Supplementary Material for
Machine Learning for Variance Reduction in Online Experiments

Yongyi Guo
Department of Operations Research and Financial Engineering
Princeton University
Princeton, NJ 08544
yongyig@princeton.edu

Dominic Coey
Facebook
1 Hacker Way, Menlo Park, CA 94025
coeysfb.com

Mikael Konutgan
Facebook
1 Hacker Way, Menlo Park, CA 94025
kmikael@fb.com

Wenting Li
Facebook
1 Hacker Way, Menlo Park, CA 94025
wentingli@fb.com

Chris Schoener
Facebook
1 Hacker Way, Menlo Park, CA 94025
chrissc@fb.com

Matt Goldman
Facebook
1 Hacker Way, Menlo Park, CA 94025
mattgoldman@fb.com

In this supplementary material, we provide the proof of all theoretical results stated in the paper.

1 Proof of Proposition 1

For any (deterministic) $g \in \mathcal{G}$, we have

$$ P[Z(g)Z(g)^\top] = M_1(g) \otimes M_2, $$

where $\otimes$ denotes the Kronecker product,

$$ M_1(g) = \begin{pmatrix} 1 & Eg(X) \\ Eg(X) & Eg(X)^2 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1 & p \\ p & p \end{pmatrix}. $$

Therefore, any eigenvalue of $P[Z(g)Z(g)^\top]$ is the product of one eigenvalue of $M_1(g)$ and one eigenvalue of $M_2$. It’s easy to verify from Assumption 1 that all eigenvalues of $M_1(g)$ and $M_2$ are nonnegative and bounded. Thus, we only need to show $\inf_{g \in \mathcal{G}} \lambda_{\min}(M_1(g)) > 0, \lambda_{\min}(M_2) > 0$.

Through some calculations, one can find out that

$$ \lambda_{\min}(M_1(g)) = \frac{1}{2} \left\{ (Eg(X)^2 + 1) - \sqrt{(Eg(X)^2 + 1)^2 - 4Var(g(X))} \right\} $$

$$ = \frac{2Var(g(X))}{(Eg(X)^2 + 1) + \sqrt{(Eg(X)^2 + 1)^2 - 4Var(g(X))}} \geq \frac{Var(g(X))}{Eg(X)^2 + 1}. $$

which leads to
\[
\inf_{g \in \mathcal{G}} \lambda_{min}(M_1(g)) \geq \inf_{g \in \mathcal{G}} \text{Var}(g(X)) \sup_{g \in \mathcal{G}} E_g(X)^2 + 1 > 0.
\]

On the other hand, \(\lambda_{min}(M_2) > 0\) can be deduced from \(p \in (0, 1)\). By combining the above two inequalities, we conclude the proof.

## 2 Proof of Proposition 2

For compactness we may write the random variables \(Z(\hat{g}_k)\) as \(\hat{Z}_k\) and \(Z(g_0)\) as \(Z\). Similarly for any observation \(i\) we write \(Z_i(\hat{g}_k)\) as \(\hat{Z}_{k,i}\) and \(Z_i(g_0)\) as \(Z_i\). We are only interested in convergence in probability, so we can assume that the inverse matrices in the definition of \(\hat{\beta}(\{\hat{g}_k\}_{k=1}^K)\) and \(\hat{\beta}(g_0)\) exist, as this happens with probability approaching 1 according to Lemma 2. We have
\[
\hat{\beta}(\{\hat{g}_k\}_{k=1}^K) - \beta(\{g_k\}_{k=1}^K) = A + B, \quad \text{where}
\]
\[
A = \left[ \frac{1}{N} \sum_k \sum_{i \in I_k} \hat{Z}_{k,i} \hat{Z}_{k,i}^\top \right]_F^{-1} - \left[ \frac{1}{K} \sum_k P[\hat{Z}_k \hat{Z}_k^\top] \right]_F^{-1} \cdot \left[ \frac{1}{N} \sum_k \sum_{i \in I_k} \hat{Z}_{k,i} Y_i \right].
\]
and
\[
B = \left[ \frac{1}{K} \sum_k P[\hat{Z}_k \hat{Z}_k^\top] \right]_F^{-1} \left[ \frac{1}{N} \sum_k \sum_{i \in I_k} [\hat{Z}_{k,i} Y_i - P[\hat{Z}_k Y]] \right].
\]
Similarly, \(\hat{\beta}(g_0) - \beta(g_0) = C + D, \quad \text{where}
\]
\[
C = \left[ \frac{1}{N} \sum_i Z_i Z_i^\top \right]_{F_1}^{-1} - \left[ P[Z Z^\top] \right]_{F_1}^{-1} \cdot \left[ \frac{1}{N} \sum_i Z_i Y_i \right]
\]
and
\[
D = \left[ P[Z Z^\top] \right]_{G_1}^{-1} \left[ \frac{1}{N} \sum_i [Z_i Y_i - P[Z Y]] \right].
\]

We can write
\[
\hat{\beta}(\{\hat{g}_k\}_{k=1}^K) - \beta(\{g_k\}_{k=1}^K) - [\hat{\beta}(g_0) - \beta(g_0)] = A - C + B - D. \quad \text{We show that}
\]
\[
\sqrt{N} ||A - C|| \to_p 0 \quad \text{and} \quad \sqrt{N} ||B - D|| \to_p 0.
\]

