Online Appendix for "Abadie's Kappa and Weighting Estimators of the Local Average Treatment Effect"

Tymon Słoczyński, S. Derya Uysal, and Jeffrey M. Wooldridge

Review of Recent Empirical Applications. In Section 1 in the main text, we included the following statement: "Our application of weighting to estimate the LATE appears to be somewhat rare in practice, although Abadie's (2003) result is more commonly used to estimate mean characteristics of compliers, as also recommended by Angrist and Pischke (2009). We analyze two samples of applications of instrumental variables to verify this claim. First, our reading of the 30 papers replicated by Young (2022), each of which uses 2SLS, suggests that none of these papers uses weighting estimators of the LATE or applies Abadie's (2003) result for any other purpose. Second, we have also examined whether any of the papers published in journals of the American Economic Association in 2019 and 2020 consider weighting estimators of the LATE. Our best assessment is that the answer is likewise negative. Still, Marx and Turner (2019), Goodman et al. (2020), Leung and O'Leary (2020), and Londoño-Vélez et al. (2020) apply Abadie's (2003) result to estimate mean characteristics of compliers, while Cohodes (2020) uses this result to estimate the control complier mean (CCM), a parameter introduced by Katz et al. (2001)." In what follows, we briefly explain how we reached these conclusions.

To examine whether any of the papers published in journals of the American Economic Association in 2019 and 2020 consider weighting estimators of the LATE, we first searched for the string "instrument" in the main text of each such paper. We retained every paper where this string appeared at least once and it was not immediately clear that the context in which it appeared had nothing to do with instrumental variables (e.g., financial instruments, Texas Instruments).

For every paper that was retained in the search described above and additionally for every paper replicated by Young (2022), we subsequently verified whether it cited any single-authored papers by Alberto Abadie, Markus Frölich, or Zhiqiang Tan, and whether any of the following strings appeared in its main text: "propensity score," "IPW," or "weighting." In the case of any such citation and any appearance of any of these strings, we subsequently read the relevant part of the paper to determine whether Abadie's (2003) result and/or weighting estimators of the LATE may have been used. Our statement in Section 1 in the main text, also restated above, summarizes our conclusions from this exercise.

Proof of Proposition 3.2. We begin with the case of translation invariance. For $\hat{\tau}_u$, we can write

$$\begin{aligned}
\hat{\tau}_u(\mathbf{Y}+k,\mathbf{W}) &= \frac{\left[\sum_{i=1}^N \frac{Z_i}{p(X_i)}\right]^{-1}\sum_{i=1}^N \frac{(Y_i+k)Z_i}{p(X_i)} - \left[\sum_{i=1}^N \frac{1-Z_i}{1-p(X_i)}\right]^{-1}\sum_{i=1}^N \frac{(Y_i+k)(1-Z_i)}{1-p(X_i)}}{\left[\sum_{i=1}^N \frac{Z_i}{p(X_i)}\right]^{-1}\sum_{i=1}^N \frac{D_iZ_i}{p(X_i)} - \left[\sum_{i=1}^N \frac{1-Z_i}{1-p(X_i)}\right]^{-1}\sum_{i=1}^N \frac{D_i(1-Z_i)}{1-p(X_i)}} \\
&= \hat{\tau}_u(\mathbf{Y},\mathbf{W}) + \frac{\left[\sum_{i=1}^N \frac{Z_i}{p(X_i)}\right]^{-1}\sum_{i=1}^N \frac{kZ_i}{p(X_i)} - \left[\sum_{i=1}^N \frac{1-Z_i}{1-p(X_i)}\right]^{-1}\sum_{i=1}^N \frac{k(1-Z_i)}{1-p(X_i)}}{\left[\sum_{i=1}^N \frac{Z_i}{p(X_i)}\right]^{-1}\sum_{i=1}^N \frac{D_iZ_i}{p(X_i)} - \left[\sum_{i=1}^N \frac{1-Z_i}{1-p(X_i)}\right]^{-1}\sum_{i=1}^N \frac{D_i(1-Z_i)}{1-p(X_i)}} \\
&= \hat{\tau}_u(\mathbf{Y},\mathbf{W}),
\end{aligned}$$

which means that $\hat{\tau}_u$ is indeed translation invariant. Similarly,

$$\begin{aligned}
\hat{\tau}_{a,10}(\mathbf{Y}+k,\mathbf{W}) &= \left[\sum_{i=1}^N \kappa_{i1}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i1}(Y_i+k)\right] - \left[\sum_{i=1}^N \kappa_{i0}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i0}(Y_i+k)\right] \\
&= \hat{\tau}_{a,10}(\mathbf{Y},\mathbf{W}) + \left[\sum_{i=1}^N \kappa_{i1}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i1}k\right] - \left[\sum_{i=1}^N \kappa_{i0}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i0}k\right] \\
&= \hat{\tau}_{a,10}(\mathbf{Y},\mathbf{W}),
\end{aligned}$$

which means that $\hat{\tau}_{a,10}$ is translation invariant, too. On the other hand, we can write

$$\begin{aligned}
\hat{\tau}_a(\mathbf{Y}+k,\mathbf{W}) &= \left[\sum_{i=1}^N \kappa_i\right]^{-1}\left[\sum_{i=1}^N \kappa_{i1}(Y_i+k)\right] - \left[\sum_{i=1}^N \kappa_i\right]^{-1}\left[\sum_{i=1}^N \kappa_{i0}(Y_i+k)\right] \\
&= \hat{\tau}_a(\mathbf{Y},\mathbf{W}) + \left[\sum_{i=1}^N \kappa_i\right]^{-1}\left[\sum_{i=1}^N \kappa_{i1}k\right] - \left[\sum_{i=1}^N \kappa_i\right]^{-1}\left[\sum_{i=1}^N \kappa_{i0}k\right] \\
&= \hat{\tau}_a(\mathbf{Y},\mathbf{W}) + k\left(\left[\sum_{i=1}^N \kappa_i\right]^{-1}\left[\sum_{i=1}^N \kappa_{i1}\right] - \left[\sum_{i=1}^N \kappa_i\right]^{-1}\left[\sum_{i=1}^N \kappa_{i0}\right]\right)
\end{aligned}$$

and

$$\begin{aligned}
\hat{\tau}_{a,1}(\mathbf{Y}+k,\mathbf{W}) &= \left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}(Y_i+k)\right] - \left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}(Y_i+k)\right] \\
&= \hat{\tau}_{a,1}(\mathbf{Y},\mathbf{W}) + \left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}k\right] - \left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}k\right] \\
&= \hat{\tau}_{a,1}(\mathbf{Y},\mathbf{W}) + k\left(1 - \left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}\right]\right)
\end{aligned}$$

and also

$$\begin{aligned}
\hat{\tau}_{a,0}(\mathbf{Y}+k,\mathbf{W}) &= \left[\sum_{i=1}^{N}\kappa_{i0}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}(Y_i+k)\right] - \left[\sum_{i=1}^{N}\kappa_{i0}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}(Y_i+k)\right] \\
&= \hat{\tau}_{a,0}(\mathbf{Y},\mathbf{W}) + \left[\sum_{i=1}^{N}\kappa_{i0}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}k\right] - \left[\sum_{i=1}^{N}\kappa_{i0}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}k\right] \\
&= \hat{\tau}_{a,0}(\mathbf{Y},\mathbf{W}) + k\left(\left[\sum_{i=1}^{N}\kappa_{i0}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}\right] - 1\right).
\end{aligned}$$

Even though $k\left(\left[\sum_{i=1}^{N}\kappa_{i0}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}\right] - 1\right) = o_p(1)$, $k\left(1 - \left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}\right]\right) = o_p(1)$, and $k\left(\left[\sum_{i=1}^{N}\kappa_{i}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}\right] - \left[\sum_{i=1}^{N}\kappa_{i}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}\right]\right) = o_p(1)$, none of these objects is generally equal to zero in finite samples, which means that $\hat{\tau}_{a,0}$, $\hat{\tau}_{a,1}$, and $\hat{\tau}_{a}$, respectively, are not translation invariant.

We next turn to the case of scale equivariance. Begin by denoting $\frac{Z_i}{p(X_i)} = \omega_{i1}$ and $\frac{1-Z_i}{1-p(X_i)} = \omega_{i0}$. Then, for $\hat{\tau}_u$, we can write

$$\begin{aligned}
\hat{\tau}_u(f(a\mathbf{Y}),\mathbf{W}) &= \frac{\left[\sum_{i=1}^{N}\omega_{i1}\right]^{-1}\left[\sum_{i=1}^{N}(\alpha_2(aY_i)^{\alpha_1}-\alpha_3)\omega_{i1}\right] - \left[\sum_{i=1}^{N}\omega_{i0}\right]^{-1}\left[\sum_{i=1}^{N}(\alpha_2(aY_i)^{\alpha_1}-\alpha_3)\omega_{i0}\right]}{\left[\sum_{i=1}^{N}\omega_{i1}\right]^{-1}\left[\sum_{i=1}^{N}D_i\omega_{i1}\right] - \left[\sum_{i=1}^{N}\omega_{i0}\right]^{-1}\left[\sum_{i=1}^{N}D_i\omega_{i0}\right]} \\
&= \frac{\left[\sum_{i=1}^{N}\omega_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\alpha_2(aY_i)^{\alpha_1}\omega_{i1}\right] - \left[\sum_{i=1}^{N}\omega_{i0}\right]^{-1}\left[\sum_{i=1}^{N}\alpha_2(aY_i)^{\alpha_1}\omega_{i0}\right]}{\left[\sum_{i=1}^{N}\omega_{i1}\right]^{-1}\left[\sum_{i=1}^{N}D_i\omega_{i1}\right] - \left[\sum_{i=1}^{N}\omega_{i0}\right]^{-1}\left[\sum_{i=1}^{N}D_i\omega_{i0}\right]} \\
&\quad - \frac{\alpha_3\left(\left[\sum_{i=1}^{N}\omega_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\omega_{i1}\right] - \left[\sum_{i=1}^{N}\omega_{i0}\right]^{-1}\left[\sum_{i=1}^{N}\omega_{i0}\right]\right)}{\left[\sum_{i=1}^{N}\omega_{i1}\right]^{-1}\left[\sum_{i=1}^{N}D_i\omega_{i1}\right] - \left[\sum_{i=1}^{N}\omega_{i0}\right]^{-1}\left[\sum_{i=1}^{N}D_i\omega_{i0}\right]} \\
&= \frac{a^{\alpha_1}\left(\left[\sum_{i=1}^{N}\omega_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\alpha_2 Y_i^{\alpha_1}\omega_{i1}\right] - \left[\sum_{i=1}^{N}\omega_{i0}\right]^{-1}\left[\sum_{i=1}^{N}\alpha_2 Y_i^{\alpha_1}\omega_{i0}\right]\right)}{\left[\sum_{i=1}^{N}\omega_{i1}\right]^{-1}\left[\sum_{i=1}^{N}D_i\omega_{i1}\right] - \left[\sum_{i=1}^{N}\omega_{i0}\right]^{-1}\left[\sum_{i=1}^{N}D_i\omega_{i0}\right]}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{a^{\alpha_1}\left(\left[\sum_{i=1}^N \omega_{i1}\right]^{-1}\left[\sum_{i=1}^N \alpha_2 Y_i^{\alpha_1}\omega_{i1}\right]-\left[\sum_{i=1}^N \omega_{i0}\right]^{-1}\left[\sum_{i=1}^N \alpha_2 Y_i^{\alpha_1}\omega_{i0}\right]\right)}{\left[\sum_{i=1}^N \omega_{i1}\right]^{-1}\left[\sum_{i=1}^N D_i\omega_{i1}\right]-\left[\sum_{i=1}^N \omega_{i0}\right]^{-1}\left[\sum_{i=1}^N D_i\omega_{i0}\right]} \\
&\quad \pm \frac{a^{\alpha_1}\alpha_3\left(\left[\sum_{i=1}^N \omega_{i1}\right]^{-1}\left[\sum_{i=1}^N \omega_{i1}\right]-\left[\sum_{i=1}^N \omega_{i0}\right]^{-1}\left[\sum_{i=1}^N \omega_{i0}\right]\right)}{\left[\sum_{i=1}^N \omega_{i1}\right]^{-1}\left[\sum_{i=1}^N D_i\omega_{i1}\right]-\left[\sum_{i=1}^N \omega_{i0}\right]^{-1}\left[\sum_{i=1}^N D_i\omega_{i0}\right]} \\
&= \frac{a^{\alpha_1}\left(\left[\sum_{i=1}^N \omega_{i1}\right]^{-1}\left[\sum_{i=1}^N(\alpha_2 Y_i^{\alpha_1}-\alpha_3)\omega_{i1}\right]-\left[\sum_{i=1}^N \omega_{i0}\right]^{-1}\left[\sum_{i=1}^N(\alpha_2 Y_i^{\alpha_1}-\alpha_3)\omega_{i0}\right]\right)}{\left[\sum_{i=1}^N \omega_{i1}\right]^{-1}\left[\sum_{i=1}^N D_i\omega_{i1}\right]-\left[\sum_{i=1}^N \omega_{i0}\right]^{-1}\left[\sum_{i=1}^N D_i\omega_{i0}\right]} \\
&\quad + \frac{a^{\alpha_1}\alpha_3\left(\left[\sum_{i=1}^N \omega_{i1}\right]^{-1}\left[\sum_{i=1}^N \omega_{i1}\right]-\left[\sum_{i=1}^N \omega_{i0}\right]^{-1}\left[\sum_{i=1}^N \omega_{i0}\right]\right)}{\left[\sum_{i=1}^N \omega_{i1}\right]^{-1}\left[\sum_{i=1}^N D_i\omega_{i1}\right]-\left[\sum_{i=1}^N \omega_{i0}\right]^{-1}\left[\sum_{i=1}^N D_i\omega_{i0}\right]} \\
&= a^{\alpha_1}\hat{\tau}_u\left(f(\mathbf{Y}),\mathbf{W}\right),
\end{aligned}
$$

which means that $\hat{\tau}_u$ is indeed scale equivariant. Similarly,

$$
\begin{aligned}
\hat{\tau}_{a,10}\left(f(a\mathbf{Y}),\mathbf{W}\right) &= \left[\sum_{i=1}^N \kappa_{i1}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i1}(\alpha_2(aY_i)^{\alpha_1}-\alpha_3)\right]-\left[\sum_{i=1}^N \kappa_{i0}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i0}(\alpha_2(aY_i)^{\alpha_1}-\alpha_3)\right] \\
&= \left[\sum_{i=1}^N \kappa_{i1}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i1}\alpha_2(aY_i)^{\alpha_1}\right]-\left[\sum_{i=1}^N \kappa_{i0}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i0}\alpha_2(aY_i)^{\alpha_1}\right] \\
&\quad - \alpha_3\left(\left[\sum_{i=1}^N \kappa_{i1}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i1}\right]-\left[\sum_{i=1}^N \kappa_{i0}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i0}\right]\right) \\
&= a^{\alpha_1}\left(\left[\sum_{i=1}^N \kappa_{i1}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i1}\alpha_2 Y_i^{\alpha_1}\right]-\left[\sum_{i=1}^N \kappa_{i0}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i0}\alpha_2 Y_i^{\alpha_1}\right]\right) \\
&\quad \pm a^{\alpha_1}\alpha_3\left(\left[\sum_{i=1}^N \kappa_{i1}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i1}\right]-\left[\sum_{i=1}^N \kappa_{i0}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i0}\right]\right) \\
&= a^{\alpha_1}\left(\left[\sum_{i=1}^N \kappa_{i1}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i1}(\alpha_2 Y_i^{\alpha_1}-\alpha_3)\right]-\left[\sum_{i=1}^N \kappa_{i0}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i0}(\alpha_2 Y_i^{\alpha_1}-\alpha_3)\right]\right) \\
&\quad + a^{\alpha_1}\alpha_3\left(\left[\sum_{i=1}^N \kappa_{i1}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i1}\right]-\left[\sum_{i=1}^N \kappa_{i0}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i0}\right]\right) \\
&= a^{\alpha_1}\hat{\tau}_{a,10}\left(f(\mathbf{Y}),\mathbf{W}\right),
\end{aligned}
$$

which means that $\hat{\tau}_{a,10}$ is scale equivariant, too. On the other hand, we can write

$$\begin{aligned}
\hat{\tau}_a\left(f(a\mathbf{Y}),\mathbf{W}\right) &= \left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}(\alpha_2(aY_i)^{\alpha_1}-\alpha_3)\right]-\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}(\alpha_2(aY_i)^{\alpha_1}-\alpha_3)\right] \\
&= \left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}\alpha_2(aY_i)^{\alpha_1}\right]-\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}\alpha_2(aY_i)^{\alpha_1}\right] \\
&\quad -\ \alpha_3\left(\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}\right]-\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}\right]\right) \\
&= a^{\alpha_1}\left(\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}\alpha_2Y_i^{\alpha_1}\right]-\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}\alpha_2Y_i^{\alpha_1}\right]\right) \\
&\quad -\ \alpha_3\left(\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}\right]-\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}\right]\right) \\
&\quad \pm\ a^{\alpha_1}\alpha_3\left(\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}\right]-\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}\right]\right) \\
&= a^{\alpha_1}\left(\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}(\alpha_2Y_i^{\alpha_1}-\alpha_3)\right]-\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}(\alpha_2Y_i^{\alpha_1}-\alpha_3)\right]\right) \\
&\quad -\ (\alpha_3-a^{\alpha_1}\alpha_3)\left(\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}\right]-\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}\right]\right) \\
&= a^{\alpha_1}\hat{\tau}_a\left(f(\mathbf{Y}),\mathbf{W}\right)\ -\ (\alpha_3-a^{\alpha_1}\alpha_3)\left(\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}\right]-\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}\right]\right)
\end{aligned}$$

and

$$\begin{aligned}
\hat{\tau}_{a.1}\left(f(a\mathbf{Y}),\mathbf{W}\right) &= \left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}(\alpha_2(aY_i)^{\alpha_1}-\alpha_3)\right]-\left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}(\alpha_2(aY_i)^{\alpha_1}-\alpha_3)\right] \\
&= \left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}\alpha_2(aY_i)^{\alpha_1}\right]-\left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}\alpha_2(aY_i)^{\alpha_1}\right] \\
&\quad -\ \alpha_3\left(\left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}\right]-\left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}\right]\right) \\
&= a^{\alpha_1}\left(\left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}\alpha_2Y_i^{\alpha_1}\right]-\left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}\alpha_2Y_i^{\alpha_1}\right]\right)
\end{aligned}$$

$$
\begin{aligned}
&- \ \alpha_3 \left(\left[\sum_{i=1}^{N} \kappa_{i1} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} \right] - \left[\sum_{i=1}^{N} \kappa_{i1} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} \right] \right) \\
&\pm \ a^{\alpha_1}\alpha_3 \left(\left[\sum_{i=1}^{N} \kappa_{i1} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} \right] - \left[\sum_{i=1}^{N} \kappa_{i1} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} \right] \right) \\
= \ & a^{\alpha_1} \left(\left[\sum_{i=1}^{N} \kappa_{i1} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1}(\alpha_2 Y_i^{\alpha_1} - \alpha_3) \right] - \left[\sum_{i=1}^{N} \kappa_{i1} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0}(\alpha_2 Y_i^{\alpha_1} - \alpha_3) \right] \right) \\
&- \ (\alpha_3 - a^{\alpha_1}\alpha_3) \left(\left[\sum_{i=1}^{N} \kappa_{i1} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} \right] - \left[\sum_{i=1}^{N} \kappa_{i1} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} \right] \right) \\
= \ & a^{\alpha_1} \hat{\tau}_{a,1} \left(f(\mathbf{Y}), \mathbf{W} \right) \ - \ (\alpha_3 - a^{\alpha_1}\alpha_3) \left(1 - \left[\sum_{i=1}^{N} \kappa_{i1} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} \right] \right)
\end{aligned}
$$

and also

$$
\begin{aligned}
\hat{\tau}_{a,0} \left(f(a\mathbf{Y}), \mathbf{W} \right) \ = \ & \left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1}(\alpha_2 (aY_i)^{\alpha_1} - \alpha_3) \right] - \left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0}(\alpha_2 (aY_i)^{\alpha_1} - \alpha_3) \right] \\
= \ & \left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1}\alpha_2 (aY_i)^{\alpha_1} \right] - \left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0}\alpha_2 (aY_i)^{\alpha_1} \right] \\
&- \ \alpha_3 \left(\left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} \right] - \left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} \right] \right) \\
= \ & a^{\alpha_1} \left(\left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1}\alpha_2 Y_i^{\alpha_1} \right] - \left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0}\alpha_2 Y_i^{\alpha_1} \right] \right) \\
&- \ \alpha_3 \left(\left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} \right] - \left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} \right] \right) \\
&\pm \ a^{\alpha_1}\alpha_3 \left(\left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} \right] - \left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} \right] \right) \\
= \ & a^{\alpha_1} \left(\left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1}(\alpha_2 Y_i^{\alpha_1} - \alpha_3) \right] - \left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0}(\alpha_2 Y_i^{\alpha_1} - \alpha_3) \right] \right) \\
&- \ (\alpha_3 - a^{\alpha_1}\alpha_3) \left(\left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} \right] - \left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} \right] \right) \\
= \ & a^{\alpha_1} \hat{\tau}_{a,0} \left(f(\mathbf{Y}), \mathbf{W} \right) \ - \ (\alpha_3 - a^{\alpha_1}\alpha_3) \left(\left[\sum_{i=1}^{N} \kappa_{i0} \right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} \right] - 1 \right).
\end{aligned}
$$

Even though $(\alpha_3 - a^{\alpha_1}\alpha_3)\left(\left[\sum_{i=1}^N \kappa_{i0}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i1}\right] - 1\right)$, $(\alpha_3 - a^{\alpha_1}\alpha_3)\left(1 - \left[\sum_{i=1}^N \kappa_{i1}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i0}\right]\right)$, and $(\alpha_3 - a^{\alpha_1}\alpha_3)\left(\left[\sum_{i=1}^N \kappa_{i}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i1}\right] - \left[\sum_{i=1}^N \kappa_{i}\right]^{-1}\left[\sum_{i=1}^N \kappa_{i0}\right]\right)$ are all $o_p(1)$, none of these objects is generally equal to zero in finite samples, which means that $\hat{\tau}_{a,0}$, $\hat{\tau}_{a,1}$, and $\hat{\tau}_a$, respectively, are not scale equivariant.

Proof of Proposition 3.5. The sample moment conditions in equation (9) in the main text can be written as

$$N^{-1}\sum_{i=1}^{N} X_i \frac{Z_i - \hat{p}_{cb}(X_i)}{\hat{p}_{cb}(X_i)\left(1 - \hat{p}_{cb}(X_i)\right)} = 0.$$

If X includes a constant, then one of these moment conditions is $N^{-1}\sum_{i=1}^N \frac{Z_i - \hat{p}_{cb}(X_i)}{\hat{p}_{cb}(X_i)(1-\hat{p}_{cb}(X_i))} = 0$, and this, together with Remark 2.2, guarantees that $N^{-1}\sum_{i=1}^N \hat{\kappa}_{i1} = N^{-1}\sum_{i=1}^N \hat{\kappa}_{i0}$, where $\hat{\kappa}_1$ and $\hat{\kappa}_0$ use the covariate-balancing instrument propensity score, $\hat{p}_{cb}(X)$. If $N^{-1}\sum_{i=1}^N \hat{\kappa}_{i1} = N^{-1}\sum_{i=1}^N \hat{\kappa}_{i0}$, then it is also the case that $\hat{\tau}_t^{cb}$ $(= \hat{\tau}_{a,1}^{cb})$, $\hat{\tau}_{a,0}^{cb}$, and $\hat{\tau}_{a,10}^{cb}$ are numerically identical to each other. They are also identical to $\hat{\tau}_u^{cb}$ following the result in Heiler (2022), which says that $\hat{\tau}_u^{cb}$ is identical to $\hat{\tau}_t^{cb}$.

