First, by the mean value equation (S9) and Assumption 1(i)–(ii), we have

$$\tilde{E}(\delta_2^2) = \tilde{E} \left[ T \left\{ Y - \tilde{m}_1^{\text{WL}}(X) \right\}^2 \right] \left\{ \frac{1}{\tilde{\pi}_1^{\text{RCAL}}(X)} - \frac{1}{\tilde{\pi}_1^{\text{CAL}}(X)} \right\}^2$$

$$\leq e^{-B_0 + 2\| \tilde{\gamma}_1^{\text{RCAL}} - \tilde{\gamma}_1^{\text{CAL}} \|_1} C_0 \tilde{E} \left[ Tw(X; \tilde{\gamma}_1^{\text{cal}}) \left\{ Y - \tilde{m}_1^{\text{WL}}(X) \right\}^2 \right] \left\{ \tilde{h}_1^{\text{RCAL}}(X) - \tilde{h}_1^{\text{CAL}}(X) \right\}^2 \right]. \quad (S29)$$

Second, writing $\left\{ \tilde{\pi}_1^{\text{RCAL}}(X) \right\}^{-1} - \left\{ \tilde{\pi}_1^{\text{CAL}}(X) \right\}^{-1} = e^{-\tilde{h}_1^{\text{CAL}}(X)} \left( e^{-\tilde{h}_1^{\text{RCAL}}(X)} + \tilde{h}_1^{\text{CAL}}(X) - 1 \right)$ and using Assumption 1(i)–(ii), we have

$$\tilde{E}(\delta_2^2) \leq e^{-B_0} \left( 1 + e^{\| \tilde{\gamma}_1^{\text{RCAL}} - \tilde{\gamma}_1^{\text{CAL}} \|_1} C_0 \right)^2 \tilde{E} \left[ Tw(X; \tilde{\gamma}_1^{\text{cal}}) \left\{ \tilde{m}_1^{\text{rwl}}(X) - \tilde{m}_1^{\text{WL}}(X) \right\}^2 \right]. \quad (S30)$$

Third, using Assumption 1(i)–(ii), we also have

$$\tilde{E}(\delta_2^2) \leq e^{-B_0} \left( \frac{T}{\tilde{\pi}_1^{\text{CAL}}(X)} \right)^2 \left\{ \frac{1}{\tilde{\pi}_1^{\text{RCAL}}(X)} - \frac{1}{\tilde{\pi}_1^{\text{CAL}}(X)} \right\}^2$$

$$\leq (1 + e^{-B_0}) \tilde{E} \left[ \left\{ \tilde{h}_1^{\text{rwl}}(X) - \tilde{h}_1^{\text{WL}}(X) \right\}^2 \right]$$

$$\leq (1 + e^{-B_0}) C_0 \tilde{E} \left[ \tilde{\alpha}_1^{\text{rwl}} - \tilde{\alpha}_1^{\text{WL}} \right]^2. \quad (S31)$$
Inequality (41) follows by collecting inequalities (S27)--(S32) and applying (37), (40), (S10), and (S13) in the event \( \Omega_0 \cap \Omega_1 \cap \Omega_2 \cap \Omega_3 \cap \Omega_4 \). If condition (35) holds, then we have in the event \( \Omega_1 \cap \Omega_5 \),

\[
\tilde{E} \left[ \{ \hat{h}_{\text{RWL}}(X) - \hat{h}_{\text{WL}}(X) \}^2 \right] \leq e^{B_0} \lambda_0 \| \hat{\alpha}_{\text{RWL}} - \hat{\alpha}_{\text{WL}} \|^2_1 \\
+ \tau_0^{-1} \left\{ \lambda_0 \| \hat{\alpha}_{\text{RWL}} - \hat{\alpha}_{\text{WL}} \|^2_1 + \tilde{E} \left[ T_w(X; \hat{\alpha}_{\text{WL}}) \{ \hat{h}_{\text{RWL}}(X) - \hat{h}_{\text{WL}}(X) \}^2 \right] \right\},
\]

(S33)

by inequalities (S1) and (S5) and similar steps as in the proof of (S13). Inequality (42) follows, similarly as (41), by combining inequalities (S27)--(S31) and (S33).