From the definitions of \(F_0\) and \(F_1\) above, we have
\[
A - C = [F_0 - F_1] \left[ \frac{1}{N} \sum_k \sum_{i \in I_k} \hat{Z}_{k,i} Y_i \right] + F_1 \left[ \frac{1}{N} \sum_k \sum_{i \in I_k} (\hat{Z}_{k,i} - Z_i) Y_i \right]. \quad \text{If}
\]
1. \(\sqrt{N} [F_0 - F_1] = o_p(1)\)
2. \(\frac{1}{N} \sum_k \sum_{i \in I_k} \hat{Z}_{k,i} Y_i = O_p(1)\)
3. \(\sqrt{N} F_1 = O_p(1)\)
4. \(\frac{1}{N} \sum_k \sum_{i \in I_k} (\hat{Z}_{k,i} - Z_i) Y_i = O_p(1)\)

then \(\sqrt{N} ||A - C|| = o_p(1)\) as desired. Similarly we write \(B - D\) as
\[
B - D = \left[ \left[ \frac{1}{K} \sum_k P[\hat{Z}_k \hat{Z}_k^\top] \right]_F^{-1} - \left[ P[Z Z^\top] \right]_{F_1}^{-1} \right] G_0 + \left[ P[Z Z^\top] \right]_{G_1}^{-1} [G_0 - G_1]. \quad \text{If}
\]
5. \(\left[ \left[ \frac{1}{K} \sum_k P[\hat{Z}_k \hat{Z}_k^\top] \right]_F^{-1} - \left[ P[Z Z^\top] \right]_{F_1}^{-1} \right] = o_p(1)\)
1 as well as the following fact: For each \( k \),
\[
\begin{align*}
\| \hat{Z}_k \|_2^2 & = 1 + \frac{1}{n} \sum_{i \in I_k} \hat{Z}_{k,i}^T \hat{Z}_{k,i} + \| \hat{Z}_k \|_2, \\
\| \hat{Z}_k \|_2 & = 1 + \frac{1}{n} \sum_{i \in I_k} \hat{Z}_{k,i}^T \hat{Z}_{k,i} + \| \hat{Z}_k \|_2.
\end{align*}
\]

Then
\[
\begin{align*}
\| \hat{Z}_k \|_2 & = 1 + \frac{1}{n} \sum_{i \in I_k} \hat{Z}_{k,i}^T \hat{Z}_{k,i} + \| \hat{Z}_k \|_2.
\end{align*}
\]

Step 1. We apply Lemma 3 by letting \( M_{1n} = \frac{1}{n} \sum_{k} \sum_{i \in I_k} \hat{Z}_{k,i} \hat{Z}_{k,i}^T, B_n = M_{2n} = \frac{1}{n} \sum_{k} \sum_{i \in I_k} \hat{Z}_{k,i} \hat{Z}_{k,i}^T, M_{3n} = \frac{1}{n} \sum_{k} \sum_{i \in I_k} \hat{Z}_{k,i} \hat{Z}_{k,i}^T. \)

Consequently, Step 1 amounts to verifying the conditions of Lemma 3. In fact, these conditions are guaranteed by Lemma 1 as well as the following fact: For each \( k = 1, \ldots, K \),
\[
\left\| \frac{1}{\sqrt{n}} \sum_{i \in I_k} \hat{Z}_{k,i} - P [ \hat{Z}_k \hat{Z}_k^T ] - Z_i Z_i^T + P [ Z Z^T ] \right\| \rightarrow_p 0.
\]

We now prove (1). Define \( W_{k,i} = \hat{Z}_{k,i} \hat{Z}_{k,i}^T - P [ \hat{Z}_k \hat{Z}_k^T ] - Z_i Z_i^T + P [ Z Z^T ] \), and note that conditional on the data in \( I_k \), the function \( \hat{g}_k \) is non-random, and the \( W_{k,i} \) are mean zero matrices, uncorrelated across observations in \( I_k \). With slight abuse of notation, we use \( E \| \cdot \big| I_k \| \) to denote expectations conditional on the observations with indices belonging to the set \( I_k \). For any \( k = 1, 2, \ldots, K \),
\[
E \left[ \left\| \frac{1}{\sqrt{n}} \sum_{i \in I_k} W_{k,i} \right\|_F^2 \right] = \frac{1}{n} E \left[ \text{tr} \left( \sum_{i,j \in I_k} W_{k,i} W_{k,j}^T \big| I_k \right) \right] \\
= \frac{1}{n} E \left[ \text{tr} \left( \sum_{i \in I_k} W_{k,i} W_{k,i}^T \big| I_k \right) \right] \\
\leq \frac{1}{n} E \left[ \left\| \sum_{i \in I_k} (\hat{Z}_{k,i} \hat{Z}_{k,i}^T - Z_i Z_i^T) \right\|_F^2 \big| I_k \right] \\
= P \left[ \left\| \hat{Z}_k \hat{Z}_k^T - P[Z Z^T] \right\|_F^2 \big| I_k \right].
\]

If the RHS of (5) is \( o_p(1) \), we can use Lemma 6.1 of [11] to conclude that \( \frac{1}{\sqrt{n}} \sum_{i \in I_k} W_{k,i} \) is \( o_p(1) \) as required. Some calculations give
\[
\left\| \hat{Z}_k \hat{Z}_k^T - P[Z Z^T] \right\|_F^2 \leq 12 \left[ (\hat{g}_k(X) - g_0(X))^2 + (\hat{g}_k(X))^2 + g_0(X)^2 \right].
\]

Then \( P \left[ (\hat{g}_k - g_0)^2 \right] \leq \sqrt{P[\hat{g}_k - g_0]^4} \rightarrow_p 0. \) Also
\[
P \left[ (\hat{g}_k - g_0)^2 \right] = \sqrt{P[\hat{g}_k - g_0]^4} \bigg/ \sqrt{P[\hat{g}_k + g_0]^4} \\
\leq \sup_{g \in \mathcal{G}} P[g^4] \bigg/ \sqrt{P[\hat{g}_k + g_0]^4} \\
\rightarrow_p 0,
\]
where the second-to-last line follows because \( \hat{g}_k + g_0 \in \mathcal{G} \) as \( \mathcal{G} \) is a vector space. We conclude from (6) that the RHS of (5) is \( o_p(1) \).