Asymptotic Derivations. As stated in Section 3.6 in the main text, all the weighting estimators considered in this paper can be represented as an M-estimator. Thus, for the asymptotic distributions of each estimator, we can rely on the results regarding the asymptotics of the M-estimator. The M-estimator, denoted as $\hat{\theta}$, for θ, a $K \times 1$ unknown parameter vector, can be derived as the solution to the sample moment equation

$$N^{-1}\sum_{i=1}^{N} \psi(O_i, \hat{\theta}) = 0,$$

where O_i is the observed data. Thus, $\hat{\theta}$ is the estimator of θ that satisfies the population relation $\mathrm{E}\left[\psi(O,\theta)\right] = 0$.[1] Under standard regularity conditions[2] and assuming that the relevant moments exist, i.e. $\mathrm{E}\left[\frac{\partial\psi(O,\theta)}{\partial\theta'}\right]$ exists and is nonsingular, and $\mathrm{E}\left[\psi(O,\theta)\psi(O,\theta)'\right]$ exists and is finite, the asymp-

[1] See, for example, Wooldridge (2010) and Boos and Stefanski (2013) for more on M-estimation.

[2] Theorem 7.2 in Boos and Stefanski (2013) states the conditions for the asymptotic normality of M-estimators. A more general treatment of these regularity conditions can be found in Newey and McFadden (1994).

totic distribution of an M-estimator is given by

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, A^{-1}VA^{-1\prime}) \tag{A.1}$$

with

$$\begin{aligned} A &= \mathrm{E}\left[\frac{\partial \psi(O,\theta)}{\partial \theta'}\right], \\ V &= \mathrm{E}\left[\psi(O,\theta)\psi(O,\theta)'\right]. \end{aligned}$$

We use different combinations of moment functions listed in Table A.1 for each of the weighting estimators. For example, if τ_{LATE} is estimated by $\hat{\tau}_a^{ml}$, then

$$\psi_a^{ml} = \begin{pmatrix} \psi_\alpha^{ml} \\ \psi_\Gamma \\ \psi_\Delta \\ \psi_{\tau_a} \end{pmatrix}$$

is used as the vector of moment functions. Under standard regularity conditions for M-estimation, all of the LATE estimators discussed above will be asymptotically normal with different asymptotic variances. A joint estimation of α and τ_{LATE} allows us to conduct inference based on the asymptotic variance-covariance matrix of an M-estimator given in (A.1) without explicitly deriving the asymptotic distribution of τ_{LATE}. At the same time, the M-estimation framework also facilitates the derivations of the asymptotic variance terms for each of the LATE estimators. In what follows, we provide asymptotic distributions of all the estimators discussed in the body of the paper.

We first introduce some additional notation in order to simplify the representation of the asymptotic variances. Let us denote the population counterpart of the numerator of the estimators $\hat{\tau}_a$, $\hat{\tau}_{a,1}$ $(= \hat{\tau}_t)$, $\hat{\tau}_{a,0}$, and $\hat{\tau}_u$ by Δ, i.e.,

$$\Delta \equiv \mathrm{E}\left[Y \frac{Z - p(X)}{p(X)\,(1 - p(X))}\right]. \tag{A.2}$$

Recall that the expectation on the right hand side is equal to $\mathrm{E}\left[(\kappa_1 - \kappa_0)\, Y\right]$; see equation (2) in the main text. Next, denote $\mathrm{E}(\kappa_1 Y)$ and $\mathrm{E}(\kappa_0 Y)$ by Δ_1 and Δ_0, respectively. Alternatively, we can write the expectation in equation (A.2) as follows:

$$\mathrm{E}\left[Y\frac{Z-p(X)}{p(X)\,(1-p(X))}\right] \;=\; \mathrm{E}\left[\frac{YZ}{p(X)}\right]-\mathrm{E}\left[\frac{Y(1-Z)}{1-p(X)}\right].$$

We denote $\mathrm{E}\left[\frac{YZ}{p(X)}\right]$ by μ_1 and $\mathrm{E}\left[\frac{Y(1-Z)}{1-p(X)}\right]$ by μ_0. Symmetrically, we denote $\mathrm{E}\left[\frac{DZ}{p(X)}\right]$ and $\mathrm{E}\left[\frac{D(1-Z)}{1-p(X)}\right]$ by m_1 and m_0. Additionally, the population proportion of compliers is denoted by Γ, Γ_1, or Γ_0, depending on which sample mean is used to estimate the population parameter, i.e., $\Gamma \equiv \mathrm{E}(\kappa)$, $\Gamma_1 \equiv \mathrm{E}(\kappa_1)$, and $\Gamma_0 \equiv \mathrm{E}(\kappa_0)$. Note that $\tau_{\mathrm{LATE}} = \frac{\Delta}{\Gamma} = \frac{\Delta}{\Gamma_1} = \frac{\Delta}{\Gamma_0} = \frac{\Delta_1}{\Gamma_1} - \frac{\Delta_0}{\Gamma_0} = \frac{\mu_1-\mu_0}{m_1-m_0}$. When the population parameters are replaced by their sample counterparts, we obtain the estimators $\hat{\tau}_a$, $\hat{\tau}_{a,1}$, $\hat{\tau}_{a,0}$, $\hat{\tau}_{a,10}$, and $\hat{\tau}_t$, respectively. If normalized weights are used to estimate μ_z and m_z for $z = 0, 1$, the resulting ratio estimator corresponds to $\hat{\tau}_u$. This is of course without taking into account how the propensity score is estimated.

In what follows, we first consider ML-based estimation of the instrument propensity score. For the estimator $\hat{\tau}_a^{ml}$, we use the moment functions ψ_α^{ml}, ψ_Δ, and ψ_Γ. Based on the result given in equation (A.1), the asymptotic distribution of $\hat{\tau}_a^{ml}$ can be derived as follows:

$$\sqrt{N}\left(\hat{\tau}_a^{ml} - \tau_{\mathrm{LATE}}\right) \xrightarrow{d} \mathcal{N}(0, V_{\tau_a^{ml}}),$$

where

$$V_{\tau_a^{ml}} \;=\; -\left(\frac{1}{\Gamma}E_{\Delta,\alpha} - \frac{\tau_{\mathrm{LATE}}}{\Gamma}E_{\Gamma,\alpha}\right)\left(-E_{H_\alpha^{ml}}\right)^{-1}\left(\frac{1}{\Gamma}E_{\Delta,\alpha} - \frac{\tau_{\mathrm{LATE}}}{\Gamma}E_{\Gamma,\alpha}\right)' \;+\; \mathrm{E}\left[\left(\frac{1}{\Gamma}\psi_\Delta - \frac{\tau_{\mathrm{LATE}}}{\Gamma}\psi_\Gamma\right)^2\right]$$

with

$$\begin{aligned}
\psi_\Delta &= \frac{Z_iY_i}{F(X_i,\alpha)} - \frac{(1-Z_i)Y_i}{1-F(X_i,\alpha)} - \Delta,\\
\psi_\Gamma &= 1 - \frac{(1-Z_i)D_i}{1-F(X_i,\alpha)} - \frac{Z_i(1-D_i)}{F(X_i,\alpha)} - \Gamma,\\
E_{\Delta,\alpha} &= \mathrm{E}\left[\frac{\partial\psi_\Delta}{\partial\alpha}\right] = \mathrm{E}\left[-\left(\frac{YZ}{F(X,\alpha)^2} + \frac{Y(1-Z)}{(1-F(X,\alpha))^2}\right)\nabla_\alpha F(X,\alpha)\right],\\
E_{\Gamma,\alpha} &= \mathrm{E}\left[\frac{\partial\psi_\Gamma}{\partial\alpha}\right] = \mathrm{E}\left[\left(\frac{(1-D)Z}{F(X,\alpha)^2} - \frac{D(1-Z)}{(1-F(X,\alpha))^2}\right)\nabla_\alpha F(X,\alpha)\right],
\end{aligned}$$

$$E_{H_\alpha^{ml}} = \mathrm{E}\left[\frac{\partial \psi_\alpha^{ml}}{\partial \alpha'}\right] = \mathrm{E}\left[H(X,\alpha)\right],$$

and $H(X,\alpha)$ denotes the Hessian of the log-likelihood of α.

The estimators $\hat{\tau}_{a,1}^{ml}$ ($=\hat{\tau}_t^{ml}$) and $\hat{\tau}_{a,0}^{ml}$ use the same moment functions for α and Δ as $\hat{\tau}_a^{ml}$. However, they estimate the population proportion of compliers using the moment functions derived from the population relations Γ_1 and Γ_0, respectively. The variances of $\hat{\tau}_{a,1}^{ml}$ and $\hat{\tau}_{a,0}^{ml}$ have the same form as $\hat{\tau}_a^{ml}$, where Γ is replaced with Γ_1 and Γ_0. Thus, the asymptotic distributions of $\hat{\tau}_{a,1}^{ml}$ and $\hat{\tau}_{a,0}^{ml}$ can be summarized as follows:

$$\sqrt{N}\left(\hat{\tau}_{a,1}^{ml} - \tau_{\mathrm{LATE}}\right) \xrightarrow{d} \mathcal{N}(0, V_{\tau_{a,1}^{ml}}),$$

where

$$V_{\tau_{a,1}^{ml}} = -\left(\frac{1}{\Gamma_1}E_{\Delta,\alpha} - \frac{\tau_{\mathrm{LATE}}}{\Gamma_1}E_{\Gamma_1,\alpha}\right)\left(-E_{H_\alpha^{ml}}\right)^{-1}\left(\frac{1}{\Gamma_1}E_{\Delta,\alpha} - \frac{\tau_{\mathrm{LATE}}}{\Gamma_1}E_{\Gamma_1,\alpha}\right)' + \mathrm{E}\left[\left(\frac{1}{\Gamma_1}\psi_\Delta - \frac{\tau_{\mathrm{LATE}}}{\Gamma_1}\psi_{\Gamma_1}\right)^2\right]$$

with

$$\begin{aligned}
\psi_{\Gamma_1} &= \frac{Z_i Y_i}{F(X_i,\alpha)} - \frac{(1-Z_i)Y_i}{1-F(X_i,\alpha)} - \Gamma_1,\\
E_{\Gamma_1,\alpha} &= \mathrm{E}\left[-\left(\frac{DZ}{F(X,\alpha)^2} + \frac{D(1-Z)}{(1-F(X,\alpha))^2}\right)\nabla_\alpha F(X,\alpha)\right],
\end{aligned}$$

and

$$\sqrt{N}\left(\hat{\tau}_{a,0}^{ml} - \tau_{\mathrm{LATE}}\right) \xrightarrow{d} \mathcal{N}(0, V_{\tau_{a,0}^{ml}}),$$

where

$$V_{\tau_{a,0}^{ml}} = -\left(\frac{1}{\Gamma_0}E_{\Delta,\alpha} - \frac{\tau_{\mathrm{LATE}}}{\Gamma_0}E_{\Gamma_0,\alpha}\right)\left(-E_{H_\alpha^{ml}}\right)^{-1}\left(\frac{1}{\Gamma_0}E_{\Delta,\alpha} - \frac{\tau_{\mathrm{LATE}}}{\Gamma_0}E_{\Gamma_0,\alpha}\right)' + \mathrm{E}\left[\left(\frac{1}{\Gamma_0}\psi_\Delta - \frac{\tau_{\mathrm{LATE}}}{\Gamma_0}\psi_{\Gamma_0}\right)^2\right]$$

with

$$\psi_{\Gamma_0} = \frac{Z_i(D_i-1)}{F(X_i,\alpha)} - \frac{(1-Z_i)(D_i-1)}{1-F(X_i,\alpha)} - \Gamma_0,$$

$$E_{\Gamma_0,\alpha} \;=\; \mathrm{E}\left[\frac{\partial\psi_{\Gamma_0}}{\partial\alpha}\right] = \mathrm{E}\left[-\left(\frac{(D-1)Z}{F(X,\alpha)^2} + \frac{(D-1)(1-Z)}{(1-F(X,\alpha))^2}\right)\nabla_\alpha F(X,\alpha)\right].$$

The estimator $\hat{\tau}^{ml}_{a,10}$ is essentially the difference of two ratio estimators whose covariance is zero. Thus, the variance of the difference is the sum of variances of the two estimators. It follows that

$$\sqrt{N}\left(\hat{\tau}^{ml}_{a,10} - \tau_{\text{LATE}}\right) \xrightarrow{d} \mathcal{N}(0, V_{\tau^{ml}_{a,10}}),$$

where

$$\begin{aligned} V_{\tau^{ml}_{a,10}} \;&=\; -\left(\frac{E_{\Delta_1,\alpha}}{\Gamma_1} - \frac{E_{\Delta_0,\alpha}}{\Gamma_0} - \frac{\Delta_1 E_{\Gamma_1,\alpha}}{\Gamma_1^2} + \frac{\Delta_0 E_{\Gamma_0,\alpha}}{\Gamma_0^2}\right)\left(-E_{H^{ml}_\alpha}\right)^{-1}\left(\frac{E_{\Delta_1,\alpha}}{\Gamma_1} - \frac{E_{\Delta_0,\alpha}}{\Gamma_0} - \frac{\Delta_1 E_{\Gamma_1,\alpha}}{\Gamma_1^2} + \frac{\Delta_0 E_{\Gamma_0,\alpha}}{\Gamma_0^2}\right)' \\ &+\; \mathrm{E}\left(\frac{1}{\Gamma_1}\psi_{\Delta_1} - \frac{\Delta_1}{\Gamma_1^2}\psi_{\Gamma_1}\right)^2 + \mathrm{E}\left(\frac{1}{\Gamma_0}\psi_{\Delta_0} - \frac{\Delta_0}{\Gamma_0^2}\psi_{\Gamma_0}\right)^2 \end{aligned}$$

with

$$\begin{aligned} \psi_{\Delta_1} \;&=\; D_i \frac{Z_i - F(X_i,\alpha)}{F(X_i,\alpha)(1-F(X_i,\alpha))} Y_i - \Delta_1, \\ \psi_{\Delta_0} \;&=\; (1-D_i)\frac{(1-Z_i) - (1-F(X_i,\alpha))}{F(X_i,\alpha)(1-F(X_i,\alpha))} Y_i - \Delta_0, \\ E_{\Delta_1,\alpha} \;&=\; \mathrm{E}\left[\frac{\partial\psi_{\Delta_1}}{\partial\alpha}\right] = \mathrm{E}\left[-\left(\frac{DYZ}{F(X,\alpha)^2} + \frac{DY(1-Z)}{(1-F(X,\alpha))^2}\right)\nabla_\alpha F(X,\alpha)\right], \\ E_{\Delta_0,\alpha} \;&=\; \mathrm{E}\left[\frac{\partial\psi_{\Delta_0}}{\partial\alpha}\right] = \mathrm{E}\left[-\left(\frac{(D-1)YZ}{F(X,\alpha)^2} + \frac{(D-1)Y(1-Z)}{(1-F(X,\alpha))^2}\right)\nabla_\alpha F(X,\alpha)\right]. \end{aligned}$$

Finally, we examine the estimators $\hat{\tau}^{ml}_u$ and $\hat{\tau}^{cb}_u$. The key distinction between them is the method used to estimate the instrument propensity score. The instrument propensity score is estimated using maximum likelihood for $\hat{\tau}^{ml}_u$, while it is estimated using covariate balancing for $\hat{\tau}^{cb}_u$. As a result, the former employs ψ^{ml}_α whereas the latter uses ψ^{cb}_α within the M-estimation framework. Thus, the moment function related to the estimation of α and the appropriate moment functions that take normalization into account can be used to obtain the asymptotic distribution:

$$\sqrt{N}\left(\hat{\tau}^{ml}_u - \tau_{\text{LATE}}\right) \xrightarrow{d} \mathcal{N}(0, V_{\tau^{ml}_u}),$$

where

$$
\begin{aligned}
V_{\tau_u^{ml}} &= -\left(\frac{1}{\Gamma}(E_{\mu_1,\alpha}-E_{\mu_0,\alpha})-\frac{\Delta}{\Gamma^2}(E_{m_1,\alpha}-E_{m_0,\alpha})\right)\left(-E_{H_\alpha^{ml}}\right)^{-1}\left(\frac{1}{\Gamma}(E_{\mu_1,\alpha}-E_{\mu_0,\alpha})-\frac{\Delta}{\Gamma^2}(E_{m_1,\alpha}-E_{m_0,\alpha})\right)' \\
&+ \mathrm{E}\left(\frac{1}{\Gamma}\psi_{\mu_1}-\frac{\Delta}{\Gamma^2}\psi_{m_1}\right)^2+\mathrm{E}\left(\frac{1}{\Gamma}\psi_{\mu_0}-\frac{\Delta}{\Gamma^2}\psi_{m_0}\right)^2
\end{aligned}
$$

with

$$
\begin{aligned}
\psi_{\mu_1} &= \frac{Z_i(Y_i-\mu_1)}{F(X_i,\alpha)}, \quad \psi_{\mu_0}=\frac{(1-Z_i)(Y_i-\mu_0)}{1-F(X_i,\alpha)}, \\
\psi_{m_1} &= \frac{Z_i(D_i-m_1)}{F(X_i,\alpha)}, \quad \psi_{m_0}=\frac{(1-Z_i)(D_i-m_0)}{1-F(X_i,\alpha)}, \\
E_{\mu_1,\alpha} &= \mathrm{E}\left[\frac{\partial\psi_{\mu_1}}{\partial\alpha}\right]=\mathrm{E}\left[-\frac{Z(Y-\mu_1)}{F(X,\alpha)^2}\nabla_\alpha F(X,\alpha)\right], \\
E_{\mu_0,\alpha} &= \mathrm{E}\left[\frac{\partial\psi_{\mu_0}}{\partial\alpha}\right]=\mathrm{E}\left[-\frac{(1-Z)(Y-\mu_1)}{(1-F(X,\alpha))^2}\nabla_\alpha F(X,\alpha)\right], \\
E_{m_1,\alpha} &= \mathrm{E}\left[\frac{\partial\psi_{m_1}}{\partial\alpha}\right]=\mathrm{E}\left[-\frac{Z(D-m_1)}{F(X,\alpha)^2}\nabla_\alpha F(X,\alpha)\right], \\
E_{m_0,\alpha} &= \mathrm{E}\left[\frac{\partial\psi_{m_0}}{\partial\alpha}\right]=\mathrm{E}\left[-\frac{(1-Z)(D-m_1)}{(1-F(X,\alpha))^2}\nabla_\alpha F(X,\alpha)\right],
\end{aligned}
$$

and

$$
\sqrt{N}\left(\hat{\tau}_u^{cb}-\tau_{\mathrm{LATE}}\right)\xrightarrow{d}\mathcal{N}(0,V_{\tau_u^{cb}}),
$$

where

$$
\begin{aligned}
V_{\tau_u^{cb}} &= \left(\frac{1}{\Gamma}(E_{\mu_1,\alpha}-E_{\mu_0,\alpha})-\frac{\Delta}{\Gamma^2}(E_{m_1,\alpha}-E_{m_0,\alpha})\right)\left(-E_{H_\alpha^{cb}}\right)^{-1}V_{\alpha,cb}\left(-E_{H_\alpha^{cb}}\right)^{-1}\left(\frac{1}{\Gamma}(E_{\mu_1,\alpha}-E_{\mu_0,\alpha})-\frac{\Delta}{\Gamma^2}(E_{m_1,\alpha}-E_{m_0,\alpha})\right)' \\
&- 2\left(\frac{1}{\Gamma}(V_{\mu_1,\alpha}-V_{\mu_0,\alpha})-\frac{\Delta}{\Gamma^2}(V_{m_1,\alpha}-V_{m_0,\alpha})\right)\left(-E_{H_\alpha^{cb}}\right)^{-1}\left(\frac{1}{\Gamma}(E_{\mu_1,\alpha}-E_{\mu_0,\alpha})-\frac{\Delta}{\Gamma^2}(E_{m_1,\alpha}-E_{m_0,\alpha})\right)' \\
&+ \mathrm{E}\left(\frac{1}{\Gamma}\psi_{\mu_1}-\frac{\Delta}{\Gamma^2}\psi_{m_1}\right)^2 + \mathrm{E}\left(\frac{1}{\Gamma}\psi_{\mu_0}-\frac{\Delta}{\Gamma^2}\psi_{m_0}\right)^2
\end{aligned}
$$

with

$$
\begin{aligned}
E_{H_\alpha^{cb}} &= \mathrm{E}\left[\frac{\partial\psi_\alpha^{cb}}{\partial\alpha}\right], \\
V_\alpha^{cb} &= \mathrm{E}\left[\psi_\alpha^{cb}(\cdot)\psi_\alpha^{cb}(\cdot)'\right],
\end{aligned}
$$

$$
\begin{aligned}
V_{\mu_1,\alpha} &= \mathrm{E}\left[\psi_{\mu_1}\psi_\alpha^{cb}\right] = \mathrm{E}\left[\frac{Z_i(Y_i-\mu_1)}{F(X_i,\alpha)^2}\frac{(Z_i-F(X_i,\alpha))}{(1-F(X_i,\alpha))}X_i\right], \\
V_{\mu_0,\alpha} &= \mathrm{E}\left[\psi_{\mu_0}\psi_\alpha^{cb}\right] = \mathrm{E}\left[\frac{(1-Z_i)(Y_i-\mu_0)}{(1-F(X_i,\alpha))^2}\frac{(Z_i-F(X_i,\alpha))}{F(X_i,\alpha)}X_i\right], \\
V_{m_1,\alpha} &= \mathrm{E}\left[\psi_{m_1}\psi_\alpha^{cb}\right] = \mathrm{E}\left[\frac{Z_i(D_i-m_1)}{F(X_i,\alpha)}\frac{Z_i-F(X_i,\alpha)}{F(X_i,\alpha)(1-F(X_i,\alpha))}X_i\right], \\
V_{m_0,\alpha} &= \mathrm{E}\left[\psi_{m_0}\psi_\alpha^{cb}\right] = \mathrm{E}\left[\frac{(1-Z_i)(D_i-m_0)}{1-F(X_i,\alpha)}\frac{Z_i-F(X_i,\alpha)}{F(X_i,\alpha)(1-F(X_i,\alpha))}X_i\right].
\end{aligned}
$$

In fact, $V_{\tau_u^{ml}}$ has the same structure as $V_{\tau_u^{cb}}$, but it enjoys some additional simplifications when the ML-based moment condition is used to estimate $p(X)$. Namely, $\mathrm{E}\left[\frac{\partial\psi_\alpha^{cb}(\cdot)}{\partial\alpha'}\right] = -\mathrm{E}\left[\psi_\alpha^{cb}(\cdot)\psi_\alpha^{cb}(\cdot)'\right]$, $\mathrm{E}\left[\frac{\partial\psi_{\mu_z}}{\partial\alpha}\right] = -\mathrm{E}\left[\psi_{\mu_z}(\cdot)\psi_\alpha^{cb}(\cdot)'\right]$, and $\mathrm{E}\left[\frac{\partial\psi_{m_z}}{\partial\alpha}\right] = -\mathrm{E}\left[\psi_{m_z}(\cdot)\psi_\alpha^{cb}(\cdot)'\right]$ for $z = 0, 1$.