III.6 Proof of (43)--(44)

Denote \( a_0 = \min \{ e^{B_0}/(1 + e^{B_0}), e^{B_0}/(1 + e^{B_0}) \} \) and \( a_1 = \max(B_1, B_1') \).

To prove (43), we show that for any \( \gamma \),

\[
E \left[ \left\{ 1/\pi^{1}_{\text{CAL}}(X) - 1/\pi^*(X) \right\}^2 \right] \leq \frac{5e^{a_1}}{3a_0^2} E \left[ \left\{ 1/\pi(X; \gamma) - 1/\pi^*(X) \right\}^2 \right].
\]

(S34)

Let \( h^*(X) = \log\{ \pi^*(X)/(1 - \pi^*(X)) \} \). Then by Propositions 1 and 3 in Tan (2017),

\[
E \left\{ \ell_{\text{CAL}}(\bar{\gamma}^{1}_{\text{CAL}}) - \kappa_{\text{CAL}}(h^*) \right\} \geq \frac{3a_0}{5} E \left[ \left\{ \frac{\pi^*(X)}{\pi(X; \bar{\gamma}^{1}_{\text{CAL}}) - 1} - 1 \right\}^2 \right]
\]

\[
\geq \frac{3a_0^3}{5} E \left[ \left\{ 1/\pi(X; \bar{\gamma}^{1}_{\text{CAL}}) - 1/\pi^*(X) \right\}^2 \right].
\]

By the definition of \( \bar{\gamma}^{1}_{\text{CAL}} \), the above implies that

\[
E \left[ \left\{ 1/\pi^{1}_{\text{CAL}}(X) - 1/\pi^*(X) \right\}^2 \right] \leq \frac{5}{3a_0^2} E \left\{ \ell_{\text{CAL}}(\bar{\gamma}^{1}_{\text{CAL}}) - \kappa_{\text{CAL}}(h^*) \right\}
\]

\[
\leq \frac{5}{3a_0^2} E \left\{ \ell_{\text{CAL}}(\gamma) - \kappa_{\text{CAL}}(h^*) \right\}.
\]

(S35)

On the other hand, by direct calculation with \( h(X) = \gamma^* f(X) \),

\[
\ell_{\text{CAL}}(\gamma) - \kappa_{\text{CAL}}(h^*) = D_{\text{CAL}}(h, h^*)
\]

\[
\leq D_{\text{CAL}}(h, h^*) = -\tilde{E} \left[ T \left\{ e^{-h(X)} - e^{-h^*(X)} \right\} \{ h(X) - h^*(X) \} \right].
\]

Taking expectations and using the mean value theorem gives

\[
E \left\{ \ell_{\text{CAL}}(\gamma) - \kappa_{\text{CAL}}(h^*) \right\} \leq e^{a_1} E \left[ \left\{ e^{-h(X)} - e^{-h^*(X)} \right\}^2 \right].
\]

(S36)

Combining (S35) and (S36) yields (S34).
To prove (44), we show that for any $\alpha^1$,
\[
E \left[ (\bar{m}_{WL}^1(X) - m^{1*}(X))^2 \right] \leq \frac{e^{B_1}}{a_0} E \left[ (m^1(X; \alpha^1) - m^{1*}(X))^2 \right]. \tag{S37}
\]

By the definition of $\alpha^1_{WL}$, we have
\[
E \left[ T \frac{1 - \bar{\pi}_{CAL}(X)}{\pi_{CAL}(X)} (m^{1*}(X) - \bar{m}_{WL}^1(X))^2 \right] \\
\leq E \left[ T \frac{1 - \bar{\pi}_{CAL}(X)}{\pi_{CAL}(X)} (m^{1*}(X) - m^1(X; \alpha^1))^2 \right].
\]

The left hand side is bounded from below by $a_0 e^{-B_1} E[|\bar{m}_{WL}^1(X) - m^{1*}(X)|^2]$, and the right hand side is bounded from above by $a_0^{-1} E[|m^1(X; \alpha^1) - m^{1*}(X)|^2]$. Combining these leads to (S37).