Step 2. By the Cauchy-Schwarz inequality,
\[
\left\| \frac{1}{N} \sum_{k} \sum_{i \in I_k} Z_i (\tilde{g}_k) Y_i \right\| \leq \sqrt{\frac{1}{N} \sum_{k} \sum_{i \in I_k} Z_i (\tilde{g}_k) Y_i^2} \sqrt{\frac{1}{N} \sum_{k} \sum_{i \in I_k} Y_i^2}.
\]

As \( E[Y^2] < \infty \), the second term on the RHS is \( O_p(1) \) by Markov’s inequality. Also for \( i \in I_k \),
\[
E \left[ |Z_i (\tilde{g}_k)|^2 \right] = E[1 + Y_i^2 + \tilde{g}_k(X_i)^2 + Y_i \tilde{g}_k(X_i)] \leq \sup_{g \in \mathcal{G}} E[2(1 + g(X_i)^2)] < \infty,
\]
and by Markov’s inequality the first term on the RHS is also \( O_p(1) \).
Step 3. By the central limit theorem, $\sqrt{N} \left( \sum_i \frac{Z_i Z_i^T}{N} - P[ZZ^T] \right)$ is asymptotically normal. By the delta method and invertibility of $P[ZZ^T]$, $\sqrt{N} \left( \sum_i \frac{Z_i Z_i^T}{N} - P[ZZ^T]^{-1} \right)$ is also, and hence its norm is $O_p(1)$.

Step 4. We show that for any $k$, $\frac{1}{n} \sum_{i \in I_k} (\hat{g}_k(X_i) - g_0(X_i)) Y_i = o_p(1)$, from which the result follows. By Cauchy-Schwarz,

$$\frac{1}{n} \sum_{i \in I_k} (\hat{g}_k(X_i) - g_0(X_i)) Y_i \leq \sqrt{\frac{1}{n} \sum_{i \in I_k} (\hat{g}_k(X_i) - g_0(X_i))^2} \sqrt{\frac{1}{n} \sum_{i \in I_k} Y_i^2}.$$  

As $Y$ has finite second moment by assumption, it remains to show the first term on the RHS is $o_p(1)$. We have

$$\frac{1}{n} \sum_{i \in I_k} (\hat{g}_k(X_i) - g_0(X_i))^2 = \frac{1}{n} \sum_{i \in I_k} [(\hat{g}_k(X_i) - g_0(X_i))^2 - P[(\hat{g}_k - g_0)^2]] + P[(\hat{g}_k - g_0)^2].$$

(12)

From Lemma 6.1 in [1], the first term on the RHS in (12) is $o_p(1)$ and by the convergence assumption on $\hat{g}_k$, the second term is too.

Step 5. By the continuous mapping theorem it suffices to show that $\| \frac{1}{K} \sum_k [P[Z(\hat{g}_k)Z(\hat{g}_k)^T] - P[Z(g_0)Z(g_0)^T]) \| = o_p(1)$. From the argument in Step 1, both $P[|\hat{g}_k - g_0|]^{2}$ and $P[|\hat{g}_k^2 - g_0^2|]^{2}$ are $o_p(1)$ for all $k$, and hence $P[|\hat{g}_k - g_0|]$ and $P[|\hat{g}_k^2 - g_0^2|]$ are both $o_p(1)$ for all $k$. The other entries in the matrix are straightforwardly $o_p(1)$.

Step 6. This follows from Step 8 and the fact that by Chebyshev’s inequality, $\| \frac{1}{\sqrt{N}} \sum_i [Z_i Y_i - P(ZY)] \| = O_p(1)$.

Step 7. $P[ZZ^T]$ is invertible by assumption.

Step 8. The reasoning here is similar to Step 1. For any $k$ and $i \in I_k$, define $W_{k,i} = \hat{Z}_{k,i} Y_i - P[\hat{Z}_i Y_i] - Z_i Y_i + P[ZY]$, and note that conditional on the data in $I_k$, the $W_{k,i}$ are mean zero matrices, uncorrelated across observations in $I_k$. Then

$$E \left( \left\| \frac{1}{\sqrt{n}} \sum_{i \in I_k} W_{k,i} \right\|_{I_k}^2 \right) \leq \frac{1}{n} E \left( \sum_{i \in I_k} \left\| \hat{Z}_{k,i} Y_i - Z_i Y_i \right\|_{I_k}^2 \right) = P \left( \| \hat{Z}_i Y - Z_i Y \|^2 \right).$$

Because $P[|\hat{g}_k(X) - g_0(X)|^2 Y]^2 \leq S \| \hat{g}_k - g_0 \|^2 \| Y \|^2 \rightarrow p 0$, the RHS of (2) is $o_p(1)$. We use Lemma 6.1 of [1] to conclude that $\left\| \frac{1}{\sqrt{n}} \sum_{i \in I_k} W_{k,i} \right\|$ is also $o_p(1)$, from which the result follows.