Table A.1: Parameters and Moment Functions

Parameter	Population Relation	Related Moment Condition
α	$\mathrm{P}(Z=1 \mid X) = F(X,\alpha)$	$\psi_\alpha^{ml} = \frac{Z_i - F(X_i,\alpha)}{F(X_i,\alpha)(1-F(X_i,\alpha))} \nabla_\alpha F(X_i,\alpha)$
		$\psi_\alpha^{cb} = \frac{Z_i - F(X_i,\alpha)}{F(X_i,\alpha)(1-F(X_i,\alpha))} X_i$
Δ	$\Delta = \mathrm{E}\left[Y \frac{Z-p(X)}{p(X)(1-p(X))}\right]$	$\psi_\Delta = \frac{Z_i Y_i}{F(X_i,\alpha)} - \frac{(1-Z_i)Y_i}{1-F(X_i,\alpha)} - \Delta$
Γ	$\Gamma = \mathrm{E}\left[1 - \frac{D(1-Z)}{1-p(X)} - \frac{(1-D)Z}{p(X)}\right]$	$\psi_\Gamma = 1 - \frac{(1-Z_i)D_i}{1-F(X_i,\alpha)} - \frac{Z_i(1-D_i)}{F(X_i,\alpha)} - \Gamma$
Γ_1	$\Gamma_1 = \mathrm{E}\left[D \frac{Z-p(X)}{p(X)(1-p(X))}\right]$	$\psi_{\Gamma_1} = \frac{Z_i D_i}{F(X_i,\alpha)} - \frac{(1-Z_i)D_i}{1-F(X_i,\alpha)} - \Gamma_1$
Γ_0	$\Gamma_0 = \mathrm{E}\left[(1-D)\frac{(1-Z)-(1-p(X))}{p(X)(1-p(X))}\right]$	$\psi_{\Gamma_0} = \frac{Z_i(D_i-1)}{F(X_i,\alpha)} - \frac{(1-Z_i)(D_i-1)}{1-F(X_i,\alpha)} - \Gamma_0$
Δ_1	$\Delta_1 = \mathrm{E}(\kappa_1 Y)$	$\psi_{\Delta_1} = D_i \frac{Z_i - F(X_i,\alpha)}{F(X_i,\alpha)(1-F(X_i,\alpha))} Y_i - \Delta_1$
Δ_0	$\Delta_0 = \mathrm{E}(\kappa_0 Y)$	$\psi_{\Delta_0} = (1-D_i)\frac{(1-Z_i)-(1-F(X_i,\alpha))}{F(X_i,\alpha)(1-F(X_i,\alpha))} Y_i - \Delta_0$
μ_1	$\mu_1 = \mathrm{E}(Y \mid Z=1)$	$\psi_{\mu_1} = \frac{Z_i(Y_i-\mu_1)}{F(X_i,\alpha)}$
μ_0	$\mu_0 = \mathrm{E}(Y \mid Z=0)$	$\psi_{\mu_0} = \frac{(1-Z_i)(Y_i-\mu_0)}{1-F(X_i,\alpha)}$
m_1	$m_1 = \mathrm{E}(D \mid Z=1)$	$\psi_{m_1} = \frac{Z_i(D_i-m_1)}{F(X_i,\alpha)}$
m_0	$m_0 = \mathrm{E}(D \mid Z=0)$	$\psi_{m_0} = \frac{(1-Z_i)(D_i-m_0)}{1-F(X_i,\alpha)}$
τ_{LATE}	$\tau_{\text{LATE}} = \frac{\Delta}{\Gamma} = \frac{\Delta}{\Gamma_1} = \frac{\Delta}{\Gamma_0} = \frac{\Delta_1}{\Gamma_1} - \frac{\Delta_0}{\Gamma_0} = \frac{\mu_1-\mu_0}{m_1-m_0}$	$\psi_{\tau_a} = \frac{\Delta}{\Gamma} - \tau_a$
		$\psi_{\tau_{a,1}} = \frac{\Delta}{\Gamma_1} - \tau_{a,1}$
		$\psi_{\tau_{a,0}} = \frac{\Delta}{\Gamma_0} - \tau_{a,0}$
		$\psi_{\tau_{a,10}} = \frac{\Delta_1}{\Gamma_1} - \frac{\Delta_0}{\Gamma_0} - \tau_{a,10}$
		$\psi_{\tau_u} = \frac{\mu_1-\mu_0}{m_1-m_0} - \tau_u$

Table B.1: Simulation Results for Design A.1

		2SLS	Normalized estimators $\hat{\tau}_u^{cb}$	Normalized estimators $\hat{\tau}_u^{ml}$	Normalized estimators $\hat{\tau}_{a,10}^{ml}$	Unnormalized estimators $\hat{\tau}_a^{ml}$	Unnormalized estimators $\hat{\tau}_t^{ml} = \hat{\tau}_{a,1}^{ml}$	Unnormalized estimators $\hat{\tau}_{a,0}^{ml}$
$\delta = 0.01$								
$N = 500$	MSE	1	2.70	2.63	1093.84	14.16	1304.62	3.12
	|B|	0.0095	0.0215	0.0216	0.1852	0.0365	0.1813	0.0333
	Coverage rate	0.96	0.88	0.92	0.93	0.94	0.94	0.93
$N = 1{,}000$	MSE	1	2.75	2.72	4.11	3.45	4.36	3.07
	|B|	0.0052	0.0090	0.0080	0.0359	0.0096	0.0357	0.0130
	Coverage rate	0.95	0.91	0.93	0.94	0.94	0.95	0.93
$N = 5{,}000$	MSE	1	2.71	2.69	3.00	2.84	3.02	2.98
	|B|	0.0003	0.0023	0.0023	0.0058	0.0018	0.0057	0.0035
	Coverage rate	0.95	0.94	0.95	0.95	0.95	0.95	0.95
$\delta = 0.02$								
$N = 500$	MSE	1	1.93	1.91	20.87	2.94	20.67	2.11
	|B|	0.0097	0.0154	0.0153	0.0492	0.0211	0.0495	0.0215
	Coverage rate	0.96	0.91	0.93	0.94	0.94	0.94	0.93
$N = 1{,}000$	MSE	1	1.89	1.88	2.14	2.00	2.18	2.03
	|B|	0.0027	0.0057	0.0056	0.0148	0.0058	0.0149	0.0082
	Coverage rate	0.95	0.93	0.94	0.95	0.95	0.95	0.94
$N = 5{,}000$	MSE	1	1.86	1.85	2.00	1.90	2.01	1.98
	|B|	0.0026	0.0032	0.0032	0.0048	0.0030	0.0048	0.0037
	Coverage rate	0.95	0.95	0.95	0.95	0.95	0.95	0.95
$\delta = 0.05$								
$N = 500$	MSE	1	1.33	1.32	1.43	1.36	1.46	1.37
	|B|	0.0016	0.0026	0.0024	0.0089	0.0025	0.0088	0.0036
	Coverage rate	0.95	0.94	0.94	0.95	0.94	0.95	0.94
$N = 1{,}000$	MSE	1	1.32	1.31	1.38	1.33	1.39	1.36
	|B|	0.0022	0.0001	0.0001	0.0024	0.0001	0.0024	0.0009
	Coverage rate	0.95	0.94	0.95	0.95	0.95	0.95	0.95
$N = 5{,}000$	MSE	1	1.31	1.31	1.35	1.32	1.35	1.36
	|B|	0.0000	0.0000	0.0000	0.0005	0.0000	0.0005	0.0001
	Coverage rate	0.95	0.95	0.95	0.95	0.95	0.95	0.95

Notes: The details of this simulation design are provided in Section 5 in the main text. "MSE" is the mean squared error of an estimator, normalized by the mean squared error of 2SLS. "|B|" is the absolute bias. "Coverage rate" is the coverage rate for a nominal 95% confidence interval. "2SLS" is the 2SLS estimator that additively controls for X. The weighting estimators are defined in Section 3 in the main text. All weighting estimators also control for X. Results are based on 10,000 replications.

Table B.2: Simulation Results for Design A.2

		2SLS	Normalized estimators $\hat{\tau}_u^{cb}$	$\hat{\tau}_u^{ml}$	$\hat{\tau}_{a,10}^{ml}$	Unnormalized estimators $\hat{\tau}_a^{ml}$	$\hat{\tau}_t^{ml} = \hat{\tau}_{a,1}^{ml}$	$\hat{\tau}_{a,0}^{ml}$
$\delta = 0.01$								
$N = 500$	MSE	1	2.75	2.78	2.30e+04	6.83	3.09	2.52e+04
	|B|	0.0023	0.0033	0.0028	0.4066	0.0046	0.0025	0.4334
	Coverage rate	0.96	0.88	0.93	0.93	0.96	0.93	0.94
$N = 1,000$	MSE	1	2.63	2.60	3.03	2.92	2.72	3.26
	|B|	0.0017	0.0013	0.0010	0.0008	0.0006	0.0011	0.0008
	Coverage rate	0.95	0.91	0.94	0.94	0.96	0.94	0.95
$N = 5,000$	MSE	1	2.72	2.71	2.76	2.76	2.73	2.79
	|B|	0.0008	0.0018	0.0018	0.0018	0.0017	0.0017	0.0017
	Coverage rate	0.95	0.94	0.95	0.95	0.95	0.95	0.95
$\delta = 0.02$								
$N = 500$	MSE	1	1.93	1.91	2.31	2.16	2.00	2.44
	|B|	0.0029	0.0027	0.0025	0.0026	0.0034	0.0028	0.0031
	Coverage rate	0.95	0.91	0.93	0.94	0.95	0.94	0.95
$N = 1,000$	MSE	1	1.86	1.84	1.92	1.90	1.88	1.96
	|B|	0.0019	0.0028	0.0032	0.0035	0.0034	0.0034	0.0035
	Coverage rate	0.95	0.93	0.94	0.95	0.95	0.95	0.95
$N = 5,000$	MSE	1	1.91	1.90	1.92	1.91	1.91	1.93
	|B|	0.0006	0.0007	0.0008	0.0008	0.0008	0.0008	0.0008
	Coverage rate	0.95	0.94	0.95	0.95	0.95	0.95	0.95
$\delta = 0.05$								
$N = 500$	MSE	1	1.32	1.31	1.36	1.34	1.32	1.39
	|B|	0.0008	0.0012	0.0013	0.0018	0.0016	0.0015	0.0017
	Coverage rate	0.95	0.94	0.94	0.94	0.94	0.94	0.95
$N = 1,000$	MSE	1	1.30	1.30	1.31	1.31	1.31	1.32
	|B|	0.0003	0.0008	0.0008	0.0007	0.0007	0.0010	0.0005
	Coverage rate	0.95	0.95	0.95	0.95	0.95	0.95	0.95
$N = 5,000$	MSE	1	1.30	1.30	1.30	1.30	1.30	1.30
	|B|	0.0005	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008
	Coverage rate	0.95	0.95	0.95	0.95	0.95	0.95	0.95

Notes: The details of this simulation design are provided in Section 5 in the main text. "MSE" is the mean squared error of an estimator, normalized by the mean squared error of 2SLS. "|B|" is the absolute bias. "Coverage rate" is the coverage rate for a nominal 95% confidence interval. "2SLS" is the 2SLS estimator that additively controls for X. The weighting estimators are defined in Section 3 in the main text. All weighting estimators also control for X. Results are based on 10,000 replications.

Table B.3: Simulation Results for Design B

		2SLS	Normalized estimators $\hat{\tau}_u^{cb}$	$\hat{\tau}_u^{ml}$	$\hat{\tau}_{a,10}^{ml}$	Unnormalized estimators $\hat{\tau}_a^{ml}$	$\hat{\tau}_t^{ml} = \hat{\tau}_{a,1}^{ml}$	$\hat{\tau}_{a,0}^{ml}$
$\delta = 0.01$								
$N = 500$	MSE	1	2.57	2.74	189.22	210.94	761.97	4.02
	|B|	0.0614	0.0140	0.0103	0.0490	0.0927	0.0059	0.0197
	Coverage rate	0.96	0.88	0.94	0.95	0.95	0.94	0.94
$N = 1{,}000$	MSE	1	2.50	2.51	6.59	3.20	7.00	2.82
	|B|	0.0551	0.0035	0.0024	0.0323	0.0094	0.0340	0.0065
	Coverage rate	0.95	0.91	0.94	0.95	0.95	0.95	0.94
$N = 5{,}000$	MSE	1	1.96	1.95	2.19	2.06	2.20	2.10
	|B|	0.0531	0.0009	0.0006	0.0046	0.0009	0.0045	0.0014
	Coverage rate	0.92	0.94	0.95	0.95	0.95	0.95	0.95
$\delta = 0.02$								
$N = 500$	MSE	1	1.92	1.93	11.76	2.61	16.46	2.09
	|B|	0.0498	0.0129	0.0117	0.0534	0.0186	0.0568	0.0142
	Coverage rate	0.95	0.91	0.93	0.95	0.95	0.95	0.94
$N = 1{,}000$	MSE	1	1.81	1.80	2.20	1.96	2.23	1.92
	|B|	0.0473	0.0063	0.0058	0.0182	0.0075	0.0180	0.0069
	Coverage rate	0.95	0.93	0.95	0.95	0.96	0.96	0.95
$N = 5{,}000$	MSE	1	1.46	1.45	1.58	1.50	1.58	1.53
	|B|	0.0436	0.0003	0.0003	0.0021	0.0004	0.0021	0.0006
	Coverage rate	0.93	0.95	0.95	0.95	0.95	0.95	0.95
$\delta = 0.05$								
$N = 500$	MSE	1	1.30	1.30	5.79	1.35	5.22	1.34
	|B|	0.0334	0.0018	0.0014	0.0141	0.0022	0.0137	0.0016
	Coverage rate	0.96	0.94	0.95	0.95	0.95	0.96	0.95
$N = 1{,}000$	MSE	1	1.29	1.29	1.36	1.31	1.37	1.33
	|B|	0.0335	0.0042	0.0040	0.0073	0.0041	0.0073	0.0041
	Coverage rate	0.95	0.94	0.95	0.95	0.95	0.95	0.94
$N = 5{,}000$	MSE	1	1.12	1.12	1.16	1.13	1.16	1.15
	|B|	0.0309	0.0008	0.0007	0.0012	0.0007	0.0013	0.0008
	Coverage rate	0.94	0.95	0.95	0.95	0.95	0.95	0.95

Notes: The details of this simulation design are provided in Section 5 in the main text. "MSE" is the mean squared error of an estimator, normalized by the mean squared error of 2SLS. "|B|" is the absolute bias. "Coverage rate" is the coverage rate for a nominal 95% confidence interval. "2SLS" is the 2SLS estimator that additively controls for X. The weighting estimators are defined in Section 3 in the main text. All weighting estimators also control for X. Results are based on 10,000 replications.

Table B.4: Simulation Results for Design C

		2SLS	Normalized estimators $\hat{\tau}_u^{cb}$	Normalized estimators $\hat{\tau}_u^{ml}$	Normalized estimators $\hat{\tau}_{a,10}^{ml}$	Unnormalized estimators $\hat{\tau}_a^{ml}$	Unnormalized estimators $\hat{\tau}_t^{ml} = \hat{\tau}_{a,1}^{ml}$	Unnormalized estimators $\hat{\tau}_{a,0}^{ml}$
$\delta = 0.01$								
$N = 500$	MSE	1	0.75	3.82	4.95e+04	2010.01	4.92e+04	219.69
	|B|	4.6994	0.1184	0.7953	7.2631	2.5598	7.2230	2.4048
	Coverage rate	0.33	0.78	0.82	0.83	0.96	0.83	0.93
$N = 1{,}000$	MSE	1	0.42	1.47	95.93	23.83	96.38	38.68
	|B|	4.7053	0.0938	0.3867	0.8364	1.4320	0.8401	1.1898
	Coverage rate	0.07	0.84	0.87	0.88	0.97	0.88	0.94
$N = 5{,}000$	MSE	1	0.09	0.30	0.34	2.24	0.34	7.35
	|B|	4.6729	0.0415	0.0568	0.0848	0.2707	0.0849	0.2319
	Coverage rate	0.00	0.92	0.94	0.94	0.96	0.94	0.95
$\delta = 0.02$								
$N = 500$	MSE	1	0.64	1.82	20.02	52.38	20.36	53.85
	|B|	3.9155	0.0580	0.4457	0.4927	1.8422	0.4896	1.5703
	Coverage rate	0.44	0.84	0.87	0.89	0.97	0.89	0.94
$N = 1{,}000$	MSE	1	0.36	0.97	1.29	7.64	1.29	24.19
	|B|	3.8732	0.0521	0.1726	0.2334	0.7182	0.2335	0.5280
	Coverage rate	0.15	0.89	0.91	0.92	0.96	0.92	0.95
$N = 5{,}000$	MSE	1	0.08	0.20	0.23	1.52	0.23	5.09
	|B|	3.8464	0.0124	0.0109	0.0589	0.1196	0.0589	0.0763
	Coverage rate	0.00	0.93	0.94	0.95	0.95	0.95	0.95
$\delta = 0.05$								
$N = 500$	MSE	1	0.62	1.13	1.44	7.88	1.44	24.77
	|B|	2.6174	0.0767	0.1027	0.1660	0.5604	0.1661	0.2451
	Coverage rate	0.66	0.91	0.93	0.94	0.97	0.94	0.95
$N = 1{,}000$	MSE	1	0.37	0.65	0.74	4.29	0.74	13.98
	|B|	2.6376	0.0319	0.0268	0.0894	0.2009	0.0894	0.1782
	Coverage rate	0.40	0.93	0.94	0.95	0.95	0.95	0.95
$N = 5{,}000$	MSE	1	0.09	0.15	0.16	0.93	0.16	3.10
	|B|	2.6232	0.0029	0.0161	0.0035	0.0294	0.0035	0.0586
	Coverage rate	0.00	0.95	0.95	0.95	0.95	0.95	0.95

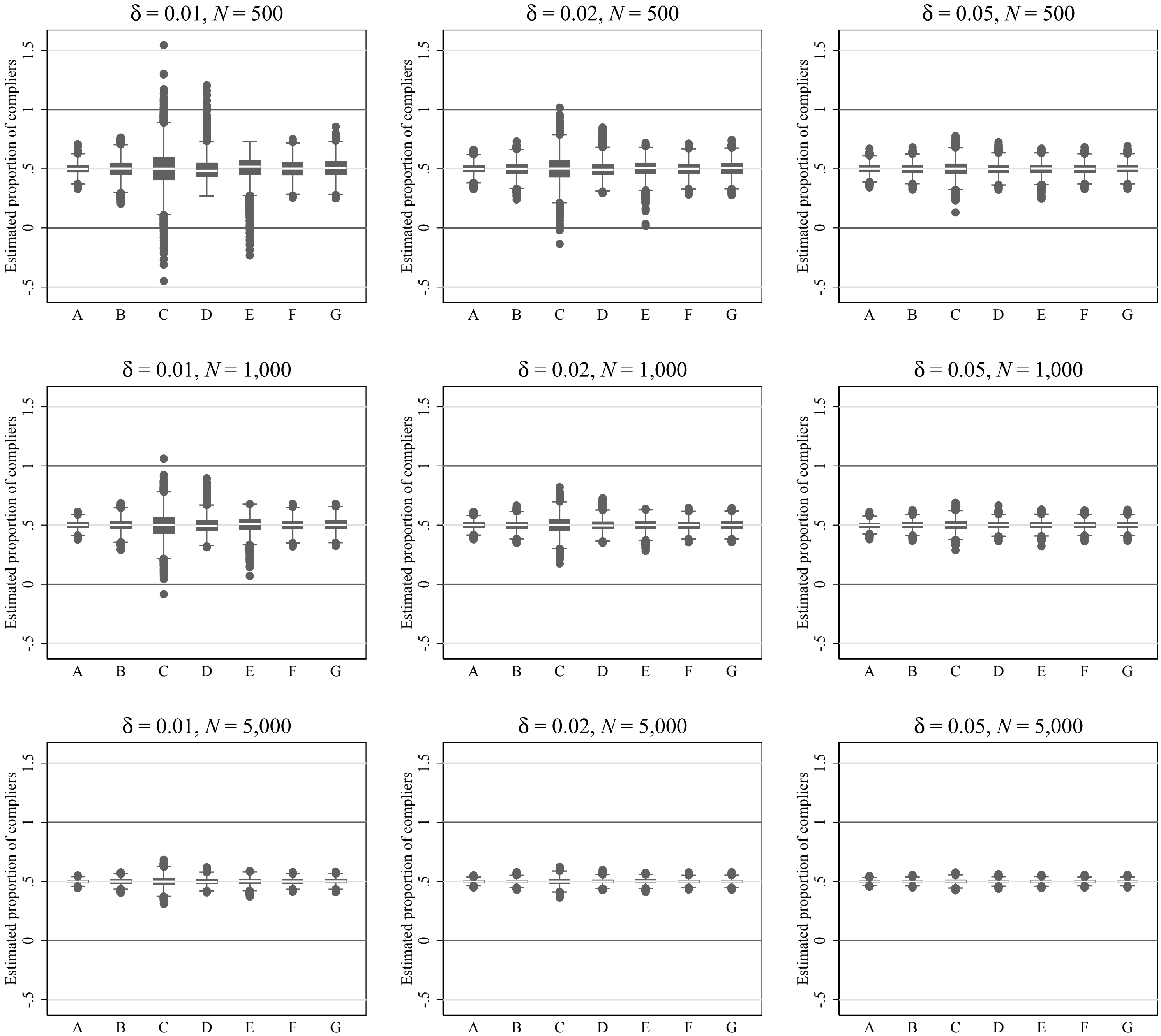
Notes: The details of this simulation design are provided in Section 5 in the main text. "MSE" is the mean squared error of an estimator, normalized by the mean squared error of 2SLS. "|B|" is the absolute bias. "Coverage rate" is the coverage rate for a nominal 95% confidence interval. "2SLS" is the 2SLS estimator that additively controls for X. The weighting estimators are defined in Section 3 in the main text. All weighting estimators also control for X. Results are based on 10,000 replications.