### III.7 Proof of Proposition 2

To prove (45), it suffices by Theorem 3 to show that
\[
|\bar{E}(\varphi - \varphi^\dagger)| \leq b_\gamma O_p(n^{-1/2}) + b_\gamma b_{\alpha 1}, \tag{S38}
\]

where $\varphi = \varphi(T, Y, X; \bar{m}_{WL}^1, \bar{\pi}_{CAL})$ and $\varphi^\dagger = \varphi(T, Y, X; m_{WL}^1, \pi^*)$. Consider the following decomposition,
\[
\varphi - \varphi^* = T\{Y - m^{1*}(X)\} \left\{ \frac{1}{\bar{\pi}_{CAL}(X)} - \frac{1}{\pi^*(X)} \right\} \\
+ \{\bar{m}_{WL}^1(X) - m^{1*}(X)\} \{ \frac{T}{\pi^*(X)} - \frac{T}{\pi_{CAL}(X)} \},
\]
denoted as $\delta_2 + \delta_3$. By Cauchy–Schwartz inequality, $|\bar{E}(\delta_2)| \leq b_\gamma b_{\alpha 1}$. Moreover, by Assumption 2(i) and unconfoundedness, we have
\[
E \left\{ \bar{E}^2(\delta_2) \mid (T_i, X_i)_{i=1,...,n} \right\} \leq (D_0^2 + D_1^2) \frac{b_\gamma^2}{n}.
\]

By Markov inequality, for any $c > 0$,
\[
P \left\{ |\bar{E}(\delta_2)| \geq c \sqrt{D_0^2 + D_1^2} b_\gamma n^{-1/2} \mid (T_i, X_i)_{i=1,...,n} \right\} \leq \frac{1}{c^2}.
\]

Then it also holds that $P \{|\bar{E}(\delta_2)| \geq c \sqrt{D_0^2 + D_1^2} b_\gamma n^{-1/2} \} \leq \frac{1}{c^2}$, and hence $\bar{E}(\delta_2) = \sqrt{D_0^2 + D_1^2} b_\gamma n^{-1/2} O_p(1)$. Consequently (S38) follows.

Similarly, to prove (46), it suffices to show that
\[
|\bar{E}(\varphi - \varphi^\dagger)| \leq b_{\alpha 1} O_p(n^{-1/2}) + b_\gamma b_{\alpha 1}, \tag{S39}
\]
where \( \varphi^+ = \varphi(T, Y, X; m^{1*}, \bar{\pi}_\text{CAL}) \). Consider the following decomposition,

\[
\bar{\varphi} - \varphi^+ = \{\bar{m}_{WL}(X) - m^{1*}(X)\} \left\{ 1 - \frac{T}{\pi^*(X)} \right\} \\
+ \{\bar{m}_{WL}(X) - m^{1*}(X)\} \left\{ \frac{T}{\pi^*(X)} - \frac{T}{\pi_\text{CAL}^1(X)} \right\},
\]

denoted as \( \delta_1 + \delta_3 \), where \( \delta_3 \) is as before. By direct calculation, we have

\[
E \left\{ \tilde{E}^2(\delta_1) \right\} (X_{i=1,...,n}) = \frac{1}{n} \tilde{E} \left\{ \frac{1 - \pi^*(X)}{\pi^*(X)} \{\bar{m}_{WL}(X) - m^{1*}(X)\}^2 \right\} \leq e^{-B_0 \frac{b^2_1}{n}}.
\]

By Markov inequality, \( \tilde{E}(\delta_1) = e^{-B_0 \frac{b^2_1}{n} n^{-1/2} O_p(1)} \). Then (S39) follows.