3 Proof of Theorem 1

We have

$$\hat{\alpha}_1(\{\hat{g}_k\}_{k=1}^K) - \hat{\alpha}_1(g_0) = \left[ \hat{\alpha}_1(\{\hat{g}_k\}_{k=1}^K) - \beta_1(\{\hat{g}_k\}_{k=1}^K) - \beta_3(\{\hat{g}_k\}_{k=1}^K) \right] \frac{1}{K} \sum_{k=1}^K \bar{P}\hat{g}_k$$

(13)

$$- \left[ \hat{\alpha}_1(g_0) - \beta_1(g_0) - \beta_3(g_0)Pg_0 \right]$$

(14)

$$= A + B,$$

(15)

where

$$A = \left[ \hat{\beta}_1(\{\hat{g}_k\}_{k=1}^K) - \beta_1(\{\hat{g}_k\}_{k=1}^K) \right] - \left[ \hat{\beta}_1(g_0) - \beta_1(g_0) \right].$$

(16)
We show this because

\[
B = \left[ \beta_3(\{\hat{g}_k\}_{k=1}^K) \frac{1}{N} \sum_i \hat{g}_{k(i)}(X_i) - \beta_3(\{\hat{g}_k\}_{k=1}^K) \frac{1}{K} \sum_{k=1}^K P\hat{g}_k \right]_C
- \left[ \beta_3(g_0) \frac{1}{N} \sum_i g_0(X_i) - \beta_3(g_0) Pg_0 \right]_D.
\]

(17)

Proposition 1 has established that \( A = o_p(1/\sqrt{N}) \). Moreover

\[
C = \left[ \beta_3(\{\hat{g}_k\}_{k=1}^K) - \beta_3(\{\hat{g}_k\}_{k=1}^K) \right] \frac{1}{N} \sum_i \hat{g}_{k(i)}(X_i) + \beta_3(\{\hat{g}_k\}_{k=1}^K) \left( \frac{1}{N} \sum_i [\hat{g}_{k(i)}(X_i) - P\hat{g}_k] \right)
\]

and

\[
D = \left[ \beta_3(g_0) - \beta_3(g_0) \right] \frac{1}{N} \sum_i g_0(X_i) + \left( \beta_3(g_0) \frac{1}{N} \sum_i [g_0(X_i) - Pg_0] \right).
\]

(19)

We show \( C_1 - D_1 \) and \( C_2 - D_2 \) are \( o_p(1/\sqrt{N}) \) to conclude. In fact

\[
C_1 - D_1 = \left[ \beta_3(\{\hat{g}_k\}_{k=1}^K) - \beta_3(\{\hat{g}_k\}_{k=1}^K) - \beta_3(g_0) + \beta_3(g_0) \right] \frac{1}{N} \sum_i \hat{g}_{k(i)}(X_i)
+ \left( \beta_3(g_0) - \beta_3(g_0) \right) \frac{1}{N} \sum_i [\hat{g}_{k(i)}(X_i) - g_0(X_i)] = o_p(1/\sqrt{N}).
\]

(20)

This is because

- \( \beta_3(\{\hat{g}_k\}_{k=1}^K) - \beta_3(\{\hat{g}_k\}_{k=1}^K) - \beta_3(g_0) + \beta_3(g_0) = o_p(1/\sqrt{N}) \) from Proposition 1;
- \( \frac{1}{N} \sum_i \hat{g}_{k(i)}(X_i) = \frac{1}{N} \sum_i g_0(X_i) + \frac{1}{N} \sum_i (\hat{g}_{k(i)}(X_i) - g_0(X_i)) = O_p(1) \) from the LLN and the same logic bounding (12) above;
- \( \beta_3(g_0) - \beta_3(g_0) = O_p(1/\sqrt{N}) \) from the CLT and the fact that \( P(Z(g_0)Z(g_0)') \) has all eigenvalues bounded away from 0;
- \( \frac{1}{N} \sum_i (\hat{g}_{k(i)}(X_i) - g_0(X_i)) = o_p(1) \) again from bounding argument applied to (12).

Similarly,

\[
C_2 - D_2 = \beta_3(\{\hat{g}_k\}_{k=1}^K) \left( \frac{1}{N} \sum_i [\hat{g}_{k(i)}(X_i) - P\hat{g}_{k(i)}] - [g_0(X_i) - Pg_0] \right)
+ \left( \beta_3(\{\hat{g}_k\}_{k=1}^K) - \beta_3(g_0) \right) \frac{1}{N} \sum_i [g_0(X_i) - Pg_0] = o_p(1/\sqrt{N}),
\]

(21)

which results from the following facts:

- \( \beta_3(\{\hat{g}_k\}_{k=1}^K) = \beta_3(g_0) + (\beta_3(\{\hat{g}_k\}_{k=1}^K) - \beta_3(g_0)) = O_p(1) \);
- \( \frac{1}{N} \sum_i [\hat{g}_{k(i)}(X_i) - P\hat{g}_{k(i)}] - [g_0(X_i) - Pg_0] = o_p(1/\sqrt{N}) \) from the same reasoning applied to bound (11);
- \( \beta_3(\{\hat{g}_k\}_{k=1}^K) - \beta_3(g_0) = o_p(1) \) due to convergence of \( \hat{g}_k \) to \( g_0 \), continuity of \( \beta_3(\cdot) \), and the continuous mapping theorem;
- \( \frac{1}{N} \sum_i [g_0(X_i) - Pg_0] = O_p(1/\sqrt{N}) \) from the CLT.