Table B.5: Simulation Results for Design D

		2SLS	Normalized estimators $\hat{\tau}_u^{cb}$	Normalized estimators $\hat{\tau}_u^{ml}$	Normalized estimators $\hat{\tau}_{a,10}^{ml}$	Unnormalized estimators $\hat{\tau}_a^{ml}$	Unnormalized estimators $\hat{\tau}_t^{ml} = \hat{\tau}_{a,1}^{ml}$	Unnormalized estimators $\hat{\tau}_{a,0}^{ml}$
$\delta = 0.01$								
$N = 500$	MSE	1	0.08	7.06	0.56	2.69e+05	0.32	1.75e+04
	|B|	17.6766	0.6047	4.2535	0.6326	102.1028	0.7343	82.6894
	Coverage rate	0.00	0.85	0.77	0.75	0.93	0.74	0.91
$N = 1{,}000$	MSE	1	0.04	3.98	2.64	1.44e+04	0.12	1.91e+05
	|B|	17.5275	0.4052	6.1212	1.9580	46.4242	2.4467	46.6583
	Coverage rate	0.00	0.88	0.80	0.79	0.86	0.79	0.82
$N = 5{,}000$	MSE	1	0.01	0.26	0.07	11.68	0.07	23.12
	|B|	17.4073	0.3154	7.9930	3.7953	55.3392	3.7955	78.2082
	Coverage rate	0.00	0.93	0.42	0.58	0.13	0.58	0.09
$\delta = 0.02$								
$N = 500$	MSE	1	0.06	0.40	0.21	7978.30	0.16	1.12e+04
	|B|	14.1078	0.3874	4.0705	1.3717	17.2726	1.3658	40.6495
	Coverage rate	0.00	0.89	0.84	0.84	0.89	0.83	0.86
$N = 1{,}000$	MSE	1	0.03	0.27	0.09	10.24	0.09	25.76
	|B|	13.9940	0.3326	4.7909	2.0492	35.2328	2.0474	51.9926
	Coverage rate	0.00	0.91	0.83	0.84	0.75	0.84	0.70
$N = 5{,}000$	MSE	1	0.01	0.18	0.05	6.64	0.05	13.56
	|B|	13.9115	0.2707	5.3737	2.5524	34.3929	2.5523	49.3305
	Coverage rate	0.00	0.95	0.36	0.61	0.02	0.61	0.01
$\delta = 0.05$								
$N = 500$	MSE	1	0.06	0.24	0.12	5.29	0.12	11.84
	|B|	9.1248	0.2697	2.2155	0.8326	16.2049	0.8327	24.8322
	Coverage rate	0.01	0.93	0.90	0.91	0.82	0.91	0.80
$N = 1{,}000$	MSE	1	0.03	0.15	0.06	4.01	0.06	8.93
	|B|	9.0882	0.2770	2.3381	0.9487	15.9970	0.9487	24.1235
	Coverage rate	0.00	0.94	0.87	0.91	0.57	0.91	0.54
$N = 5{,}000$	MSE	1	0.01	0.09	0.02	3.28	0.02	7.27
	|B|	9.0474	0.2702	2.4706	1.0592	15.9694	1.0591	23.7925
	Coverage rate	0.00	0.95	0.46	0.79	0.01	0.79	0.00

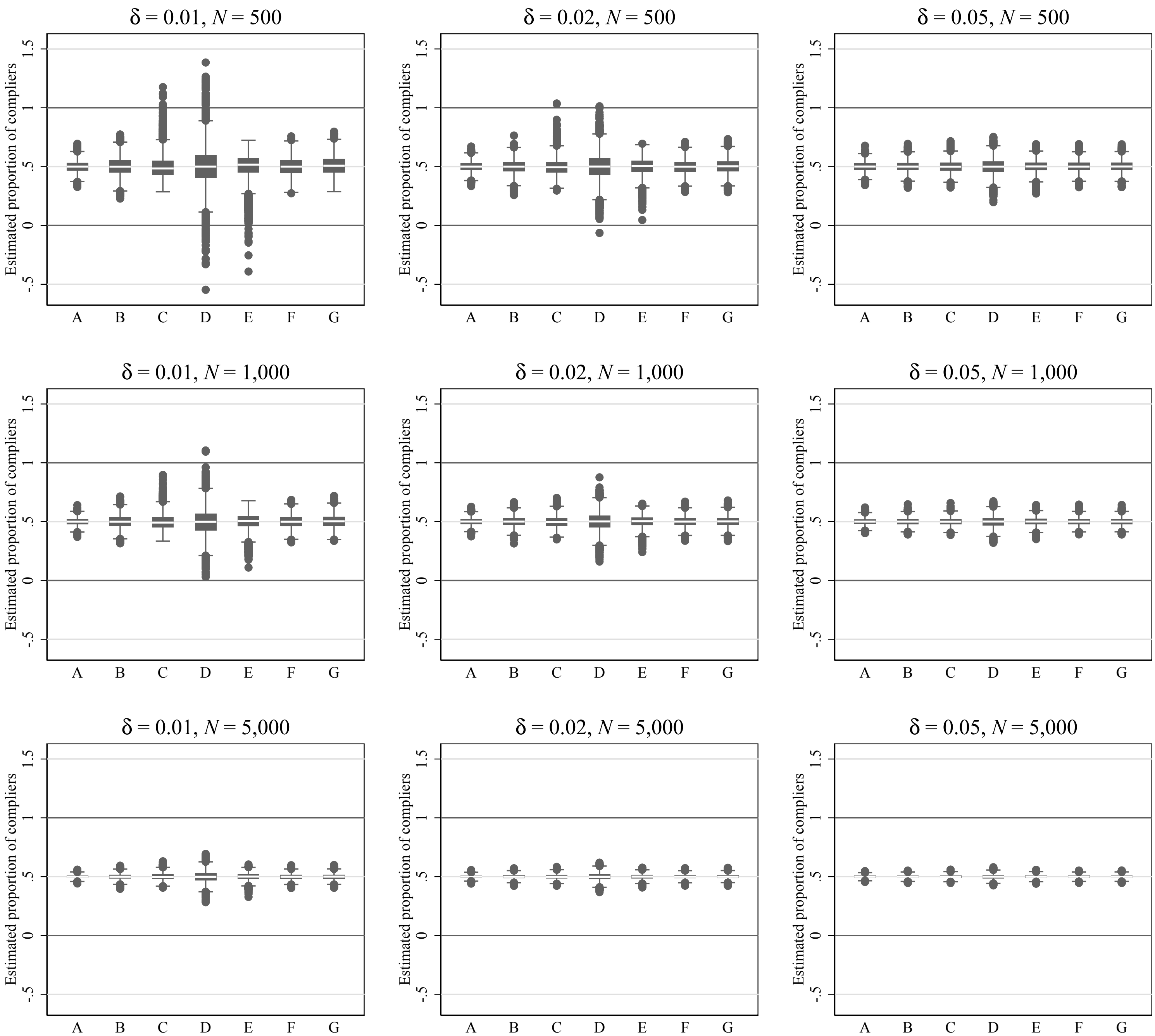
Notes: The details of this simulation design are provided in Section 5 in the main text. "MSE" is the mean squared error of an estimator, normalized by the mean squared error of 2SLS. "|B|" is the absolute bias. "Coverage rate" is the coverage rate for a nominal 95% confidence interval. "2SLS" is the 2SLS estimator that additively controls for X. The weighting estimators are defined in Section 3 in the main text. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure B.1: Simulation Results for the Proportion of Compliers in Design A.1



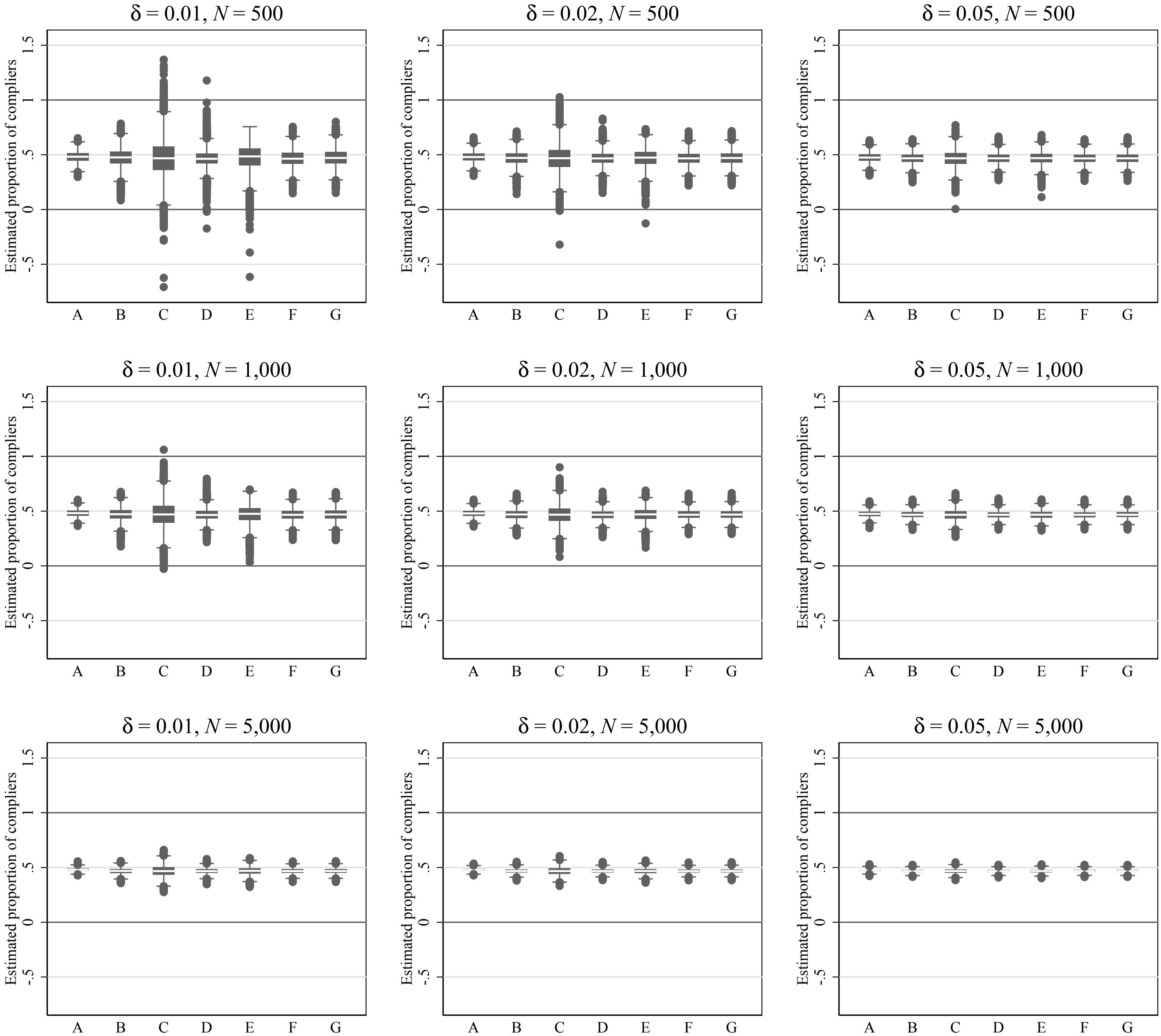
Notes: The details of this simulation design are provided in Section 5 in the main text. "A" corresponds to the first-stage coefficient on Z in 2SLS, controlling additively for X. "B" corresponds to the denominator of $\hat{\tau}_u^{ml}$. "C," "D," and "E" correspond to $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1}$, $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, and $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i}$, respectively. These estimators, as well as the denominator of $\hat{\tau}_u^{ml}$, are based on an instrument propensity score, which is estimated using logit ML, also controlling for X. "F" corresponds to the denominator of $\hat{\tau}_u^{cb}$. "G" corresponds to $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1} = N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, where the instrument propensity score is estimated using the logit model and the moment conditions in equation (9) in the main text, also controlling for X, as in the case of the denominator of $\hat{\tau}_u^{cb}$. Results are based on 10,000 replications.

Figure B.2: Simulation Results for the Proportion of Compliers in Design A.2



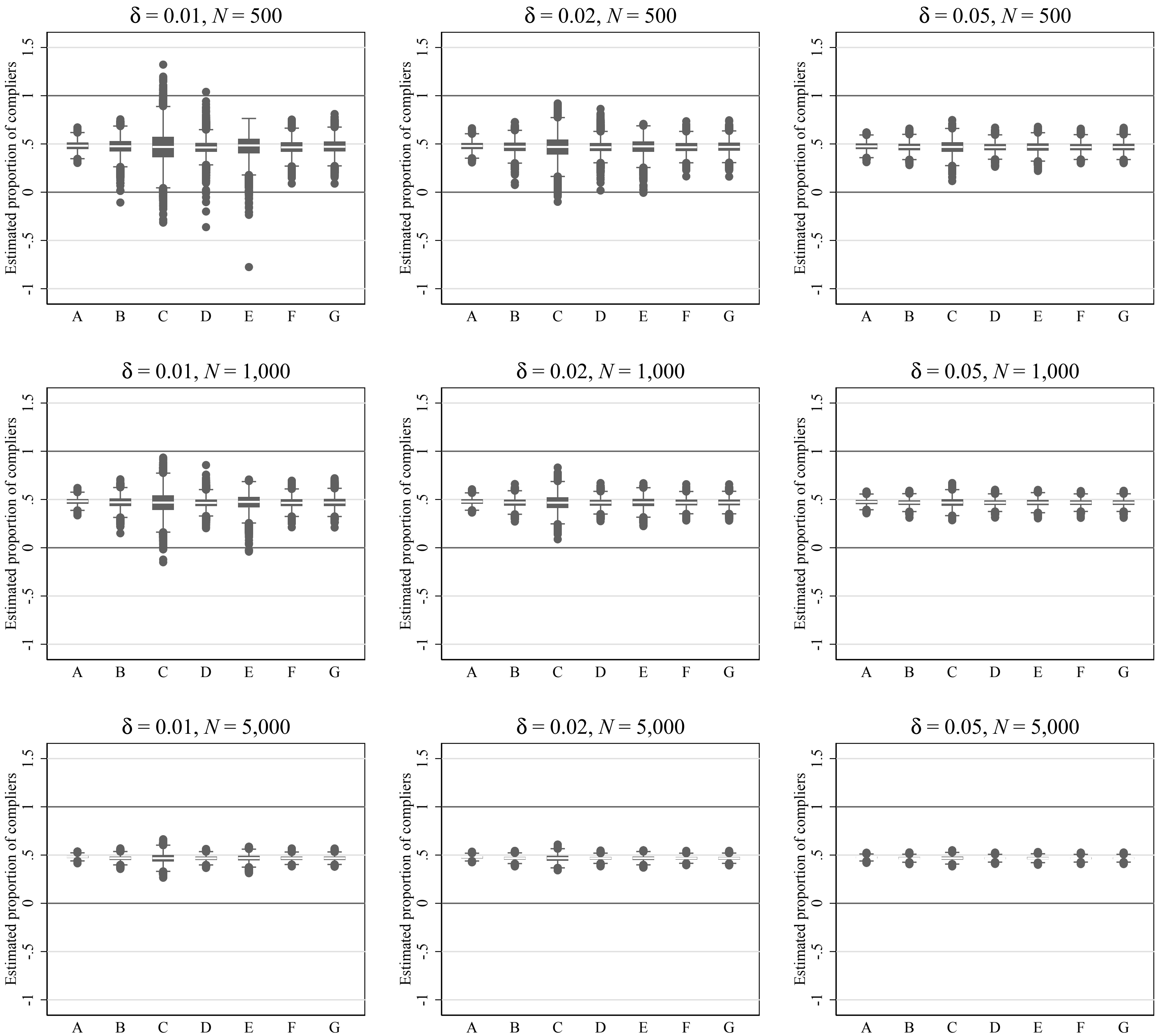
Notes: The details of this simulation design are provided in Section 5 in the main text. "A" corresponds to the first-stage coefficient on Z in 2SLS, controlling additively for X. "B" corresponds to the denominator of $\hat{\tau}_u^{ml}$. "C," "D," and "E" correspond to $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1}$, $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, and $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i}$, respectively. These estimators, as well as the denominator of $\hat{\tau}_u^{ml}$, are based on an instrument propensity score, which is estimated using logit ML, also controlling for X. "F" corresponds to the denominator of $\hat{\tau}_u^{cb}$. "G" corresponds to $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1} = N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, where the instrument propensity score is estimated using the logit model and the moment conditions in equation (9) in the main text, also controlling for X, as in the case of the denominator of $\hat{\tau}_u^{cb}$. Results are based on 10,000 replications.

Figure B.3: Simulation Results for the Proportion of Compliers in Design B



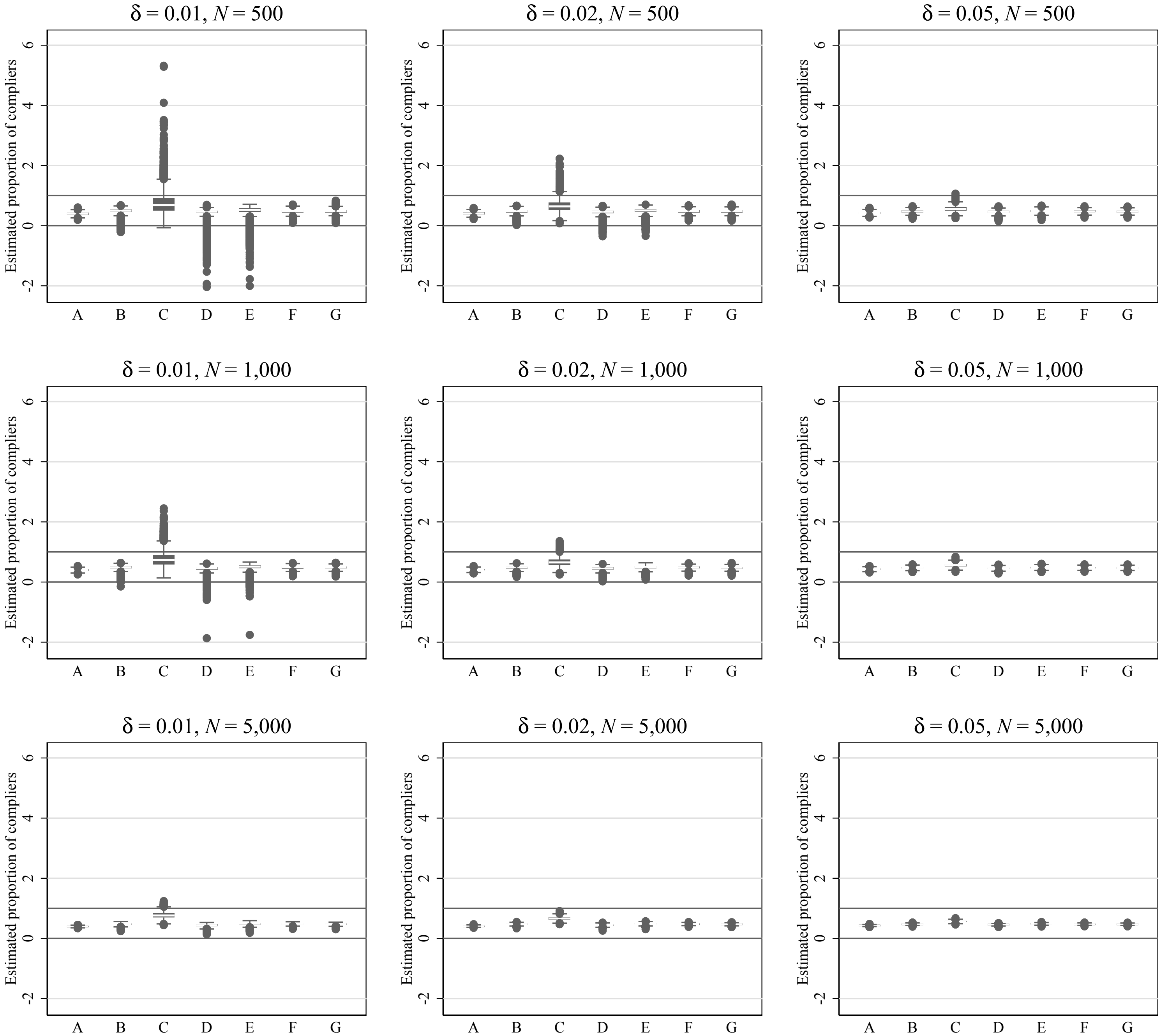
Notes: The details of this simulation design are provided in Section 5 in the main text. "A" corresponds to the first-stage coefficient on Z in 2SLS, controlling additively for X. "B" corresponds to the denominator of $\hat{\tau}_u^{ml}$. "C," "D," and "E" correspond to $N^{-1} \sum_{i=1}^{N} \hat{\kappa}_{i1}$, $N^{-1} \sum_{i=1}^{N} \hat{\kappa}_{i0}$, and $N^{-1} \sum_{i=1}^{N} \hat{\kappa}_i$, respectively. These estimators, as well as the denominator of $\hat{\tau}_u^{ml}$, are based on an instrument propensity score, which is estimated using logit ML, also controlling for X. "F" corresponds to the denominator of $\hat{\tau}_u^{cb}$. "G" corresponds to $N^{-1} \sum_{i=1}^{N} \hat{\kappa}_{i1} = N^{-1} \sum_{i=1}^{N} \hat{\kappa}_{i0}$, where the instrument propensity score is estimated using the logit model and the moment conditions in equation (9) in the main text, also controlling for X, as in the case of the denominator of $\hat{\tau}_u^{cb}$. Results are based on 10,000 replications.

Figure B.4: Simulation Results for the Proportion of Compliers in Design C



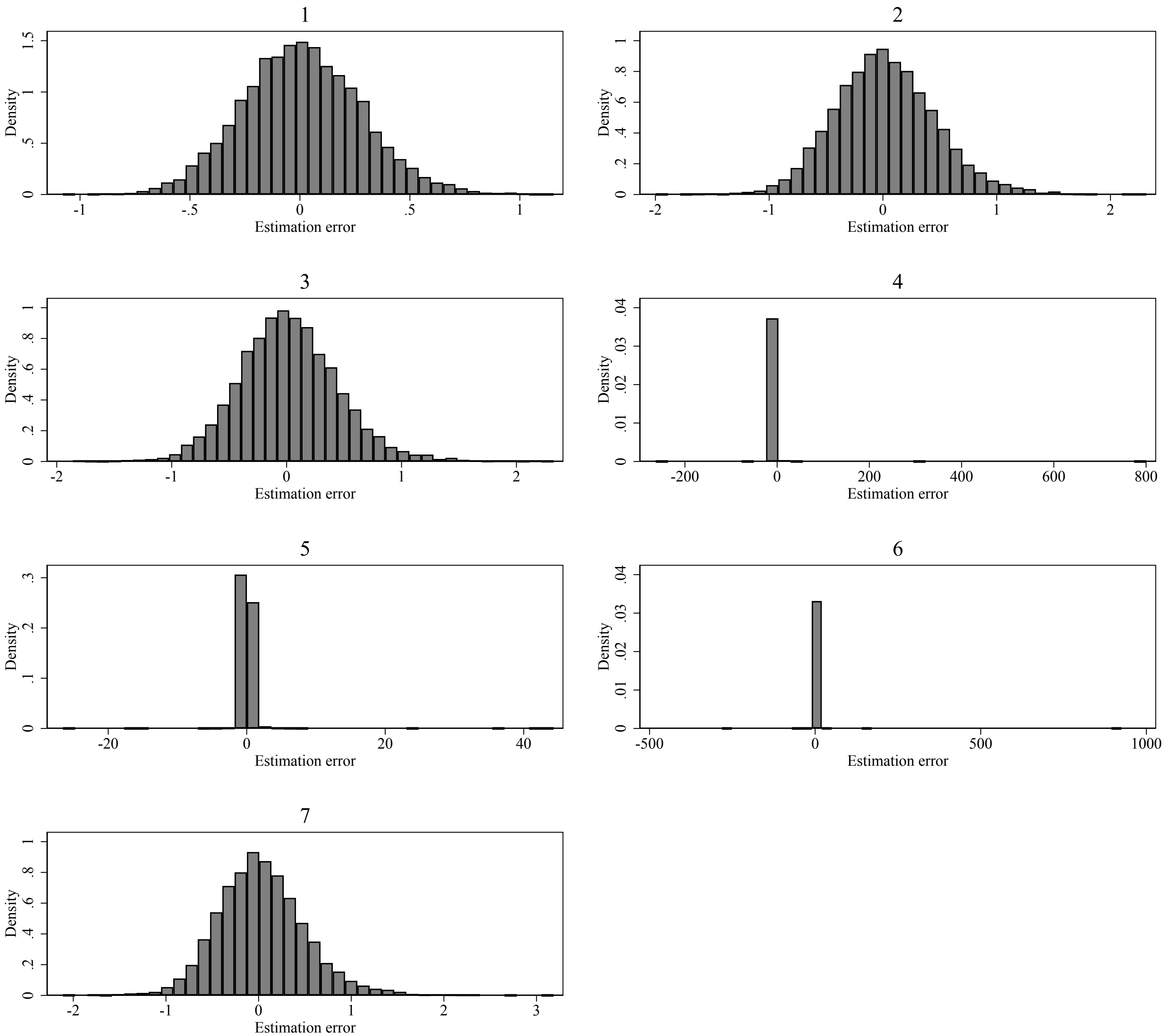
Notes: The details of this simulation design are provided in Section 5 in the main text. "A" corresponds to the first-stage coefficient on Z in 2SLS, controlling additively for X. "B" corresponds to the denominator of $\hat{\tau}_u^{ml}$. "C," "D," and "E" correspond to $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1}$, $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, and $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i}$, respectively. These estimators, as well as the denominator of $\hat{\tau}_u^{ml}$, are based on an instrument propensity score, which is estimated using logit ML, also controlling for X. "F" corresponds to the denominator of $\hat{\tau}_u^{cb}$. "G" corresponds to $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1} = N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, where the instrument propensity score is estimated using the logit model and the moment conditions in equation (9) in the main text, also controlling for X, as in the case of the denominator of $\hat{\tau}_u^{cb}$. Results are based on 10,000 replications.