### III.8 Proof of Theorem 6

We use the decomposition (S22) and handle \( \delta_1 \), \( \delta_2 \), and \( \delta_3 \) separately. The term \( \tilde{E}(\delta_2) \) can be bounded by (S24)–(S26) as in the proof of Theorem 3. By the mean value equation (S9) and the Cauchy–Schwartz inequality, \( \tilde{E}(\delta_3) \) can be bounded as

\[
\left| \tilde{E}(\delta_3) \right| \leq e^{C_0 \|\hat{\alpha}_{WL}^1 - \hat{\alpha}_{\text{CAL}}^1\|_1} \tilde{E}^{1/2} \left[ T w(X; \hat{\gamma}_{CAL}^1)\{\hat{h}_{WL}^1(X) - \hat{h}_{\text{CAL}}^1(X)\}^2 \right] \\
\times \tilde{E}^{1/2} \left[ T w(X; \hat{\gamma}_{CAL}^1)\{\hat{m}_{\text{CAL}}^1(X) - \bar{m}_{\text{WL}}^1(X)\}^2 \right].
\]

(S40)

Similarly as in Lemma 10 but arguing in the reverse direction by Assumptions 1(i) and 3(iii), we find

\[
\tilde{E} \left[ T w(X; \hat{\gamma}_{CAL}^1)\{\hat{m}_{\text{CAL}}^1(X) - \bar{m}_{\text{WL}}^1(X)\}^2 \right] \leq e^{C_4 \|\hat{\alpha}_{WL}^1 - \hat{\alpha}_{CAL}^1\|_1} \\
\times \tilde{E} \left[ T w(X; \hat{\gamma}_{CAL}^1)\psi_2\{\hat{h}_{WL}^1(X)\}\{\hat{m}_{\text{CAL}}^1(X) - \bar{m}_{\text{WL}}^1(X)\}\{\hat{h}_{\text{CAL}}^1(X) - \hat{h}_{WL}^1(X)\} \right] \\
\leq e^{C_4 \|\hat{\alpha}_{WL}^1 - \hat{\alpha}_{CAL}^1\|_1} C_1 D_{\text{WL}}^\frac{1}{2}(\hat{m}_{\text{CAL}}^1(X); \bar{m}_{\text{WL}}^1; \hat{\gamma}_{CAL}^1),
\]

(S41)

where the second inequality follows from Assumption 3(i). In the following, we derive two different bounds on \( \tilde{E}(\delta_1) \), leading to (56) and (57) respectively.

First, suppose that condition (35) holds. Consider the following decomposition

\[
\tilde{E}(\delta_1) = \tilde{E} \left[ \psi_2\{\hat{h}_{WL}(X)\}\{\hat{h}_{WL}(X) - \hat{h}_{\text{CAL}}^1(X)\} \left\{ 1 - \frac{T}{\pi_\text{CAL}^1(X)} \right\} \right] \\
+ \tilde{E} \left[ \tilde{\psi}_2(X)\{\hat{h}_{WL}(X) - \hat{h}_{\text{CAL}}^1(X)\} \left\{ 1 - \frac{T}{\pi_\text{CAL}^1(X)} \right\} \right],
\]

(S42)

denoted as \( \Delta_{11} + \Delta_{12} \), where

\[
\tilde{\psi}_2(X) = \int_0^1 \left( \psi_2(\hat{h}_{WL}(X) + u\{\hat{h}_{WL}(X) - \hat{h}_{\text{CAL}}^1(X)\}) - \psi_2(\hat{h}_{\text{CAL}}^1(X)) \right) du.
\]
Denote by $\Omega_6$ the event that
\[
\sup_{j=0,1,\ldots,p} \left| (\bar{E} - E) \left[ \psi_2(\bar{h}_WL(X))f_j(X) \left( 1 - \frac{T}{\pi_{CAL}(X)} \right) \right] \right| \leq 2C_1\lambda_0.
\]