Combining the above arguments, we conclude that \( B = o_p(1/\sqrt{N}) \).
4 Proof of Proposition 4

We first show that \( \text{Var}(\hat{g}_{k(i)}(X_i)) \rightarrow_p \sigma_g^2 \). We have

\[
\text{Var}(\hat{g}_{k(i)}(X_i)) = \frac{1}{K} \sum_k \frac{1}{n} \sum_{i \in I_k} \hat{g}_k(X_i)^2 - \left[ \frac{1}{K} \sum_k \frac{1}{n} \sum_{i \in I_k} \hat{g}_k(X_i) \right]^2.
\]  

(22)

By the same logic as in Step 1 of the proof of Proposition 1, for each \( k = 1, 2, \ldots, K \),

\[
E \left[ \left\| \frac{1}{n} \sum_{i \in I_k} [\hat{g}_k(X_i)^2 - P\hat{g}_k^2] \right\|_F^2 \right] \rightarrow_p 0,
\]

and so \( \frac{1}{n} \sum_{i \in I_k} \hat{g}_k(X_i)^2 - P\hat{g}_k^2 \rightarrow_p 0 \). Since \( P\hat{g}_k^2 \rightarrow_p P\hat{g}_0^2 \), it follows that \( \frac{1}{n} \sum_{i \in I_k} \hat{g}_k(X_i)^2 \rightarrow_p P\hat{g}_0^2 \). Similarly \( \frac{1}{n} \sum_{i \in I_k} \hat{g}_k(X_i) \rightarrow_p P\hat{g}_0 \). Hence \( \text{Var}(\hat{g}_{k(i)}(X_i)) \rightarrow_p \sigma_g^2 \). Also, by Proposition 1,

\[
\left\| \hat{\beta}(\{\hat{g}_k\}_{k=1}^K) - \beta(\{g_k\}_{k=1}^K) \right\| \rightarrow_p 0
\]

(23)

and by continuity of \( \beta(\cdot) \) and the continuous mapping theorem,

\[
\left\| \beta(\{\hat{g}_k\}_{k=1}^K) - \beta(g_0) \right\| \rightarrow_p 0.
\]

(24)

Consequently \( \left\| \hat{\beta}(\{\hat{g}_k\}_{k=1}^K) - \beta(g_0) \right\| \rightarrow_p 0 \). By the continuous mapping theorem, we conclude that \( \sigma^2 \rightarrow_p \sigma^2 \).

5 Proof of auxiliary lemmas

Lemma 1. Given Assumption 1,

\[
\left\| \frac{1}{N} \sum_k \sum_{j \in I_k} \hat{Z}_{k,j} \hat{Z}_{k,j}^T - \frac{1}{K} \sum_k P(\hat{Z}_k \hat{Z}_k^T) \right\| = O_p(1/\sqrt{n}).
\]

Proof. Since the number of splits \( K \) is bounded, we only need to verify for any \( k \in \{1, 2, \ldots, K\} \),

\[
\left\| \frac{1}{n} \sum_{j \in I_k} \hat{Z}_{k,j} \hat{Z}_{k,j}^T - P(\hat{Z}_k \hat{Z}_k^T) \right\| = O_p(1/\sqrt{n}).
\]

Below we’ll prove

\[
\frac{1}{n} \sum_{j \in I_k} T_j^2 \hat{g}_k^2(X_j) - E[T_j^2 \hat{g}_k^2(X_j)|I_k] = O_p(1/\sqrt{n}).
\]

(25)

The other terms can be derived in the similar manner.

First, since \( P(\hat{g}_k - g_0)^4 \rightarrow_p 0 \) as \( n \rightarrow \infty \), we know that for any subsequence \( \{n_l\} \) of \( \mathbb{N} \), it further has a subsequence \( \{n_l'\} \), such that \( P(\hat{g}_k - g_0)^4 \rightarrow 0 \) a.s. as \( l \rightarrow \infty \). Our next step is to prove

\[
\frac{1}{\sqrt{n_l'}} \sum_{j \in I_k} T_j^2 \hat{g}_k^2(X_j) - E[T_j^2 \hat{g}_k^2(X_j)|I_k] = O_p(1)
\]

(26)

as \( l \rightarrow \infty \).

For notational simplicity, define \( V_{k,j} = T_j^2 \hat{g}_k^2(X_j) - E[T_j^2 \hat{g}_k^2(X_j)|I_k] \). Since \( \{V_{k,j}\}_{j \in I_k} \) are independent conditioned on \( I_k \), for any \( t \in \mathbb{R} \) we have

\[
E \exp \left( it/\sqrt{n_l'} \sum_{j \in I_k} V_{k,j} \right) = E \left[ E \left[ \exp \left( it/\sqrt{n_l'} \sum_{j \in I_k} V_{k,j} \right) \right] | I_k \right] = E \left[ E \left[ \exp \left( it/\sqrt{n_l'} \sum_{j \in I_k} V_{k,j} \right) \right] | I_k \right]^{n_l'}.
\]
Furthermore,
\[
\lim_{t \to \infty} E \exp \left( \frac{it}{\sqrt{n_l^t}} \cdot \sum_{j \in I_k} V_{k,j} \right) = \lim_{t \to \infty} E \left\{ E \left[ \exp \left( \frac{it}{\sqrt{n_l^t}} \cdot V_{k,j} \right) \right] I_k^c \right\}^{n_l^t} = E \left( E \left[ \exp \left( \frac{it}{\sqrt{n_l^t}} \cdot V_{k,j} \right) \right] I_k^c \right)^{n_l^t}.
\]

(27)

Our goal is now to derive the limit in the last term so that we can infer the limiting distribution of \(1/\sqrt{n_l^t} \cdot \sum_{j \in I_k} V_{k,j}\).