Figure B.5: Simulation Results for the Proportion of Compliers in Design D



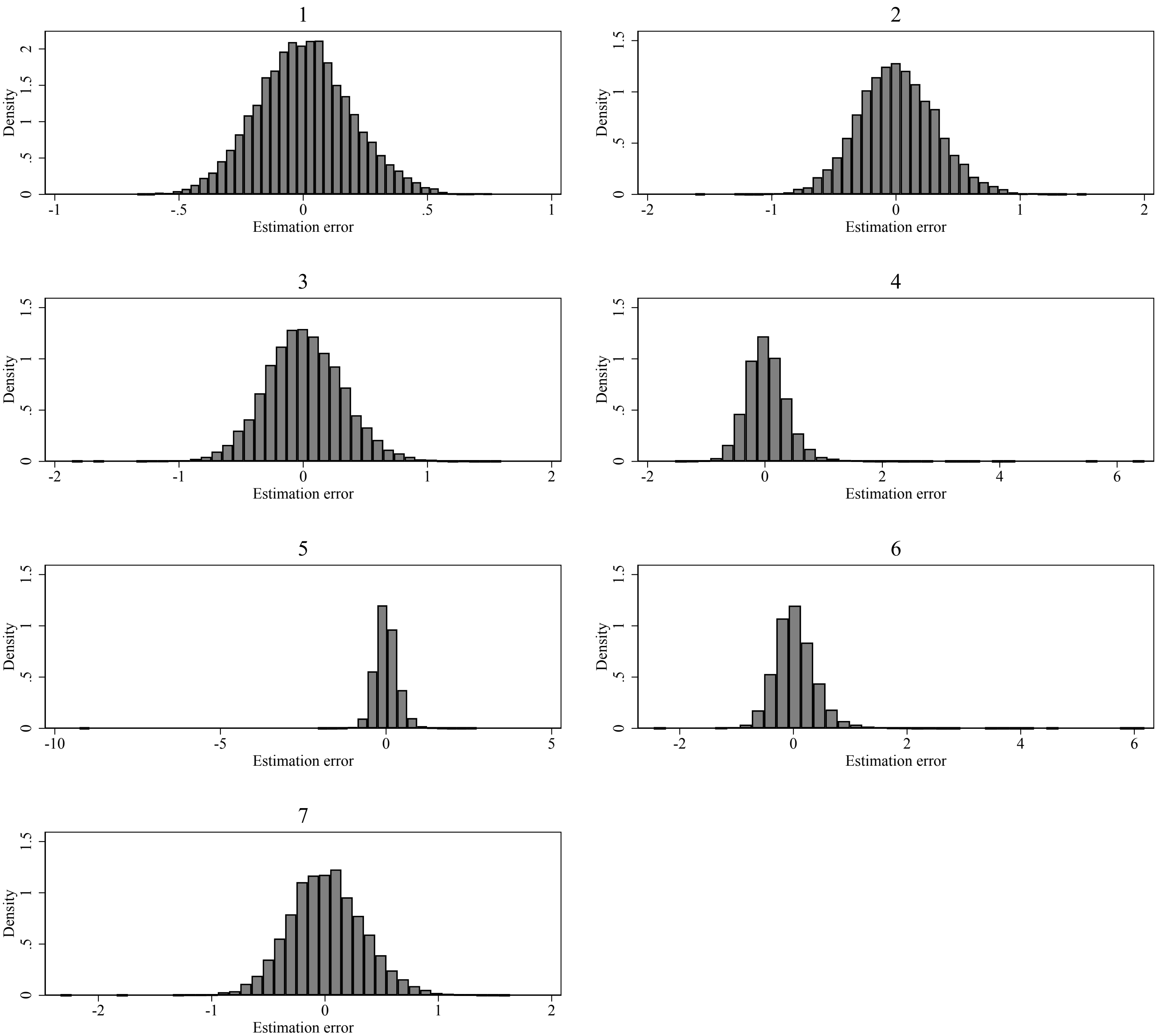
Notes: The details of this simulation design are provided in Section 5 in the main text. “A” corresponds to the first-stage coefficient on Z in 2SLS, controlling additively for X. “B” corresponds to the denominator of $\hat{\tau}_u^{ml}$. “C,” “D,” and “E” correspond to $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1}$, $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, and $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i}$, respectively. These estimators, as well as the denominator of $\hat{\tau}_u^{ml}$, are based on an instrument propensity score, which is estimated using logit ML, also controlling for X. “F” corresponds to the denominator of $\hat{\tau}_u^{cb}$. “G” corresponds to $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1} = N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, where the instrument propensity score is estimated using the logit model and the moment conditions in equation (9) in the main text, also controlling for X, as in the case of the denominator of $\hat{\tau}_u^{cb}$. Results are based on 10,000 replications.

Figure C.1: Simulation Results for Design A.1, $\delta = 0.01$, $N = 500$



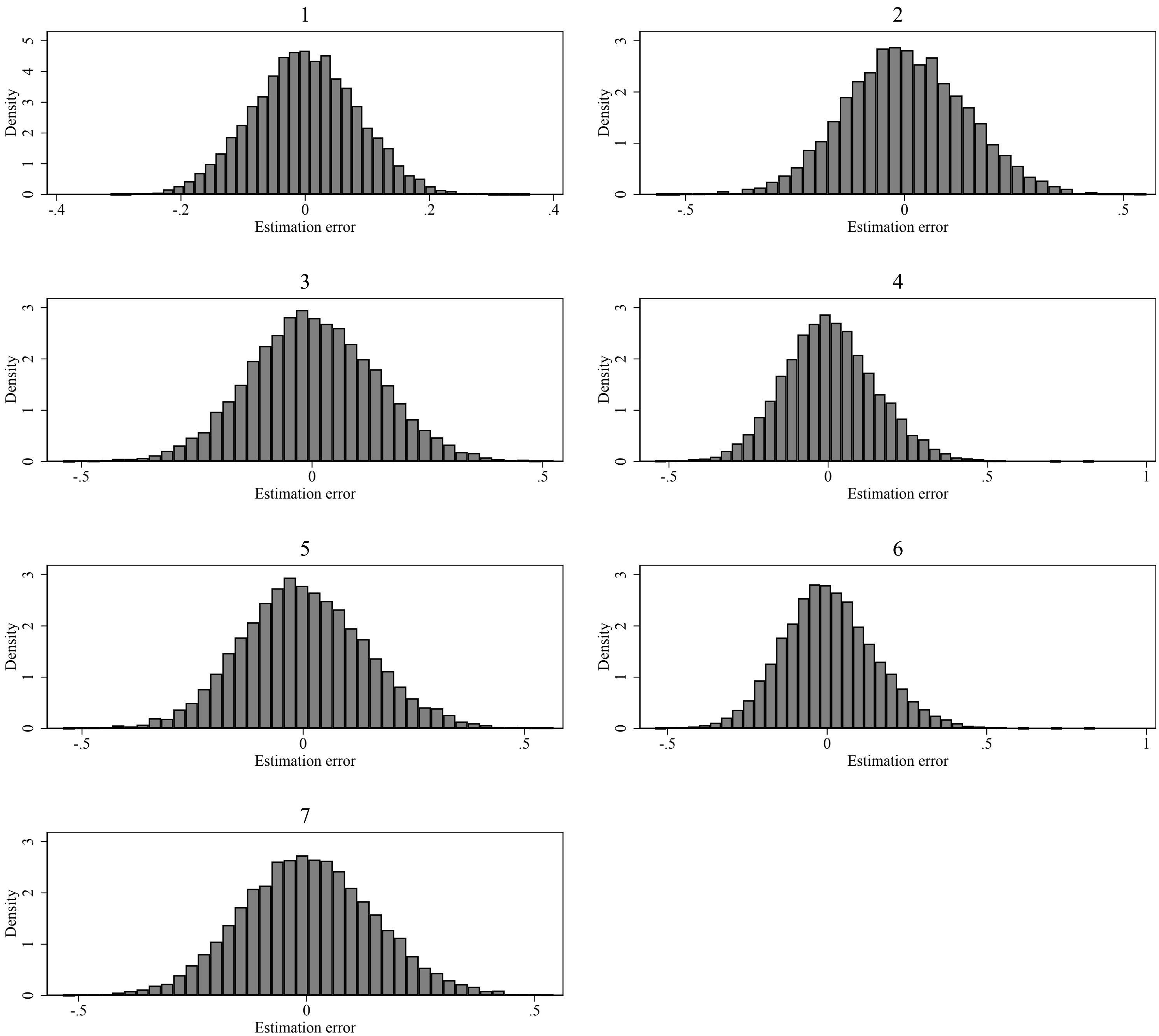
Notes: The details of this simulation design are provided in Section 5 in the main text. “1” corresponds to the 2SLS estimator that additively controls for X. “2” corresponds to $\hat{\tau}_u^{cb}$. “3” corresponds to $\hat{\tau}_u^{ml}$. “4” corresponds to $\hat{\tau}_{a,10}^{ml}$. “5” corresponds to $\hat{\tau}_a^{ml}$. “6” corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). “7” corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.2: Simulation Results for Design A.1, $\delta = 0.01$, $N = 1{,}000$



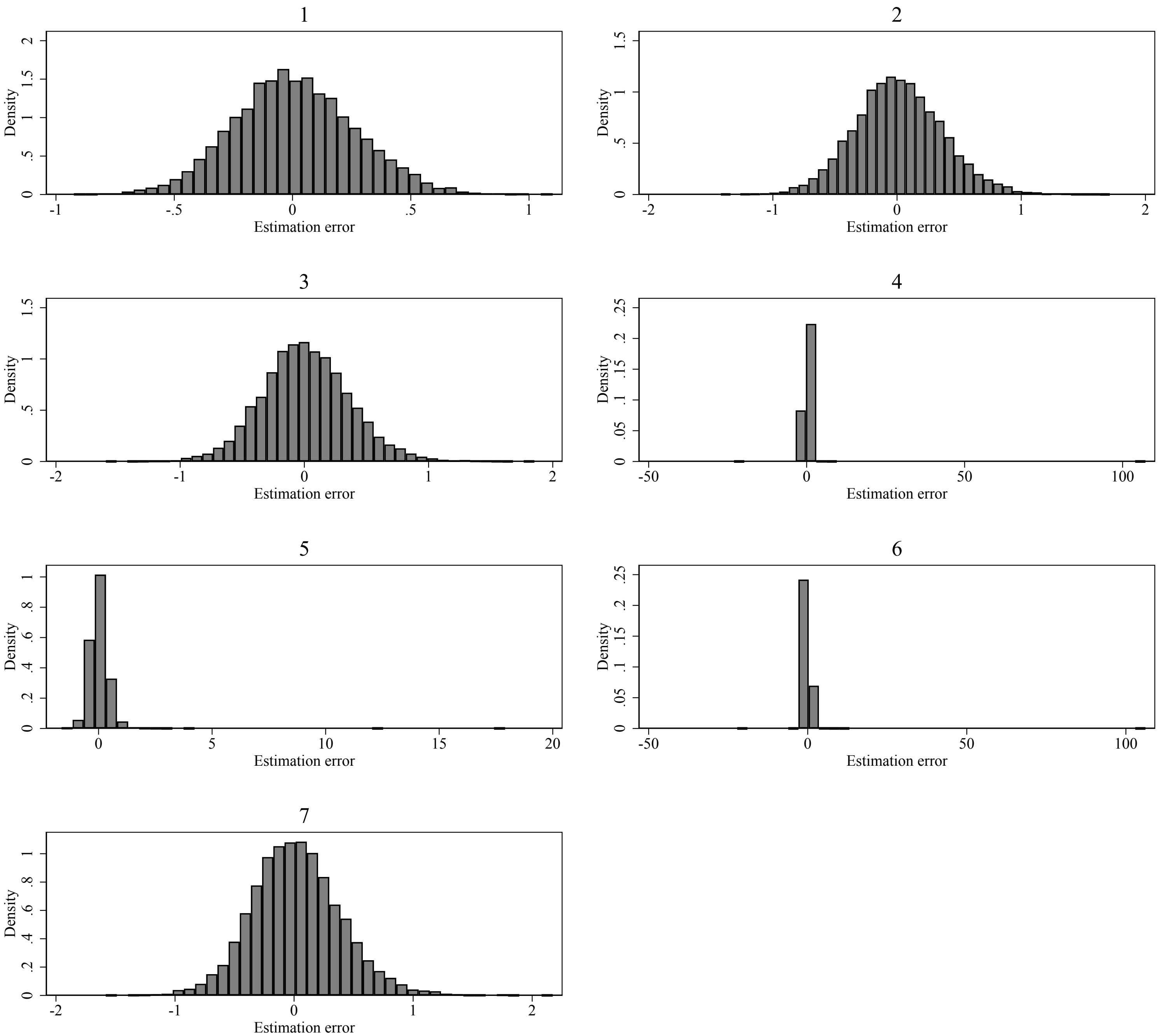
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.3: Simulation Results for Design A.1, $\delta = 0.01$, $N = 5{,}000$



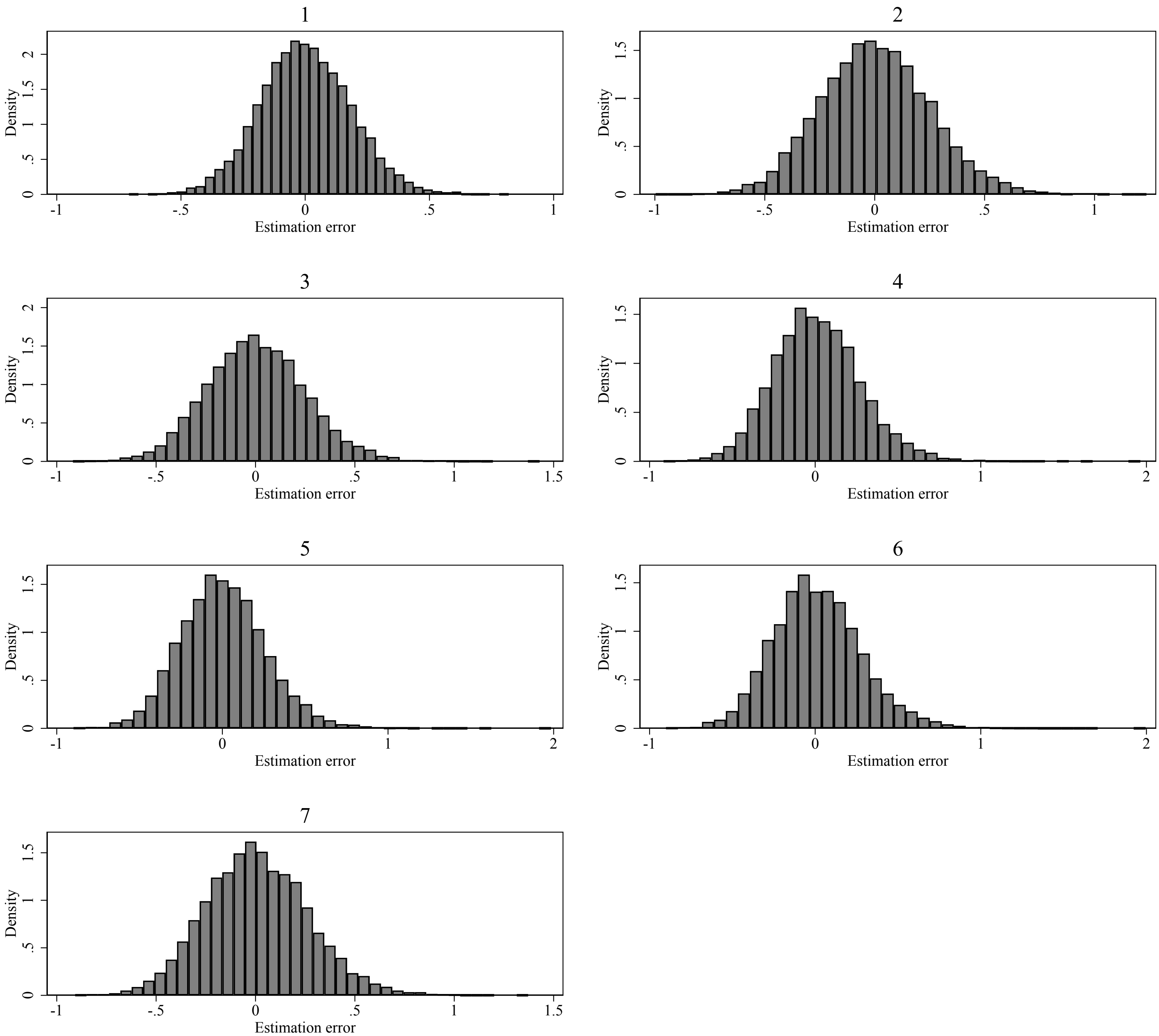
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.4: Simulation Results for Design A.1, δ = 0.02, N = 500



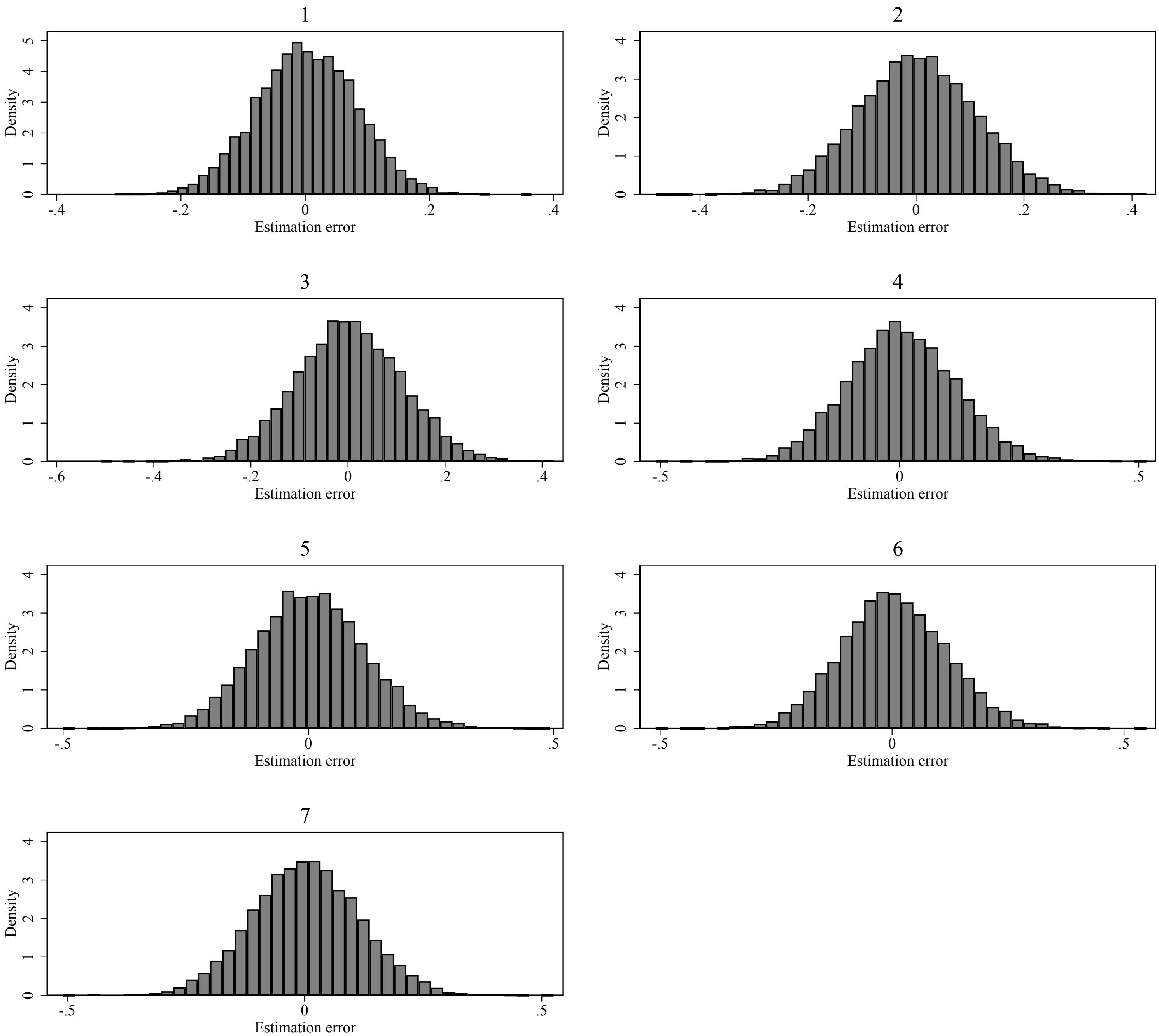
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ (= $\hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.5: Simulation Results for Design A.1, $\delta = 0.02$, $N = 1{,}000$