Then $P(\Omega_6) \geq 1 - 2\epsilon$ similarly as in Lemma 1(i). In the event $\Omega_6$, we have
\[
|\Delta_{11}| \leq \| \hat{\alpha}_{rwl}^1 - \bar{\alpha}_{wl}^1 \|_1 \sup_{j=0,1,\ldots,p} \left| \bar{E} \left[ \psi_2(\bar{h}_WL(X))f_j(X) \left( 1 - \frac{T}{\pi_{CAL}(X)} \right) \right] \right|
\leq \| \hat{\alpha}_{rwl}^1 - \bar{\alpha}_{wl}^1 \|_1 (\Lambda_1 + 2C_1\lambda_0).
\]  \hspace{1cm} (S43)

To bound $\Delta_{12}$, we have by Assumption 3(iii),
\[
|\bar{\psi}_2(X)| \leq \psi_2(\bar{h}_WL(X)) \left( e^{C_3|\bar{h}_WL(X) - \bar{h}_WL(X)|} - 1 \right)
\leq \psi_2(\bar{h}_WL(X)) C_3|\bar{h}_WL(X) - \bar{h}_WL(X)| e^{C_3|\bar{h}_WL(X) - \bar{h}_WL(X)|},
\]  \hspace{1cm} (S44)

where the second inequality follows because $(e^c - 1)/c = \int_0^c e^{ac} \, du \leq e^c$ for $c \geq 0$. As a result, we find from Assumptions 1(i) and 3(i),
\[
|\Delta_{12}| \leq (1 + e^{-B_0})C_1C_3|\hat{\alpha}_{rwl}^1 - \bar{\alpha}_{wl}^1| \sup_{j=0,1,\ldots,p} \left| \bar{E} \left[ (\bar{h}_WL(X) - \bar{h}_WL(X))^2 \right] \right|.
\]  \hspace{1cm} (S45)

By condition (35), $\bar{E}[(\bar{h}_WL(X) - \bar{h}_WL(X))^2]$ can be bounded as (S33) in the event $\Omega_1 \cap \Omega_5$. Then (57) follows by collecting inequalities (S24)–(S26) and (S40)–(S45) and applying (55), (S7), and (S10) in the event $\Omega_0 \cap \Omega_1 \cap \Omega_2 \cap \Omega_3 \cap \Omega_4$.

Now suppose that (35) may not hold. Denote $h(X;\alpha^1) = \alpha^{12}f(X)$. Then $\bar{E}(\delta_1)$ can be decomposed as
\[
\bar{E}(\delta_1) = (\bar{E} - E) \left[ \psi(\bar{h}_WL(X)) - \psi(\bar{h}_WL(X)) \right] \left( 1 - \frac{T}{\pi_{CAL}(X)} \right)
+ E \left[ \psi(\bar{h}_WL(X)) - \psi(\bar{h}_WL(X)) \right] \left( 1 - \frac{T}{\pi_{CAL}(X)} \right),
\]
denoted as $\Delta_{13} + \Delta_{14}$. In the event $\Omega_0 \cap \Omega_1 \cap \Omega_2 \cap \Omega_3 \cap \Omega_4$, we have $\| \hat{\alpha}_{rwl}^1 - \bar{\alpha}_{wl}^1 \|_1 \leq \eta_{11}$ from (55) and hence by the mean value theorem,
\[
|\Delta_{14}| \leq \eta_{11} \sup_{j=0,1,\ldots,p} \left| E \left[ \psi_2(h(X;\alpha^1))f_j(X) \left( 1 - \frac{T}{\pi_{CAL}(X)} \right) \right] \right| \leq \eta_{11}\lambda_0(\eta_{11}),
\]  \hspace{1cm} (S46)

where $\bar{\alpha}^1$ lies between $\hat{\alpha}_{rwl}^1$ and $\bar{\alpha}_{wl}^1$. Moreover, in the event $(\Omega_0 \cap \Omega_1 \cap \Omega_2 \cap \Omega_3 \cap \Omega_4) \cap \Omega_7$, applying Lemma 13 below yields
\[
|\Delta_{13}| \leq 2C_1(1 + C_3e^{C_4\eta_{11}})\eta_{11}\lambda_0.
\]  \hspace{1cm} (S47)