First, we conduct the Taylor expansion
\[
\exp \left( \frac{it}{\sqrt{n_l^t}} \cdot V_{k,j} \right) = 1 + it/\sqrt{n_l^t} \cdot V_{k,j} - \frac{t^2}{2n_l^t} V_{k,j}^2 + R_{k,j}.
\]

Here
\[
R_{k,j} = \exp \left( \frac{it}{\sqrt{n_l^t}} \cdot V_{k,j} \right) - \left[ 1 + it/\sqrt{n_l^t} \cdot V_{k,j} - \frac{t^2}{2n_l^t} V_{k,j}^2 \right].
\]

Thus
\[
E \left[ \exp \left( \frac{it}{\sqrt{n_l^t}} \cdot V_{k,j} \right) \right| I_k^c = 1 + it/\sqrt{n_l^t} \cdot E[V_{k,j} | I_k^c] - \frac{t^2}{2n_l^t} E[V_{k,j}^2 | I_k^c] + E[R_{k,j} | I_k^c]
\]

(28)

First, with probability 1,
\[
\lim_{t \to \infty} E[V_{k,j}^2 | I_k^c] = \lim_{t \to \infty} \left\{ E[T_j^2 \hat{g}_k^2(X_j) | I_k^c] - E[T_j^2 \hat{g}_k^2(X_j) | I_k^c]^2 \right\}
\]

\[
= p \cdot P \left( \frac{X_j}{\hat{g}_k} \right)^2 - \left( P \frac{X_j}{\hat{g}_k} \right)^2.
\]

(29)

Next, we bound \( |E[R_{k,j} | I_k^c]|. \) In fact,
\[
R_{k,j} \leq \begin{cases} \frac{2t^3}{n_l^t} V_{k,j}^3 & \text{when } |V_{k,j}| \leq \frac{\sqrt{n_l^t}}{2t}, \\ 2 + \frac{t}{\sqrt{n_l^t}} |V_{k,j}| + \frac{t^2}{2n_l^t} |V_{k,j}|^2 & \text{otherwise.} \end{cases}
\]

This means
\[
|E[R_{k,j} | I_k^c]| \leq E[R_{k,j}^{(1)} | I_k^c] + E[R_{k,j}^{(2)} | I_k^c],
\]

where
\[
R_{k,j}^{(1)} = \frac{2t^3}{n_l^t} |V_{k,j}|^3 1(|V_{k,j}| \leq \frac{\sqrt{n_l^t}}{2t})
\]

\[
R_{k,j}^{(2)} = (2 + \frac{t}{\sqrt{n_l^t}} |V_{k,j}| + \frac{t^2}{2n_l^t} |V_{k,j}|^2) 1(|V_{k,j}| > \frac{\sqrt{n_l^t}}{2t}).
\]

On the one hand,
\[
E[R_{k,j}^{(1)} | I_k^c] \leq \frac{2t^3}{n_l^t} E \left[ |V_{k,j}|^{2+\delta/2} \left( \frac{\sqrt{n_l^t}}{2t} \right)^{1-\delta/2} | I_k^c \right]
\]

\[
= \frac{2^{\delta/2} t^{2+\delta/2}}{n_l^t \delta/4} E \left[ |T_j^2 \hat{g}_k^2(X_j) - ET_j^2 \hat{g}_k^2(X_j)|^{2+\delta/2} | I_k^c \right] \leq \frac{2^{2+\delta/2} t^{2+\delta/2}}{n_l^t \delta/4} | P \hat{g}_k |^{4+\delta}.
\]

On the other hand, by Markov’s inequality,
\[
E[R_{k,j}^{(2)} | I_k^c] \leq 2E \left[ \left( \frac{2t}{\sqrt{n_l^t}} \right)^{2+\delta/2} |V_{k,j}|^{2+\delta/2} | I_k^c \right] + t/\sqrt{n_l^t},
\]

\[
E \left[ |V_{k,j}| \left( \frac{2t}{\sqrt{n_l^t}} \right)^{1+\delta/2} | V_{k,j} |^{1+\delta/2} | I_k^c \right] + \frac{t^2}{2n_l^t},
\]

\[
E \left[ |V_{k,j}|^2 \left( \frac{2t}{\sqrt{n_l^t}} \right)^{\delta/2} | V_{k,j} |^{\delta/2} | I_k^c \right] \leq \frac{2^{\delta/2} t^{2+\delta/2}}{n_l^t \delta/4} | P \hat{g}_k |^{4+\delta}.
\]
Combining the above two bounds, we deduce that

$$|E[R_{k,j}|I_k]| \leq \frac{9^{7+\delta} \lambda^{2+\delta/2}}{n_l^{1+\delta/4}} P[\hat{g}_k]^{4+\delta}.$$ 

Thus with probability 1, $E[R_{k,j}|I_k] = o(1/n_l')$.

Combining the above bound, (28) and (29), we obtain that with probability 1,

$$\lim_{l \to \infty} n_l' \log E \left[ \exp \left( \frac{it}{\sqrt{n_l'} \cdot V_{k,j}} \right) \right] = \lim_{l \to \infty} n_l' \log \left( 1 - \frac{t^2}{2n_l'} E[V_{k,j}^2|I_k] + E[R_{k,j}|I_k] \right) = -\frac{t^2}{2n_l'} [p \cdot P g_0^4 - p^2 \cdot (P g_0^2)^2].$$

Finally we plug the above into (27) and conclude that

$$\lim_{l \to \infty} E \exp \left( \frac{it}{\sqrt{n_l'} \cdot \sum_{j \in I_k} V_{k,j}} \right) = \exp \left\{ -\frac{t^2}{2n_l'} [p \cdot P g_0^4 - p^2 \cdot (P g_0^2)^2] \right\}.$$ 

This implies that $\frac{1}{\sqrt{n_l'}} \sum_{j \in I_k} V_{k,j}$ converges in distribution to a centered normal random variable with variance $p \cdot P g_0^4 - p^2 \cdot (P g_0^2)^2$, and (26) follows.