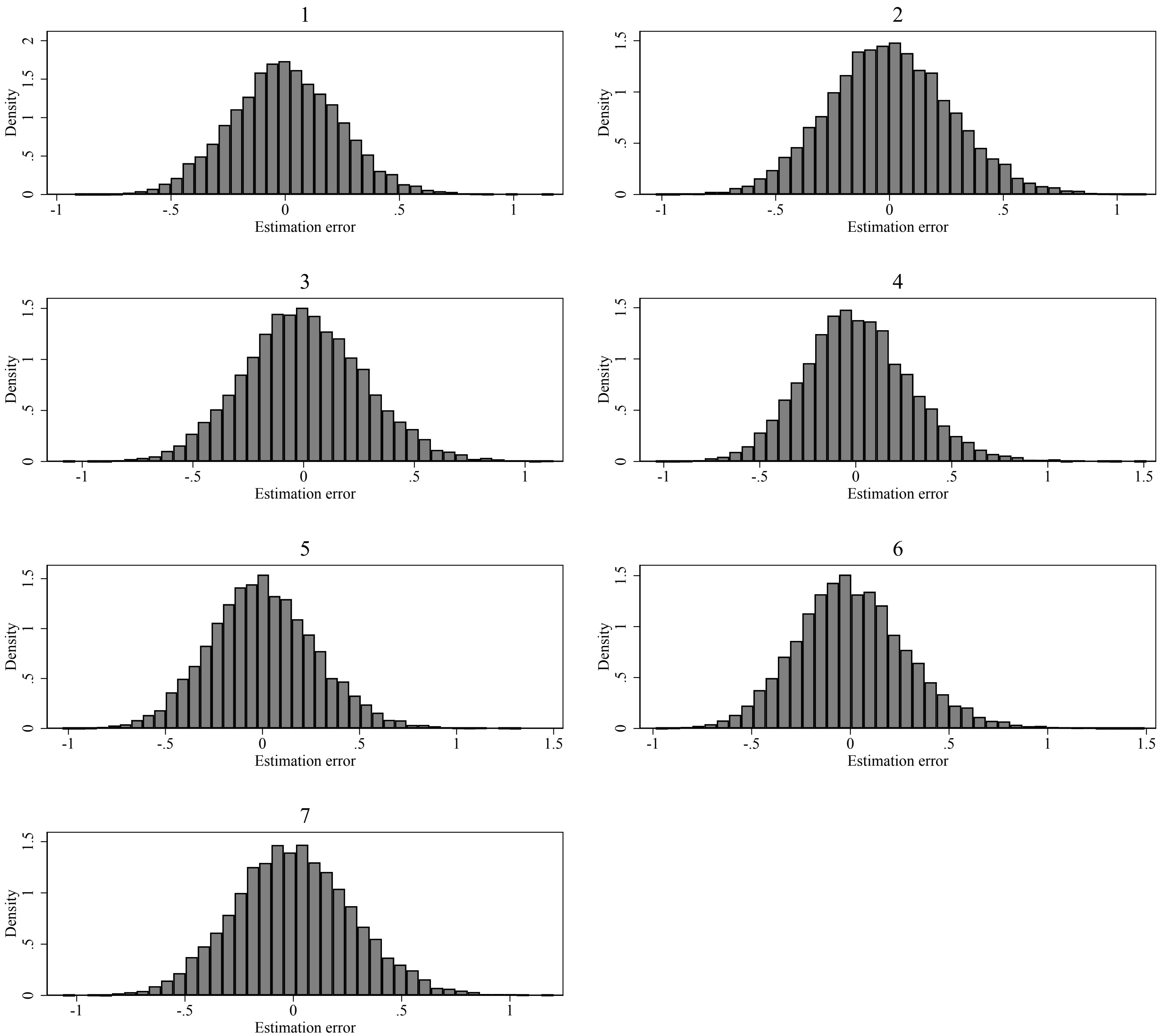
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.6: Simulation Results for Design A.1, $\delta = 0.02$, $N = 5{,}000$



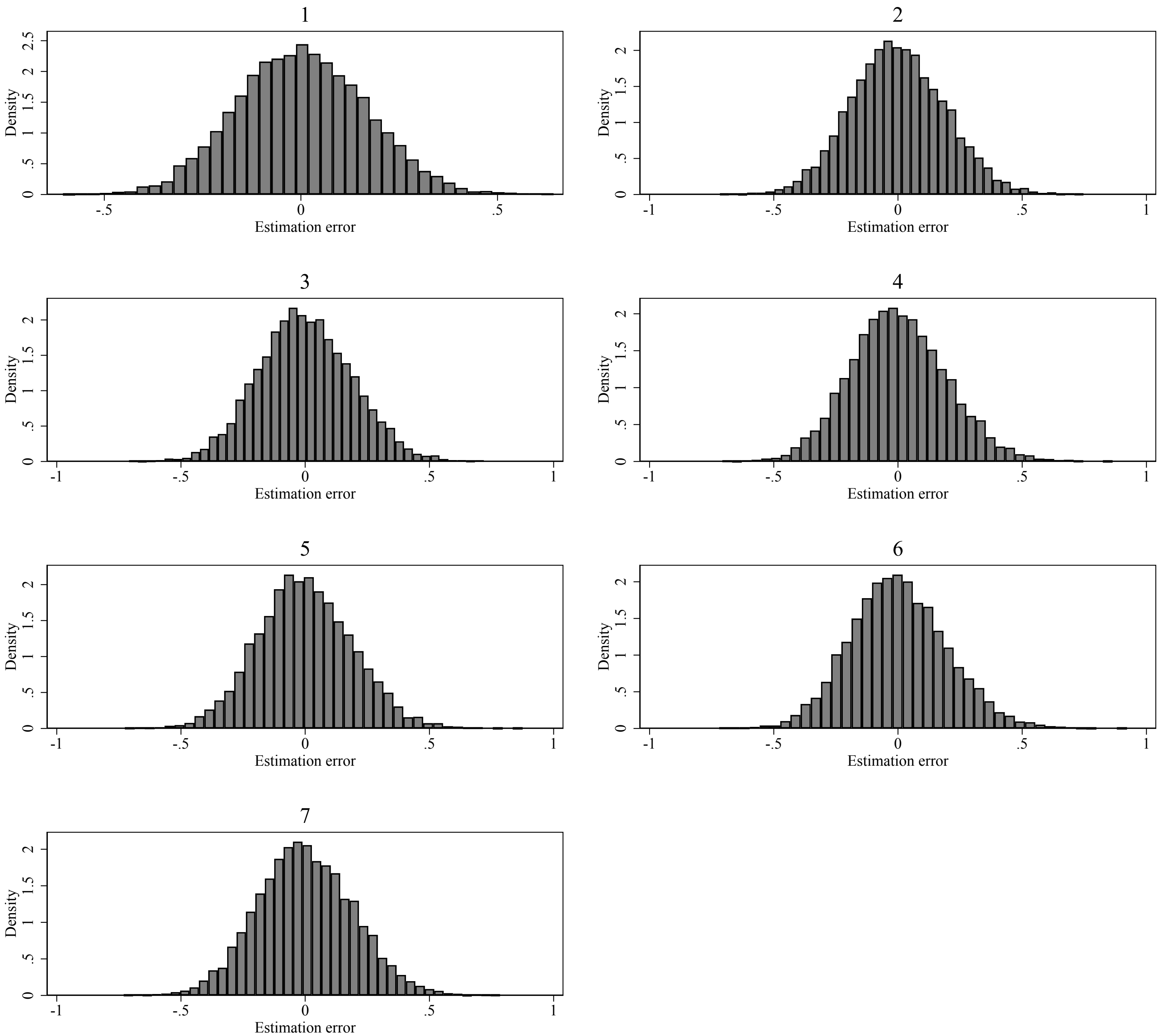
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.7: Simulation Results for Design A.1, δ = 0.05, N = 500



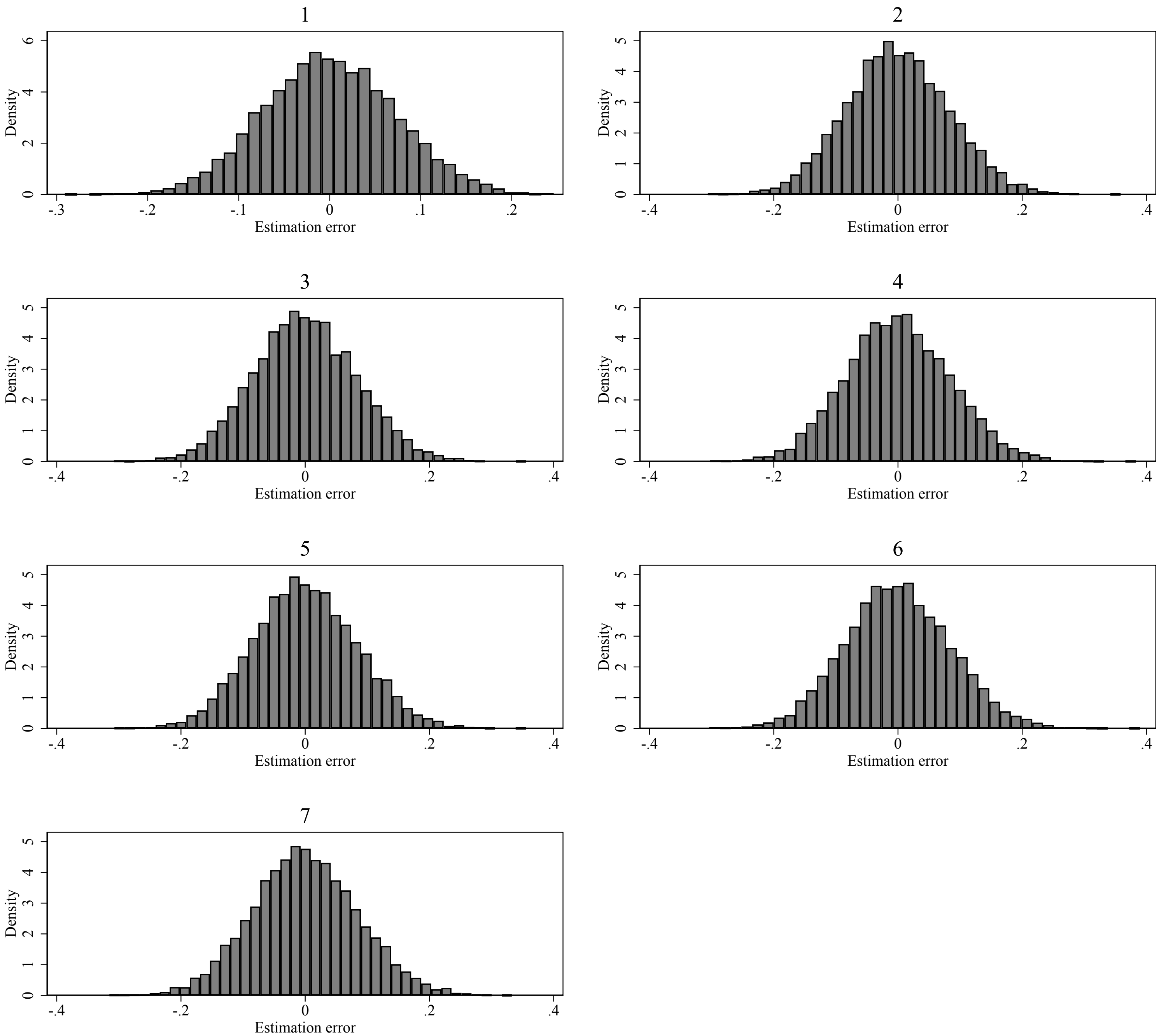
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ (= $\hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.8: Simulation Results for Design A.1, $\delta = 0.05$, $N = 1{,}000$



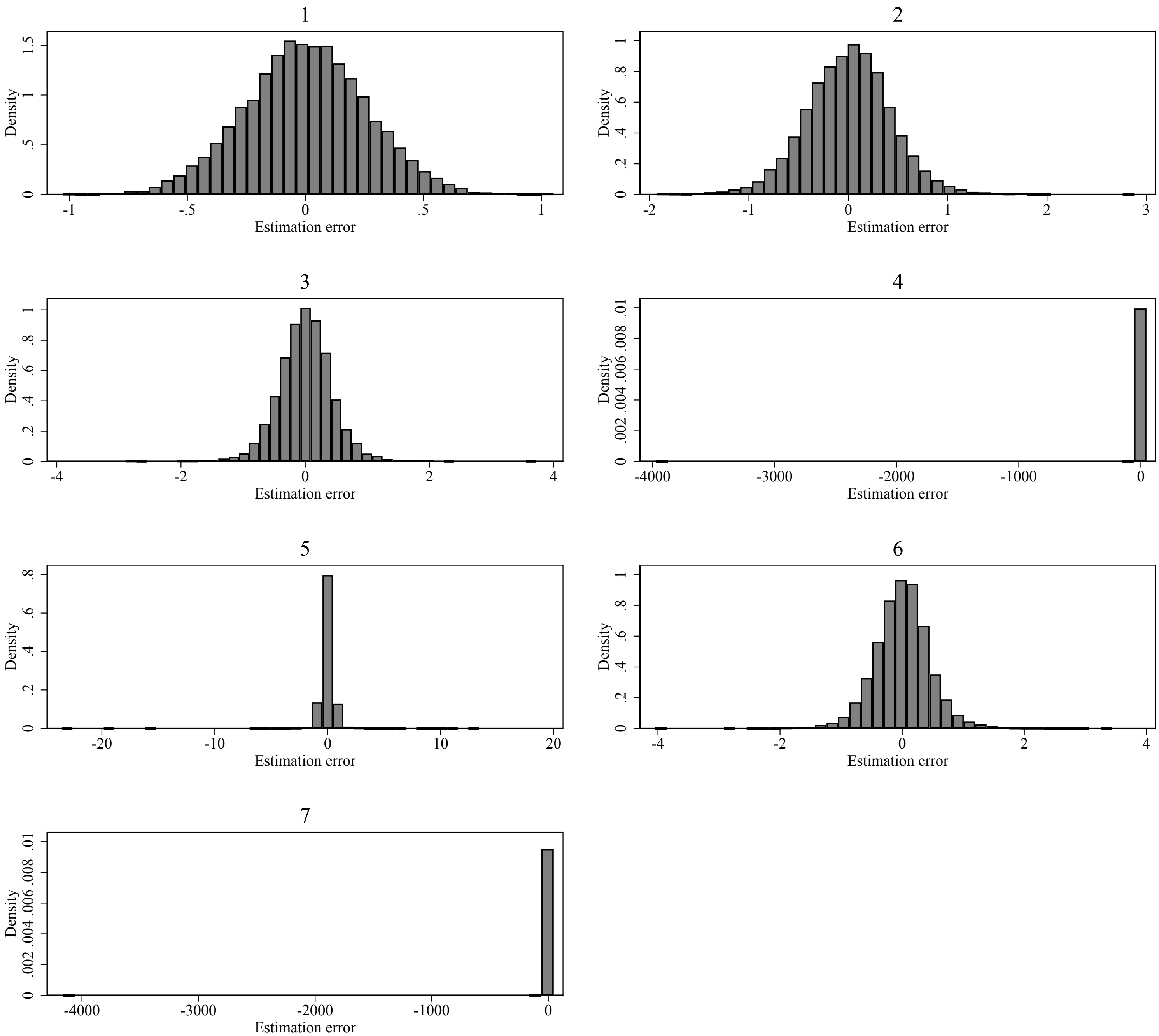
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.9: Simulation Results for Design A.1, δ = 0.05, N = 5,000



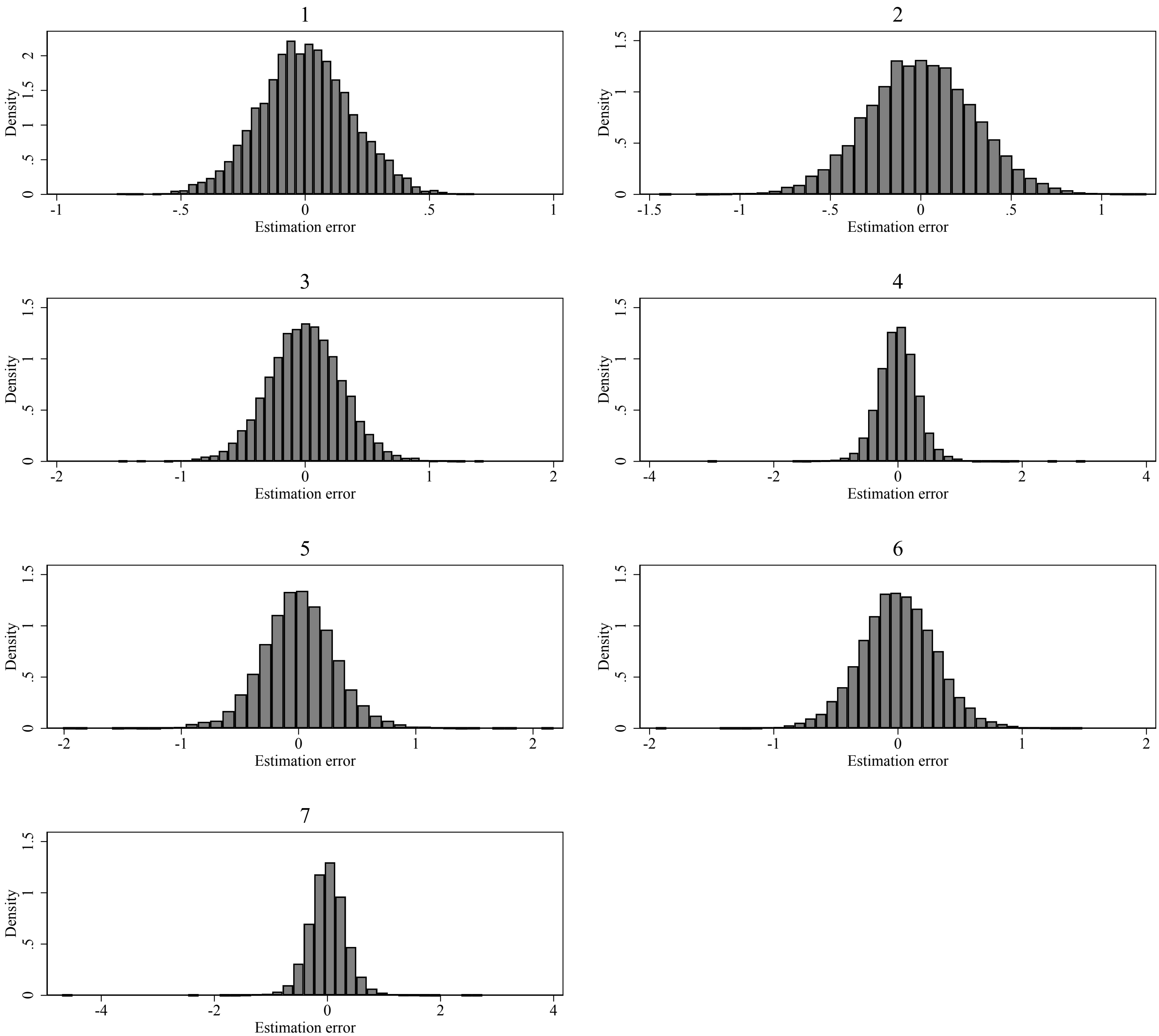
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ (= $\hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.10: Simulation Results for Design A.2, $\delta = 0.01$, $N = 500$



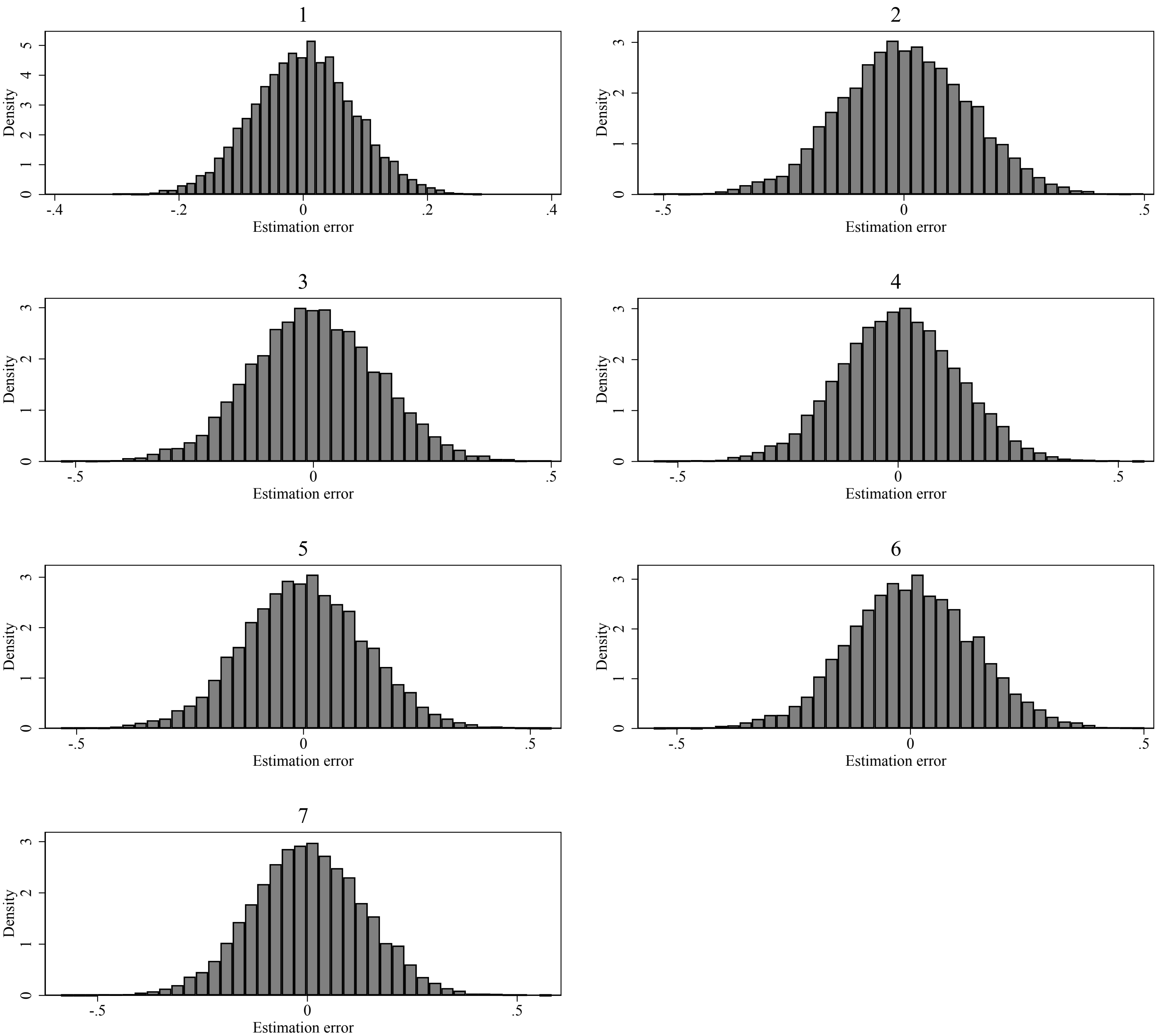
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ (= $\hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.11: Simulation Results for Design A.2, $\delta = 0.01$, $N = 1{,}000$