Then (57) follows by combining (S46)–(S47) and other aforementioned inequalities.
Lemma 13. For \( r \geq 0 \), denote by \( \Omega_7 \) the event that

\[
\sup_{\|\alpha^1 - \hat{\alpha}_{WL}^1\|_1 \leq r} \left| (\hat{E} - E) \left( \psi\{h(X; \alpha^1)\} - \psi\{\hat{h}_{WL}^1(X)\} \right) \right| \leq \sqrt{8}C_1(1 + C_3e^{C_4r})r\lambda_0.
\]

Under Assumptions 1(i)–(ii), 3(i) and 3(iii), if \( \lambda_0 \geq \sqrt{2}(e^{-B_0} + 1)C_0\sqrt{\log\{(1 + p)/\epsilon\}/n} \), then \( P(\Omega_6) \geq 1 - 2\epsilon \).

**Proof.** Denote

\[
g(T, X; \alpha^1) = \left[ \psi\{h(X; \alpha^1)\} - \psi\{\hat{h}_{WL}^1(X)\} \right] \left( 1 - \frac{T}{\pi_{CAL}^1(X)} \right).
\]

For \( \|\alpha^1 - \hat{\alpha}_{WL}^1\|_1 \leq r \), similar manipulation as in (S42) and (S44) using Assumptions 1(i), 3(i) and 3(iii) yields

\[
\left| \psi\{h(X; \alpha^1)\} - \psi\{\hat{h}_{WL}^1(X)\} \right| \leq \psi_2\{\hat{h}_{WL}^1(X)\}|h(X; \alpha^1) - \hat{h}_{WL}^1(X)| + \psi_2\{\hat{h}_{WL}^1(X)\}C_3|h(X; \alpha^1) - \hat{h}_{WL}^1(X)|e^{C_3|h(X; \alpha^1) - \hat{h}_{WL}^1(X)|} \leq C_1(1 + C_3e^{C_4r})|h(X; \alpha^1) - \hat{h}_{WL}^1(X)|,
\]

that is, \( \psi() \) satisfies a Lipschitz condition. By the symmetrization and contraction theorems (e.g., Buhlmann & van de Geer 2011, Theorems 14.3 and 14.4), we have

\[
E \left[ \sup_{\|\alpha^1 - \hat{\alpha}_{WL}^1\|_1 \leq r} \left| (\hat{E} - E)\{g(T, X; \alpha^1)\} \right| \right] \leq 2E \sup_{\|\alpha^1 - \hat{\alpha}_{WL}^1\|_1 \leq r} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i g(T_i, X_i; \alpha^1) \right|
\]

\[
\leq 2C_1(1 + C_3e^{C_4r}) \times E \sup_{\|\alpha^1 - \hat{\alpha}_{WL}^1\|_1 \leq r} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i \{h(X_i; \alpha^1) - \hat{h}_{WL}^1(X_i)\} \right| \left( 1 - \frac{T_i}{\pi_{CAL}^1(X_i)} \right)
\]

\[
\leq 2C_1(1 + C_3e^{C_4r})r \times E \sup_{j=0,1,\ldots,p} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i f_j(X_i) \right| \left( 1 - \frac{T_i}{\pi_{CAL}^1(X_i)} \right),
\]

where \( (\sigma_1, \ldots, \sigma_n) \) are independent Rademacher variables with \( P(\sigma_i = 1) = P(\sigma_i = -1) = 1/2 \) for each \( i \). By Hoeffding’s moment inequality (Buhlmann & van de Geer 2011, Lemma 14.14), we find from the preceding inequality

\[
E \left[ \sup_{\|\alpha^1 - \hat{\alpha}_{WL}^1\|_1 \leq r} \left| (\hat{E} - E)\{g(T, X; \alpha^1)\} \right| \right] \leq 2C_1(1 + C_3e^{C_4r})r \times C_0(e^{B_0} + 1)\sqrt{\frac{2\log(2 + 2p)}{n}},
\]