Finally, since for any subsequence $\{n_l\}$ of $\mathbb{N}$, it further has a subsequence $\{n_l'\}$ such that (26) holds, it can only be the case that (25) is true.

\[\square\]

**Lemma 2.** The following hold with probability tending to 1:

$$\lambda_{\min} \left( \frac{1}{n} \sum_{i \in I_k} \hat{Z}_{k,i} \hat{Z}_{k,i}^T \right) \geq \frac{1}{2} \inf_{g \in \mathcal{G}} \lambda_{\min}(P[Z(g)Z(g)^T]) \quad \forall k \in \{1, 2, \ldots, K\};$$

$$\lambda_{\min} \left( \frac{1}{N} \sum_{i=1}^{N} \hat{Z}_{i} \hat{Z}_{i}^T \right) \geq \frac{1}{2} \inf_{g \in \mathcal{G}} \lambda_{\min}(P[Z(g)Z(g)^T]).$$

**Proof.** According to Weyl’s inequality,

$$\lambda_{\min} \left( \frac{1}{n} \sum_{i \in I_k} \hat{Z}_{k,i} \hat{Z}_{k,i}^T \right) \geq \lambda_{\min}(P(\hat{Z}_k \hat{Z}_k^T)) - \left\| \frac{1}{n} \sum_{j \in I_k} \hat{Z}_{k,j} \hat{Z}_{k,j}^T - P(\hat{Z}_k \hat{Z}_k^T) \right\| \geq \inf_{g \in \mathcal{G}} \lambda_{\min}(P[Z(g)Z(g)^T]) - \left\| \frac{1}{n} \sum_{j \in I_k} \hat{Z}_{k,j} \hat{Z}_{k,j}^T - P(\hat{Z}_k \hat{Z}_k^T) \right\|.$$ 

On the other hand, from the proof of Lemma 1 we know

$$\left\| \frac{1}{n} \sum_{j \in I_k} \hat{Z}_{k,j} \hat{Z}_{k,j}^T - P(\hat{Z}_k \hat{Z}_k^T) \right\| = O_p(1/\sqrt{n}).$$ 

This implies that

$$\lim_{n \to \infty} P \left( \left\| \frac{1}{n} \sum_{j \in I_k} \hat{Z}_{k,j} \hat{Z}_{k,j}^T - P(\hat{Z}_k \hat{Z}_k^T) \right\| \geq \frac{1}{2} \inf_{g \in \mathcal{G}} \lambda_{\min}(P[Z(g)Z(g)^T]) \right) = 0.$$ 

Combining the above, we obtain (30), (31) can be proved in a similar way. \[\square\]
Lemma 3. Let \{M_{1n}\}, \{M_{2n}\}, \{M_{3n}\}, \{M_{4n}\}, \{A_n\}, \{B_n\} be sequences of random real symmetric matrices of fixed dimension. Assume that with probability 1, \(\lambda_0 := \inf_n \lambda_{\min}(B_n) > 0\), and \|A_n - B_n\| = o_p(1). Moreover, assume that
\[
\|M_{1n} - A_n\| = O_p(1/\sqrt{n}), \|M_{3n} - A_n\| = O_p(1/\sqrt{n}),
\|M_{2n} - B_n\| = O_p(1/\sqrt{n}), \|M_{4n} - B_n\| = O_p(1/\sqrt{n}).
\]
If in addition,
\[
\sqrt{n}\|M_{1n} + M_{2n} - M_{3n} - M_{4n}\| \to_p 0,
\]
then
\[
\sqrt{n}\|M_{1n}^{-1} + M_{2n}^{-1} - M_{3n}^{-1} - M_{4n}^{-1}\| \to_p 0.
\]

Proof. Define the event
\[
E_n := \{\|A_n - B_n\| \geq \lambda_0/2\} \cup \{\max\{\|M_{1n} - A_n\|, \|M_{3n} - A_n\|\} \geq \lambda_0/2\}
\cup \{\max\{\|M_{2n} - B_n\|, \|M_{4n} - B_n\|\} \geq \lambda_0/2\}.
\]
Then \(\lim_{n \to \infty} P(E_n) = 0\). Now on \(E_n^c\), according to a Neumann series expansion,
\[
M_{1n}^{-1} = [A_n + (M_{1n} - A_n)]^{-1}
= A_n^{-1/2}[I - A_n^{-1/2}(M_{1n} - A_n)A_n^{-1/2} + D_{1n}]A_n^{-1/2}.
\]
Here \(D_{1n} = \sum_{j \geq 2}[-A_n^{-1/2}(M_{1n} - A_n)A_n^{-1/2}]^j\), and we have on \(E_n^c\)
\[
\|D_{1n}\| \leq \sum_{j \geq 2} \|A_n^{-1/2}(M_{1n} - A_n)A_n^{-1/2}\|^j
\leq \frac{\|A_n^{-1}\|^2\|M_{1n} - A_n\|^2}{1 - \|A_n^{-1}\|\|M_{1n} - A_n\|} \leq \frac{8}{\lambda_0^2} \|M_{1n} - A_n\|^2. \quad (32)
\]
Here we use the fact that on \(E_n^c\)
\[
\|A_n^{-1/2}(M_{1n} - A_n)A_n^{-1/2}\| \leq \|A_n^{-1/2}\|^2\|M_{1n} - A_n\|^2 < \frac{2}{\lambda_0} \cdot \frac{\lambda_0}{2} = 1.
\]