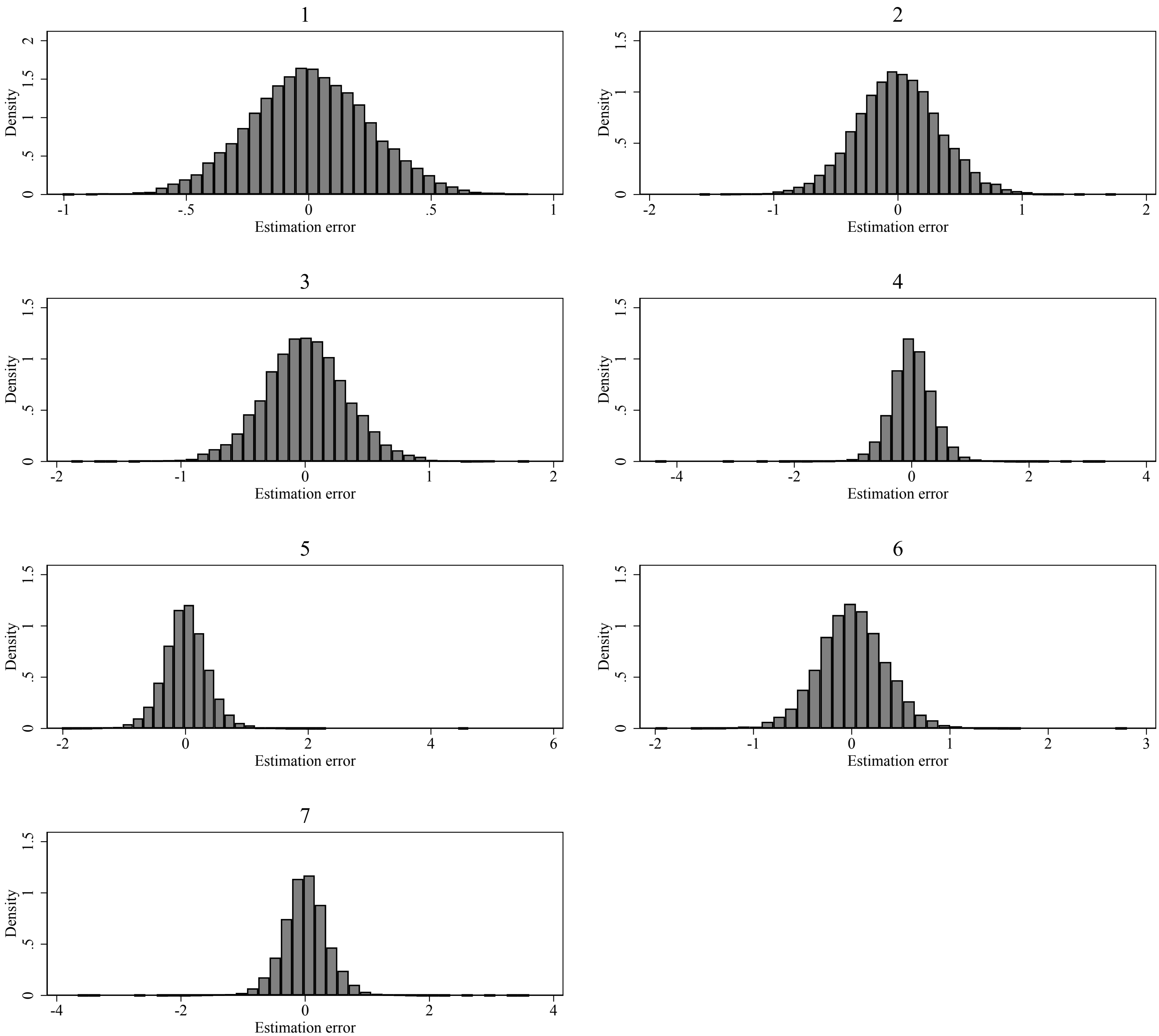
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.12: Simulation Results for Design A.2, $\delta = 0.01$, $N = 5{,}000$



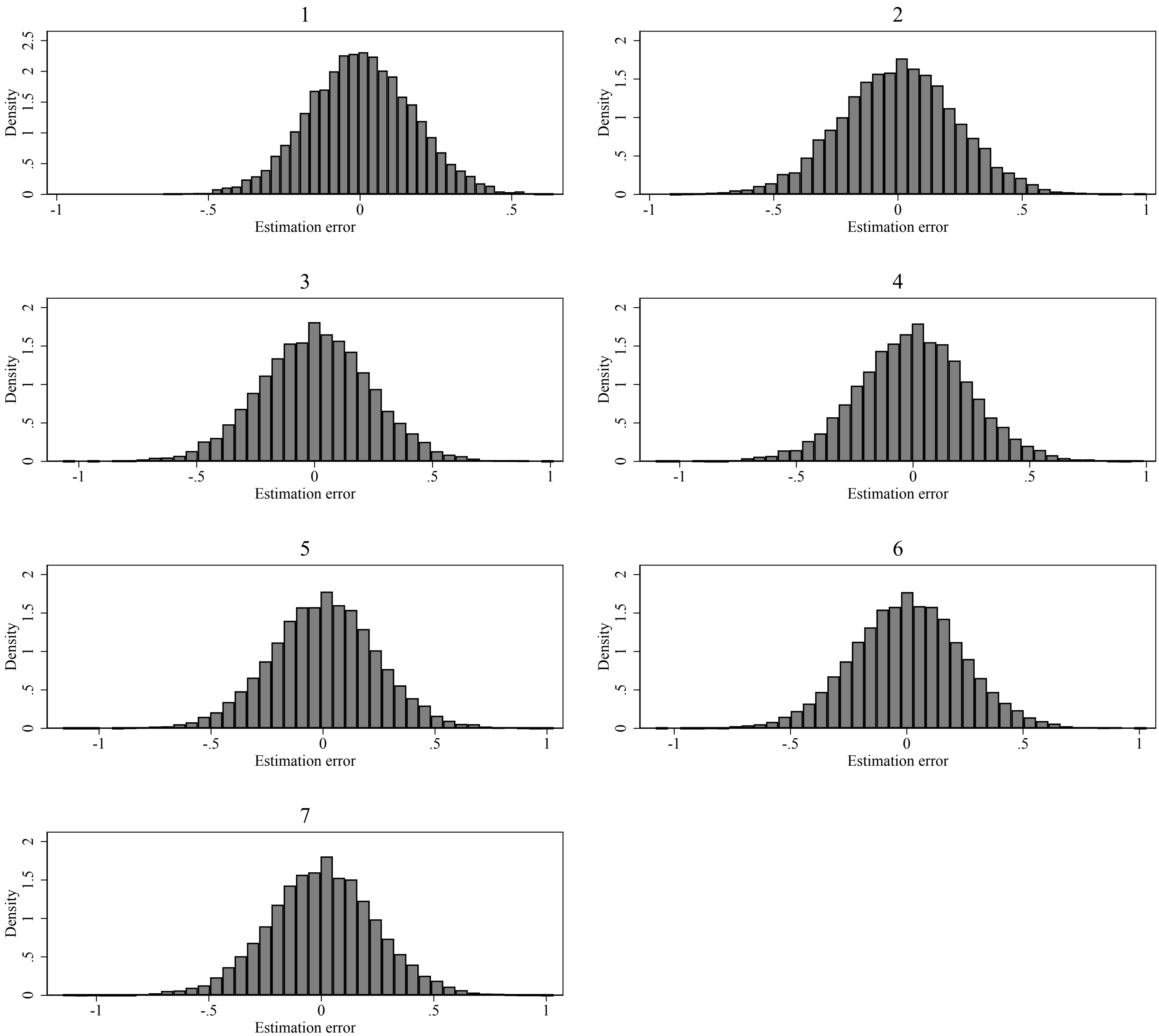
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.13: Simulation Results for Design A.2, $\delta = 0.02$, $N = 500$



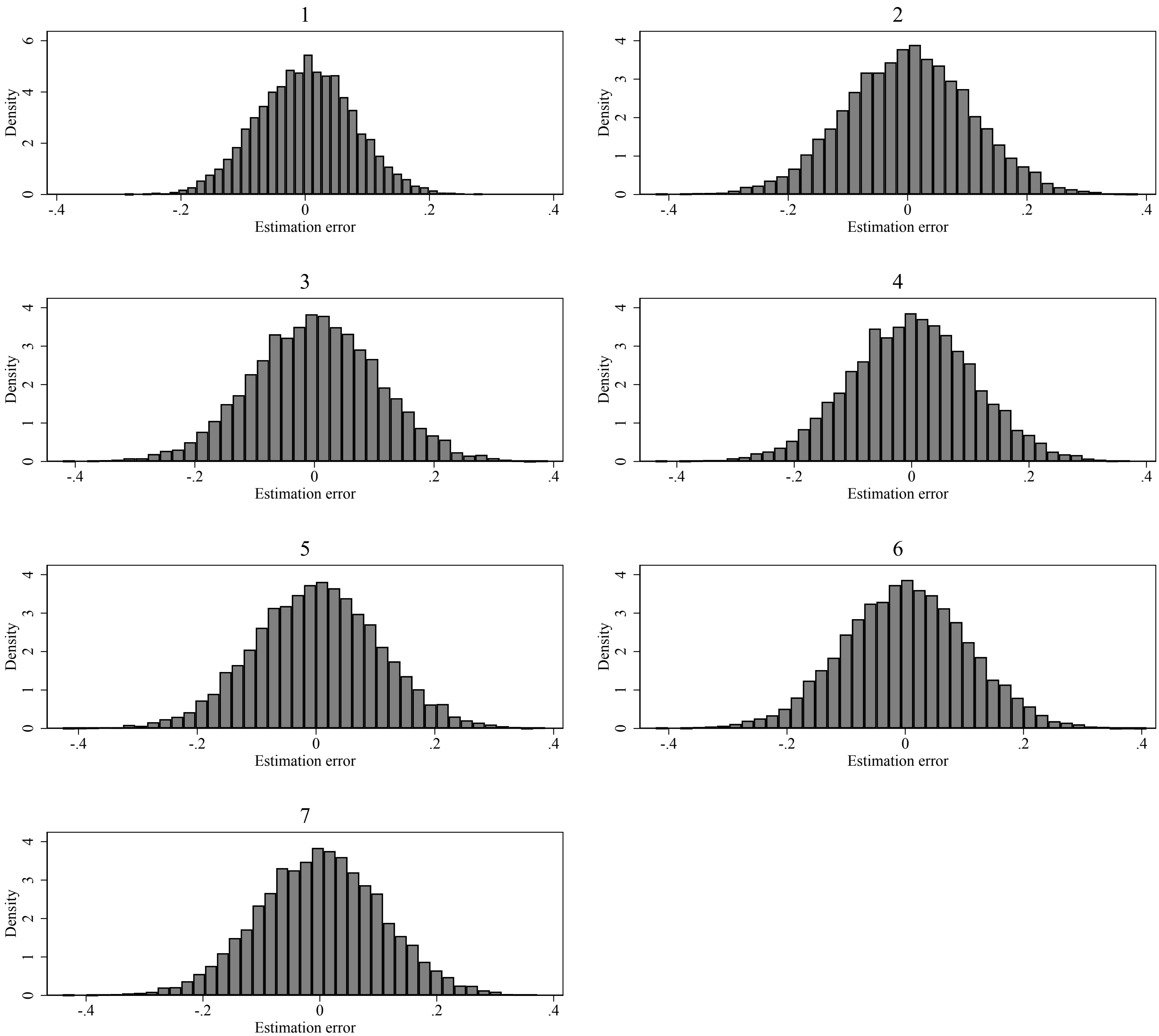
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ $(= \hat{\tau}_{a,1}^{ml})$. "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.14: Simulation Results for Design A.2, $\delta = 0.02$, $N = 1{,}000$



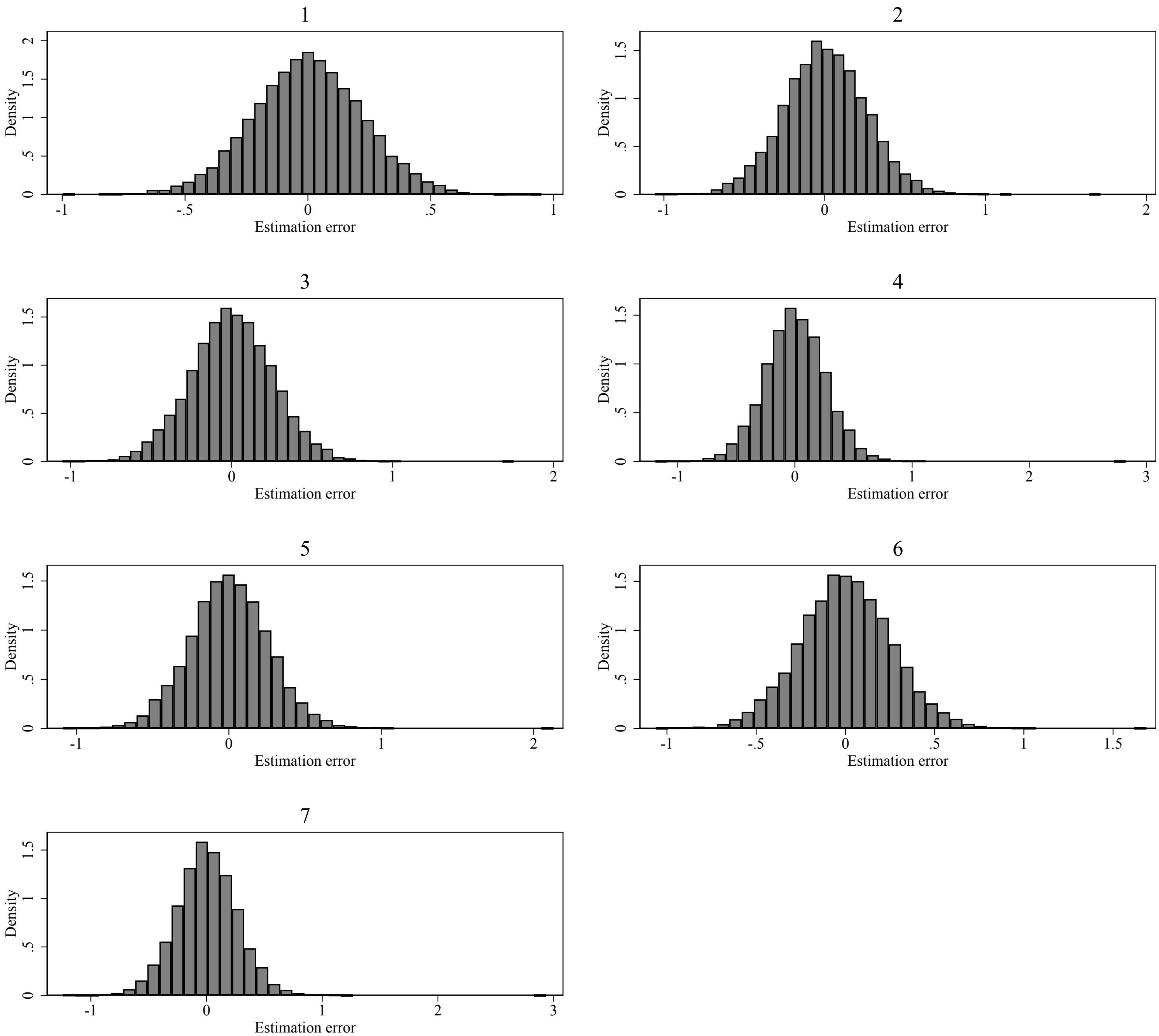
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.15: Simulation Results for Design A.2, $\delta = 0.02$, $N = 5{,}000$



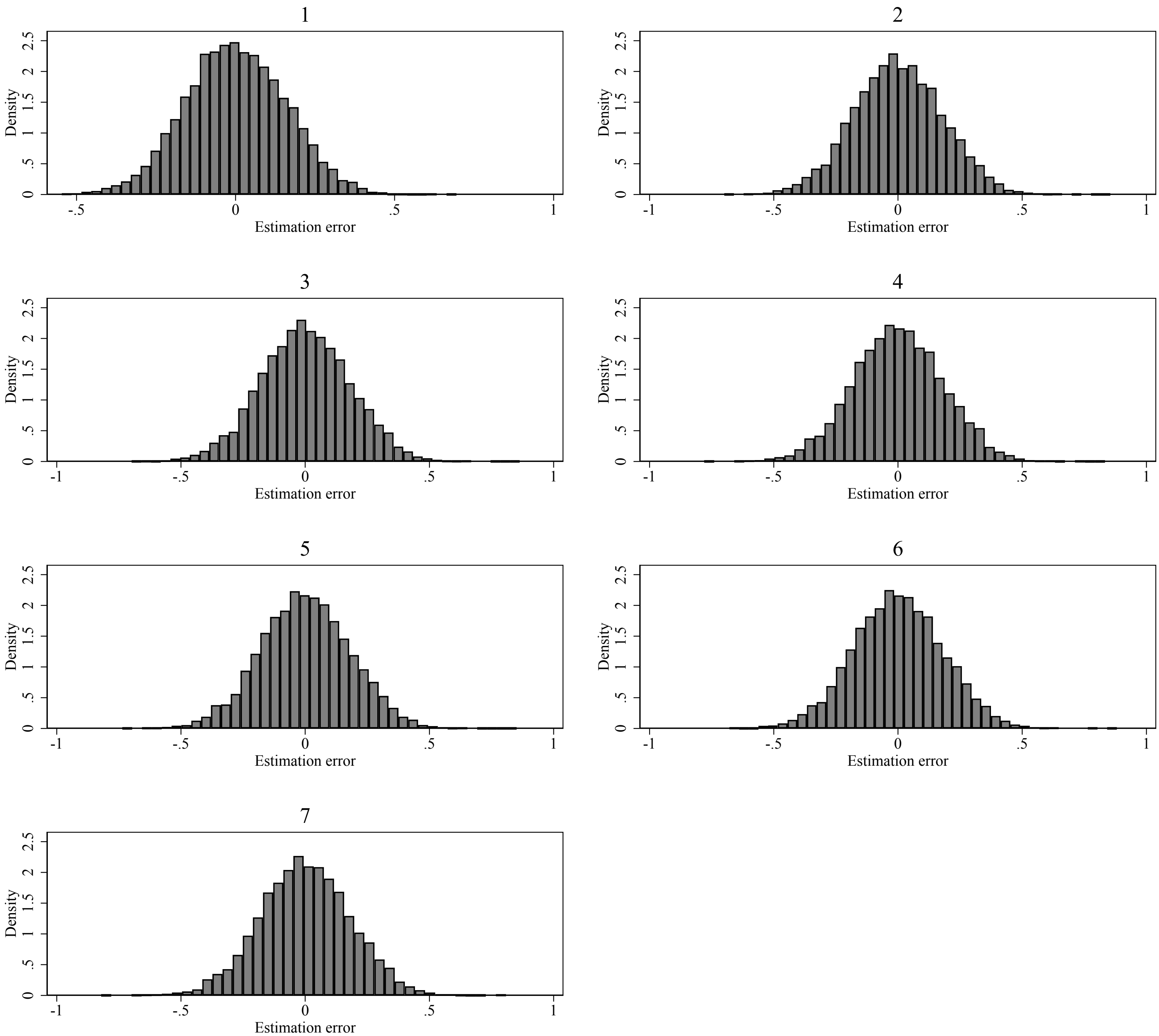
Notes: The details of this simulation design are provided in Section 5 in the main text. “1” corresponds to the 2SLS estimator that additively controls for X. “2” corresponds to $\hat{\tau}_u^{cb}$. “3” corresponds to $\hat{\tau}_u^{ml}$. “4” corresponds to $\hat{\tau}_{a,10}^{ml}$. “5” corresponds to $\hat{\tau}_a^{ml}$. “6” corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). “7” corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.16: Simulation Results for Design A.2, $\delta = 0.05$, $N = 500$



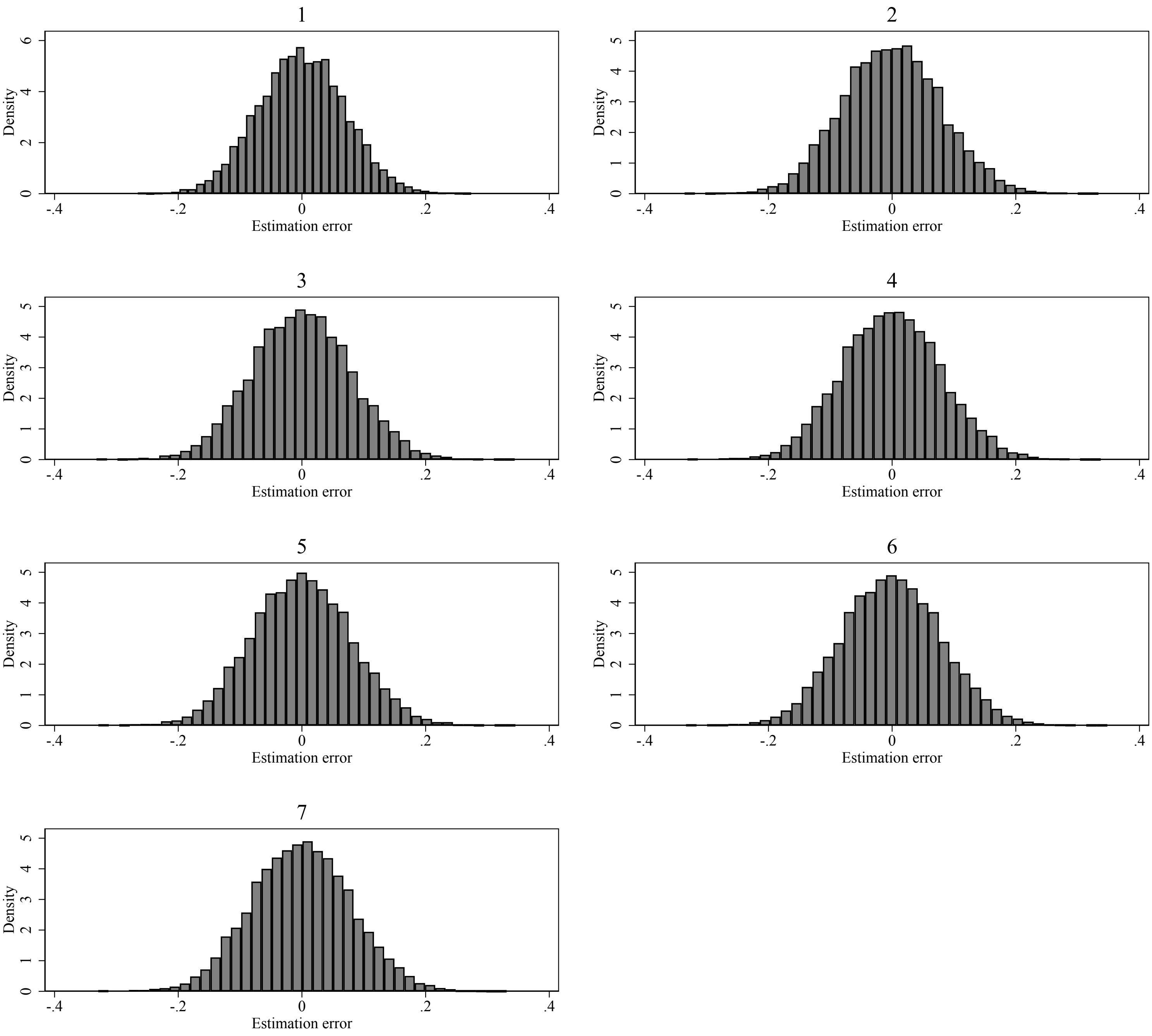
Notes: The details of this simulation design are provided in Section 5 in the main text. “1” corresponds to the 2SLS estimator that additively controls for X. “2” corresponds to $\hat{\tau}_u^{cb}$. “3” corresponds to $\hat{\tau}_u^{ml}$. “4” corresponds to $\hat{\tau}_{a,10}^{ml}$. “5” corresponds to $\hat{\tau}_a^{ml}$. “6” corresponds to $\hat{\tau}_t^{ml}$ $(= \hat{\tau}_{a,1}^{ml})$. “7” corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.17: Simulation Results for Design A.2, $\delta = 0.05$, $N = 1{,}000$