by Assumption 1(i)–(ii). For \( \|\alpha^1 - \hat{\alpha}_{WL}^1\|_1 \leq r \), inequality (S48) also shows that \( |g(T_i, X_i; \alpha^1)| \leq C_1(1 + C_3e^{C_4r})C_0(e^{B_0} + 1)r \). By Massart’s inequality (Buhlmann & van de Geer 2001, Theo-
rem 14.2), we have with probability at least $1 - 2\epsilon$,

$$\sup_{\|\alpha^{1} - \bar{\alpha}_{WL}^{1}\| \leq r} \left| \tilde{E} - E \{ g(T, X; \alpha^{1}) \} \right|$$

$$\leq C_{0}(e^{B_{0}} + 1)C_{1}(1 + C_{3}e^{C_{4}r})r \left\{ 2\sqrt{\frac{2\log(2 + 2p)}{n}} + \sqrt{\frac{8\log\{1/(2\epsilon)\}}{n}} \right\}$$

$$\leq C_{0}(e^{B_{0}} + 1)C_{1}(1 + C_{3}e^{C_{4}r})r \sqrt{\frac{16 \log\{(1 + p)/\epsilon\}}{n}},$$

where the second inequality uses $\sqrt{a} + \sqrt{b} \leq \sqrt{2(a + b)}$. \(\square\)

### III.9 Proof of Theorem 7

The proof is similar to that of Theorem 4. First, (S29) for $\tilde{E}(\delta^{2}_{2})$ remains valid. Second, combining (S30) and (S41) yields

$$\tilde{E}(\delta^{2}_{3}) \leq e^{-B_{0}} \left( 1 + e^{\|\tilde{\alpha}_{rwl}^{1} - \tilde{\alpha}_{wl}^{1}\|_{1}C_{0}^{2}} \right)^{2} C_{4}^{2}\|\tilde{\alpha}_{rwl}^{1} - \tilde{\alpha}_{wl}^{1}\|_{1} C_{1}D_{wl}^{\dagger}(\tilde{h}_{rwl}^{1}, \tilde{h}_{wl}^{1}; \tilde{\gamma}_{CAL}^{1})].$$

Third, similarly as in (S32) and (S41), we have

$$\tilde{E}(\delta^{1}_{2}) = \tilde{E} \left( \{ \hat{m}_{rwl}^{1}(X) - \bar{m}_{wl}^{1}(X) \}^{2} \left\{ 1 - \frac{T}{\bar{\pi}_{CAL}(X)} \right\} \right) \leq (1 + e^{-B_{0}})^{2} e^{2C_{4}\|\tilde{\alpha}_{rwl}^{1} - \tilde{\alpha}_{wl}^{1}\|_{1}C_{1}^{2}} \tilde{E} \left( \{ \hat{h}_{rwl}^{1}(X) - \tilde{h}_{wl}^{1}(X) \}^{2} \right)$$

$$\leq (1 + e^{-B_{0}})^{2} e^{2C_{4}\|\tilde{\alpha}_{rwl}^{1} - \tilde{\alpha}_{wl}^{1}\|_{1}C_{1}^{2}} \tilde{E} \left( \{ \hat{h}_{rwl}^{1}(X) - \tilde{h}_{wl}^{1}(X) \}^{2} \right)$$

Inequality (58) follows by collecting the aforementioned inequalities and applying (55), (56), (S10), and (S13) in the event $\Omega_{0} \cap \Omega_{1} \cap \Omega_{2} \cap \Omega_{3} \cap \Omega_{4}$. If condition (35) holds, then in the event $\Omega_{1} \cap \Omega_{5}$, combining (S33) and (S19) and using $(1 - e^{-c})/c \geq e^{-c}$ for $c \geq 0$ yields

$$\tilde{E} \left( \{ \hat{h}_{rwl}^{1}(X) - \tilde{h}_{wl}^{1}(X) \}^{2} \right) \leq e^{B_{0}} \lambda_{0}\|\tilde{\alpha}_{rwl}^{1} - \tilde{\alpha}_{wl}^{1}\|^{2}$$