Similar expansions hold for \(M_{2n}, M_{3n}\) and \(M_{4n}\), and we define \(D_{2n}, D_{3n}\) and \(D_{4n}\) accordingly. Using some simple algebra, we deduce that on \(E_n^c\),
\[
M_{1n}^{-1} + M_{2n}^{-1} - M_{3n}^{-1} - M_{4n}^{-1} = J_{1n} + J_{2n} + J_{3n} + J_{4n},
\]
where
\[
J_{1n} = -A_n^{-1}[M_{1n} + M_{2n} - M_{3n} - M_{4n}]A_n^{-1},
J_{2n} = -A_n^{-1}(M_{4n} - M_{2n})A_n^{-1} + B_n^{-1}(M_{4n} - M_{2n})B_n^{-1},
J_{3n} = A_n^{-1/2}(D_{1n} - D_{3n})A_n^{-1/2},
J_{4n} = B_n^{-1/2}(D_{2n} - D_{4n})B_n^{-1/2}.
\]
For any \(\epsilon > 0\),
\[
P(\sqrt{n}\|M_{1n}^{-1} + M_{2n}^{-1} - M_{3n}^{-1} - M_{4n}^{-1}\| > \epsilon) < P(E_n) + \sum_{\ell = 1}^{4} P(E_n^c \cap \{\sqrt{n}\|J_{\ell n}\| > \epsilon/4\}). \quad (33)
\]
Combining the fact that \(\lim_{n \to \infty} P(E_n) = 0\), we only need to prove that each of the rest of the terms on the right-hand side of (33) has limit 0.

First, \(\lim_{n \to \infty} P(E_n^c \cap \{\sqrt{n}\|J_{1n}\| > \epsilon/4\}) = 0\) follows from our assumption. For \(J_{2n}\), observe that \(J_{2n} = J_{2n}^{(1)} + J_{2n}^{(2)}\), where
\[
J_{2n}^{(1)} = (B_n^{-1} - A_n^{-1})(M_{4n} - M_{2n})A_n^{-1}, J_{2n}^{(2)} = B_n^{-1}(M_{4n} - M_{2n})(B_n^{-1} - A_n^{-1}).
\]
We bound the limit of \(\|J_{2n}^{(1)}\|\) as follows: For any \(\delta > 0\), there exists \(M > 0\) such that \(\forall n, P(\sqrt{n}\|M_{4n} - M_{2n}\| > M) < \frac{\delta}{2}\). According to our assumption, there further exists \(N \in \mathbb{N}\) such that for all \(n > N\), \(P(\|A_n - B_n\| > \frac{\lambda_0^4}{32M}) < \frac{\delta}{2}\). Therefore for all \(n > N\),

\[
P(E_n^c \cap \{\sqrt{n}\|J_{2n}^{(1)}\| > \epsilon/8\}) \leq P(E_n^c \cap \{\sqrt{n}\|A_n^{-1}(A_n - B_n)B_n^{-1}(M_{4n} - M_{2n})A_n^{-1}\| > \epsilon/8\})
\]

\[
\leq P(E_n^c \cap \{\|A_n - B_n\|\sqrt{n}\|M_{4n} - M_{2n}\| > \lambda_0^3\epsilon/32\})
\]

\[
\leq P(\sqrt{n}\|M_{4n} - M_{2n}\| > M) + P(\|A_n - B_n\| > \lambda_0^3\epsilon/(32M)) < \delta.
\]

The above argument implies that \(\lim_{n \to +\infty} P(E_n^c \cap \{\sqrt{n}\|J_{2n}^{(1)}\| > \epsilon/8\}) = 0\). Similarly we have \(\lim_{n \to +\infty} P(E_n^c \cap \{\sqrt{n}\|J_{2n}^{(2)}\| > \epsilon/8\}) = 0\). Thus

\[
\lim_{n \to +\infty} P(E_n^c \cap \{\sqrt{n}\|J_{2n}\| > \epsilon/4\}) \leq \lim_{n \to +\infty} P(E_n^c \cap \{\sqrt{n}\|J_{2n}^{(1)}\| > \epsilon/8\}) + \lim_{n \to +\infty} P(E_n^c \cap \{\sqrt{n}\|J_{2n}^{(2)}\| > \epsilon/8\}) = 0.
\]

Now we proceed to bound the limit of \(\|J_{3n}\|\). In fact we have

\[
P(E_n^c \cap \{\sqrt{n}\|J_{3n}\| > \epsilon/4\}) \leq P(E_n^c \cap \{\sqrt{n}\|D_{1n} - D_{3n}\| > \epsilon\lambda_0/8\})
\]

\[
\leq P(E_n^c \cap \{\sqrt{n}\|D_{1n}\| > \epsilon\lambda_0/16\}) + P(E_n^c \cap \{\sqrt{n}\|D_{3n}\| > \epsilon\lambda_0/16\})
\]

\[
\leq P(\sqrt{n}\|M_{1n} - A_n\|^2 > \epsilon\lambda_0^3/128) + P(\sqrt{n}\|M_{3n} - A_n\|^2 > \epsilon\lambda_0^3/128).
\]

In the last inequality we utilize (32). Combining our assumptions, we have

\[
\lim_{n \to \infty} P(E_n^c \cap \{\sqrt{n}\|J_{3n}\| > \epsilon/4\}) = 0.
\]

Similarly

\[
\lim_{n \to \infty} P(E_n^c \cap \{\sqrt{n}\|J_{4n}\| > \epsilon/4\}) = 0.
\]

We conclude our proof. \(\Box\)

References

[1] Chernozhukov, Victor; Chetverikov, Denis; Demirer, Mert; Duflo, Esther; Hansen, Christian; Newey, Whitney; Robins, James: Double/debiased machine learning for treatment and structural parameters. In: The Econometrics Journal 21 (2018), Nr. 1