$$+ \tau_{0}^{-1} \left\{ \lambda_{0}\|\tilde{\alpha}_{rwl}^{1} - \tilde{\alpha}_{wl}^{1}\|^{2} + e^{C_{4}\|\tilde{\alpha}_{rwl}^{1} - \tilde{\alpha}_{wl}^{1}\|_{1}C_{1}^{2}} D_{wl}^{\dagger}(\hat{h}_{rwl}^{1}, \tilde{h}_{wl}^{1}; \tilde{\gamma}_{CAL}^{1}) \right\}.$$ 

Inequality (59) follows by combining (S49)–(S50) and other aforementioned inequalities except that inequality (56) is replaced by (57).
III.10 Technical tools

For completeness, we state the following concentration inequalities, which can be obtained from Buhlmann & van de Geer (2011), Lemmas 14.11, 14.16, and 14.9.

**Lemma 14.** Let \((Y_1, \ldots, Y_n)\) be independent variables such that \(E(Y_i) = 0\) for \(i = 1, \ldots, n\) and \(\max_{i=1,\ldots,n} |Y_i| \leq c_0\) for some constant \(c_0\). Then for any \(t > 0\),

\[
P\left(\left| \frac{1}{n} \sum_{i=1}^{n} Y_i \right| > t \right) \leq 2 \exp \left( -\frac{nt^2}{2c_0^2} \right).
\]

**Lemma 15.** Let \((Y_1, \ldots, Y_n)\) be independent variables such that \(E(Y_i) = 0\) for \(i = 1, \ldots, n\) and \((Y_1, \ldots, Y_n)\) are uniformly sub-gaussian: \(\max_{i=1,\ldots,n} c_1^2 E\{\exp(Y_i^2/c_1^2)\} - 1 \leq c_2^2\) for some constants \((c_1, c_2)\). Then for any \(t > 0\),

\[
P\left(\left| \frac{1}{n} \sum_{i=1}^{n} Y_i \right| > t \right) \leq 2 \exp \left\{ -\frac{nt^2}{8(c_1^2 + c_2^2)} \right\}.
\]

**Lemma 16.** Let \((Y_1, \ldots, Y_n)\) be independent variables such that \(E(Y_i) = 0\) for \(i = 1, \ldots, n\) and

\[
\frac{1}{n} \sum_{i=1}^{n} E(|Y_i|^k) \leq \frac{k!}{2} c_3^{k-2} c_4^2, \quad k = 2, 3, \ldots,
\]

for some constants \((c_3, c_4)\). Then for any \(t > 0\),

\[
P\left(\left| \frac{1}{n} \sum_{i=1}^{n} Y_i \right| > c_3 t + c_4 \sqrt{2t} \right) \leq 2 \exp(-nt).
\]

The following results about sub-gaussian variables can be deduced from Buhlmann & van de Geer (2011), Lemmas 14.3 and 14.5.

**Lemma 17.** Suppose that \(Y\) is sub-gaussian: \(c_1^2 E\{\exp(X^2/c_1^2)\} - 1 \leq c_2^2\) for some constants \((c_1, c_2)\). Then

\[
E(|Y|^k) \leq \Gamma \left( \frac{k}{2} + 1 \right) (c_1^2 + c_2^2) c_1^{k-2}, \quad k = 2, 3, \ldots,
\]

**Lemma 18.** Suppose that \(X\) is bounded: \(|X| \leq c_0\) for a constant \(c_0\), and \(Y\) is sub-gaussian: \(c_1^2 E\{\exp(X^2/c_1^2)\} - 1 \leq c_2^2\) for some constants \((c_1, c_2)\). Then \(Z = XY^2\) satisfies

\[
E\left\{ |Z - E(Z)|^k \right\} \leq \frac{k!}{2} c_3^{k-2} c_4^2, \quad k = 2, 3, \ldots,
\]

for \(c_3 = 2c_0c_1^2\) and \(c_4 = 2c_0c_1c_2\).