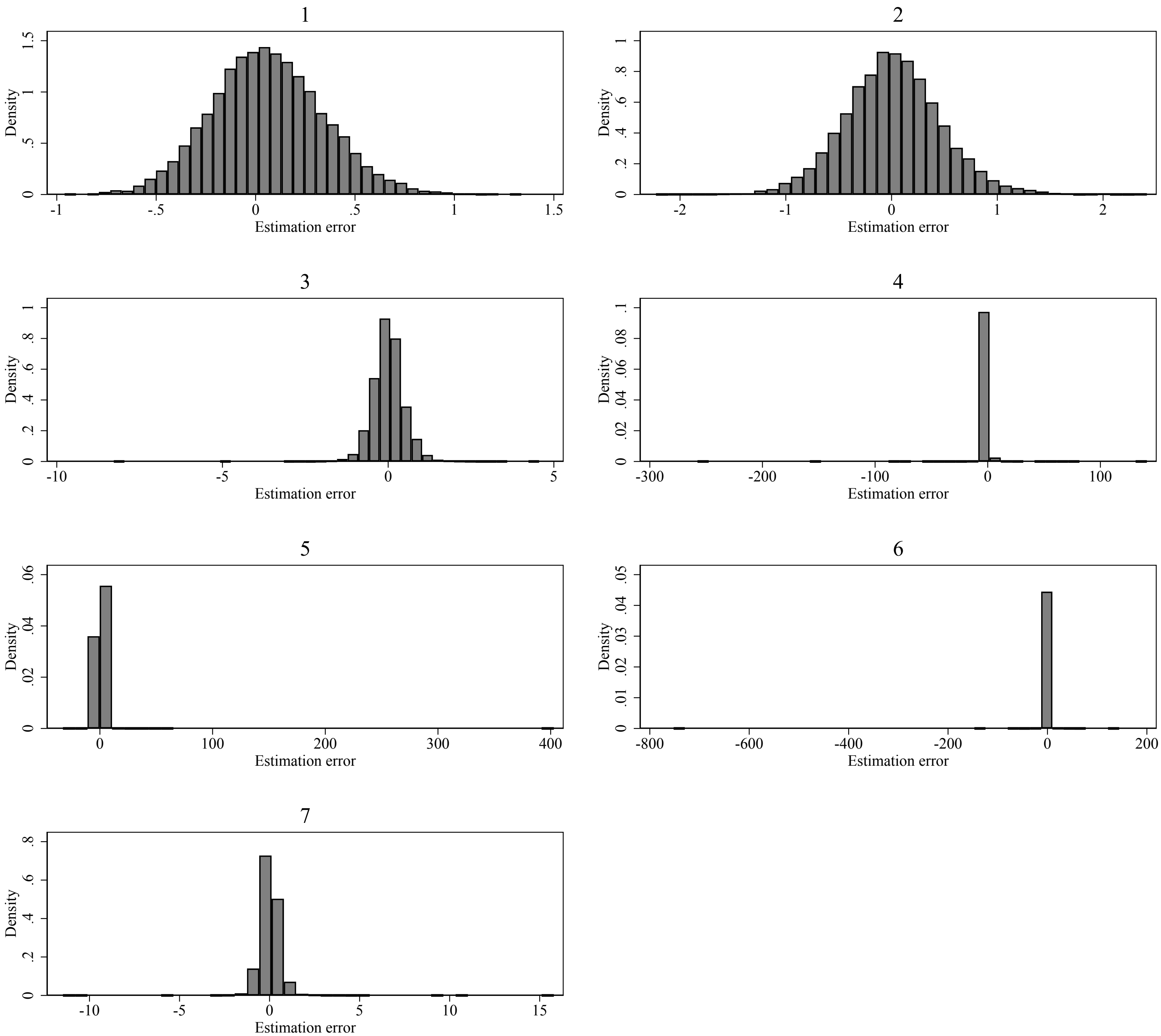
Notes: The details of this simulation design are provided in Section 5 in the main text. “1” corresponds to the 2SLS estimator that additively controls for X. “2” corresponds to $\hat{\tau}_u^{cb}$. “3” corresponds to $\hat{\tau}_u^{ml}$. “4” corresponds to $\hat{\tau}_{a,10}^{ml}$. “5” corresponds to $\hat{\tau}_a^{ml}$. “6” corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). “7” corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.18: Simulation Results for Design A.2, δ = 0.05, N = 5,000



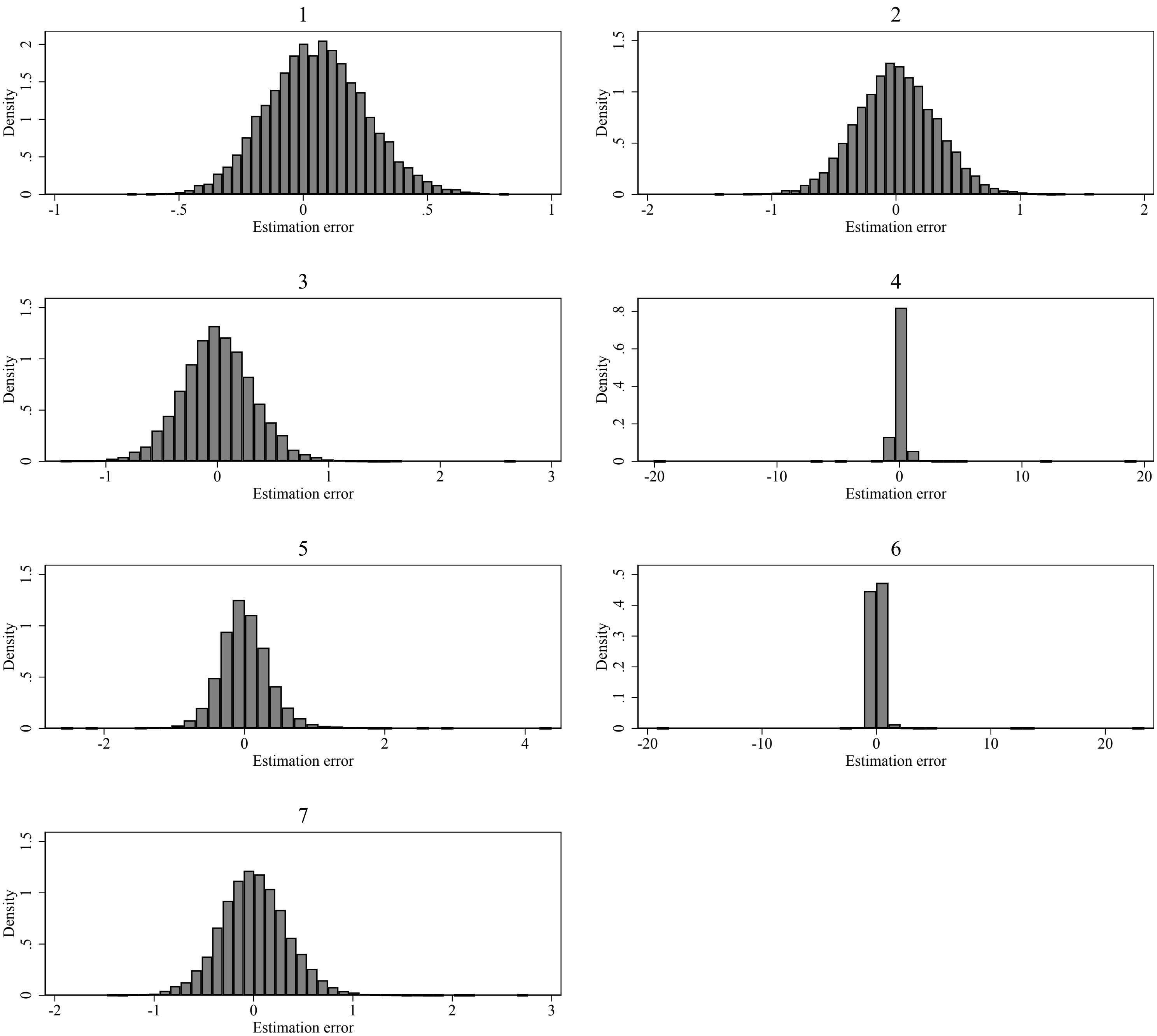
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ (= $\hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.19: Simulation Results for Design B, $\delta = 0.01$, $N = 500$



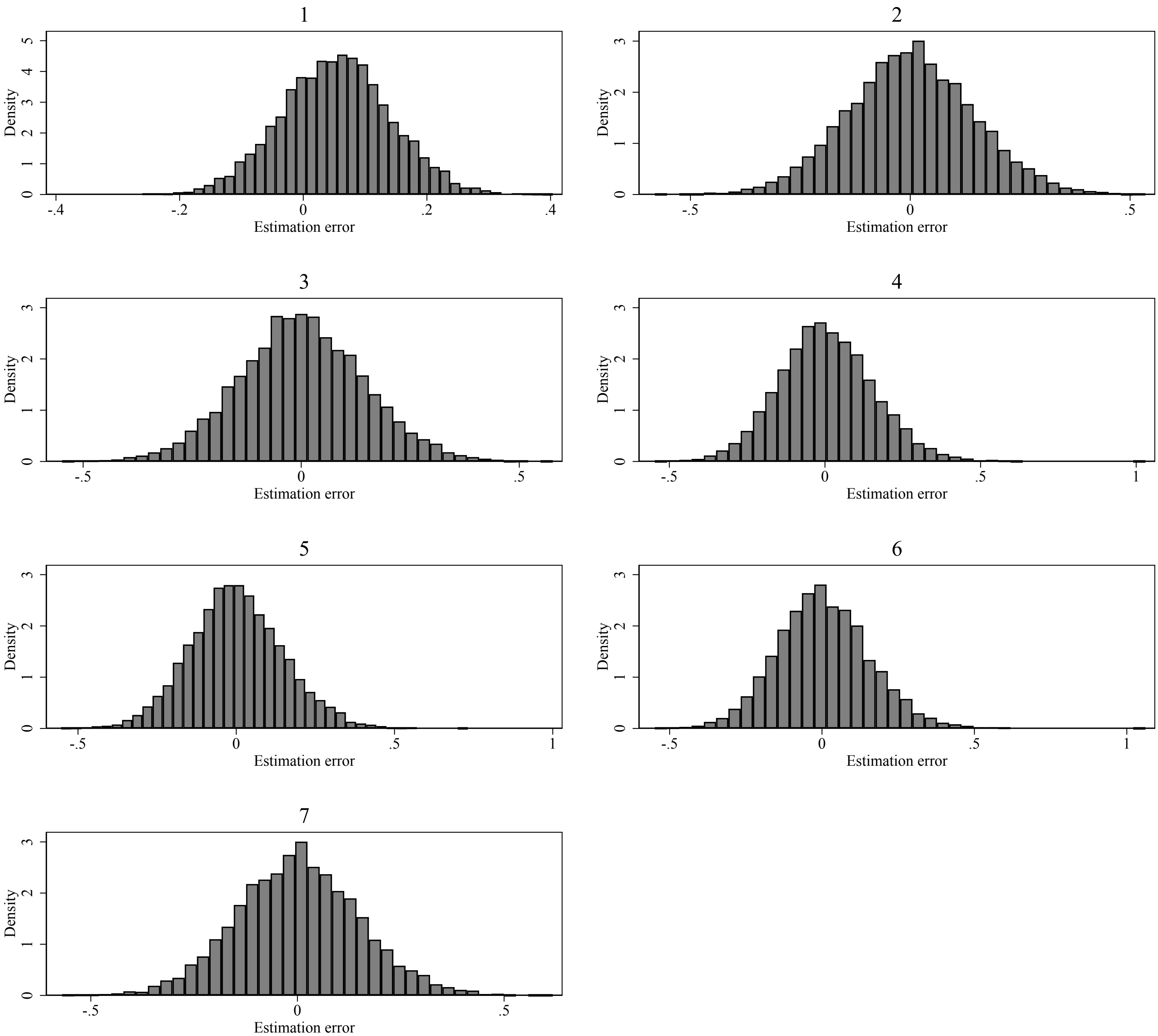
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.20: Simulation Results for Design B, $\delta = 0.01$, $N = 1{,}000$



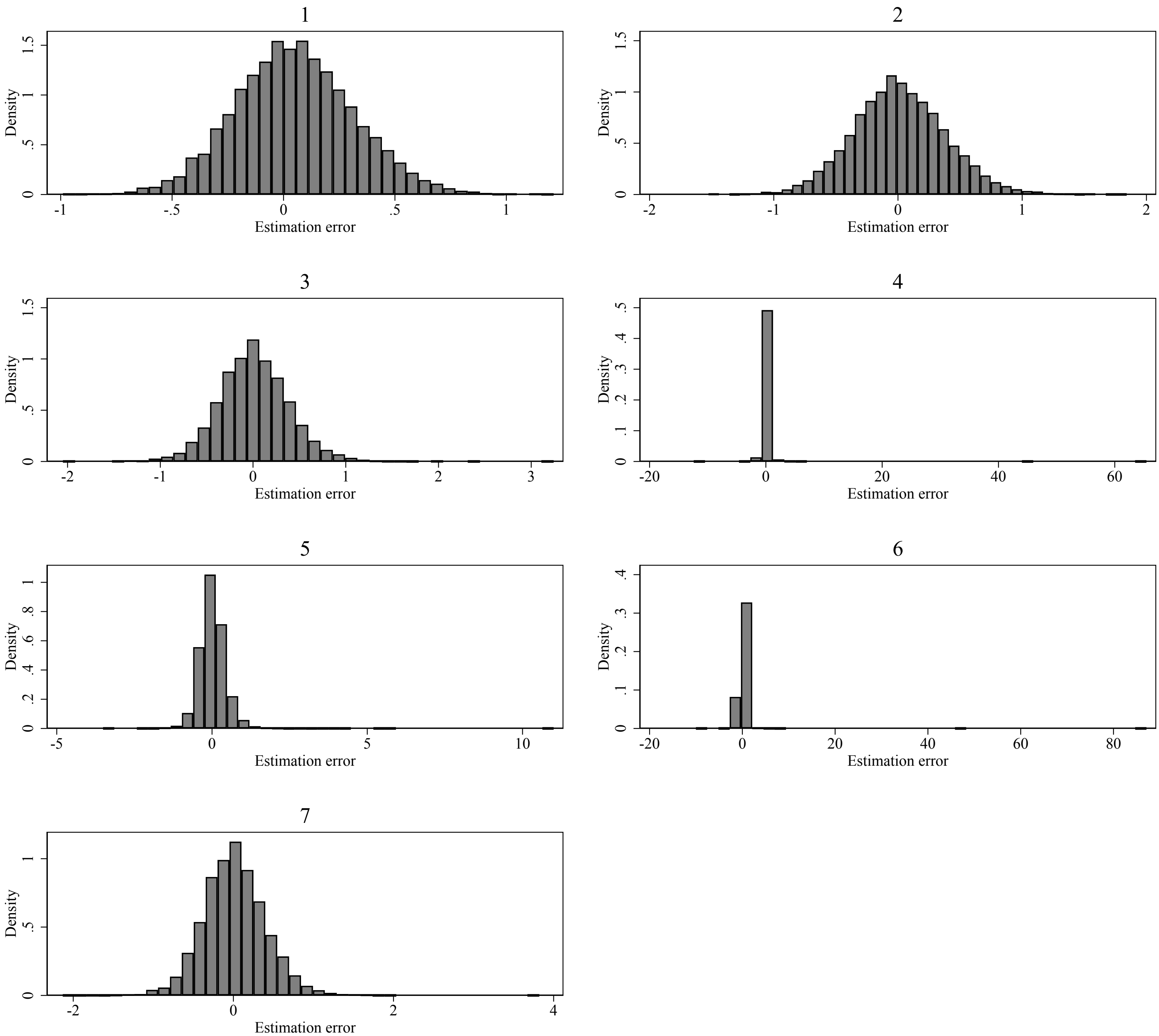
Notes: The details of this simulation design are provided in Section 5 in the main text. “1” corresponds to the 2SLS estimator that additively controls for X. “2” corresponds to $\hat{\tau}_u^{cb}$. “3” corresponds to $\hat{\tau}_u^{ml}$. “4” corresponds to $\hat{\tau}_{a,10}^{ml}$. “5” corresponds to $\hat{\tau}_a^{ml}$. “6” corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). “7” corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.21: Simulation Results for Design B, $\delta = 0.01$, $N = 5{,}000$



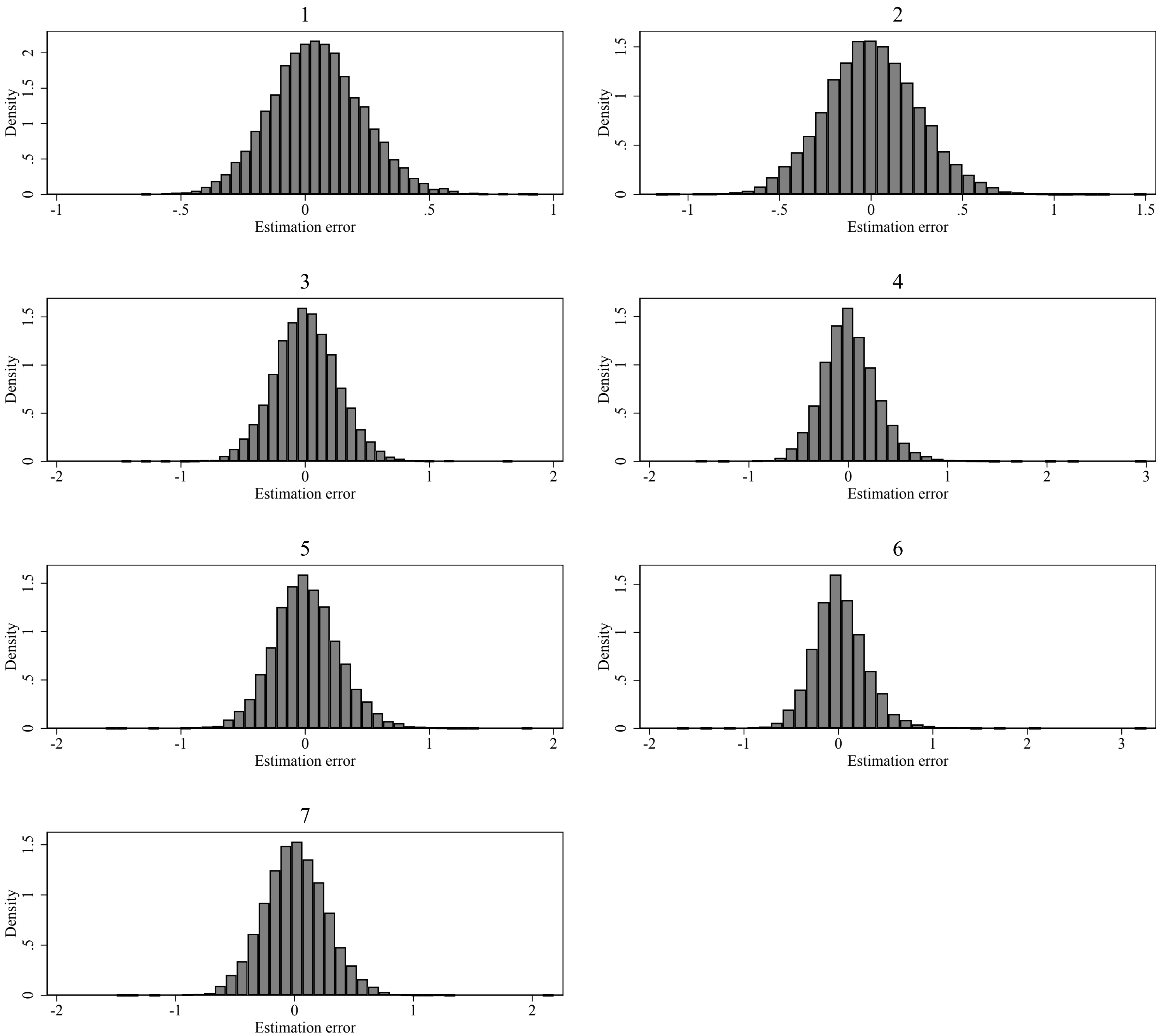
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.22: Simulation Results for Design B, $\delta = 0.02$, $N = 500$



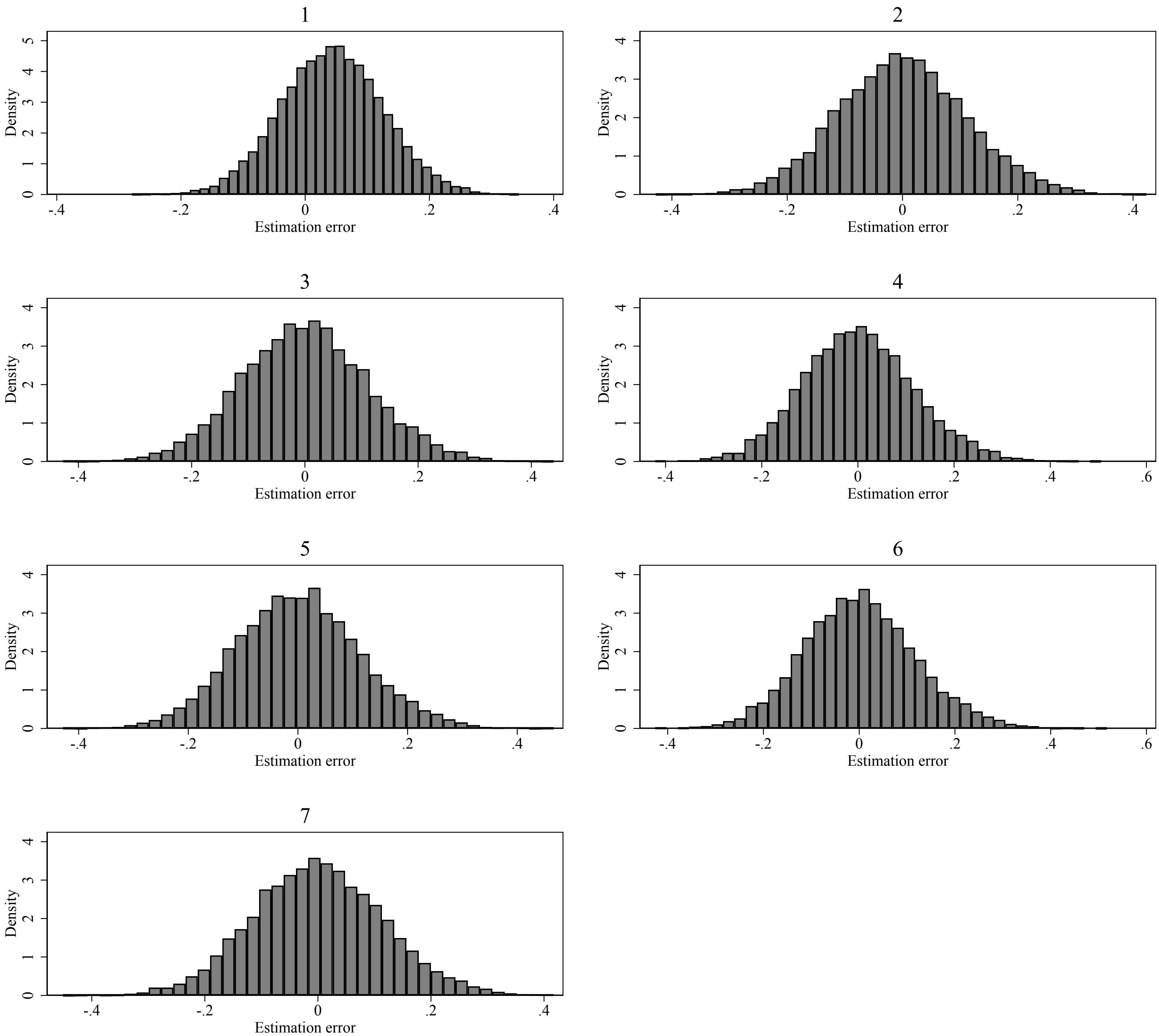
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.23: Simulation Results for Design B, $\delta = 0.02$, $N = 1{,}000$



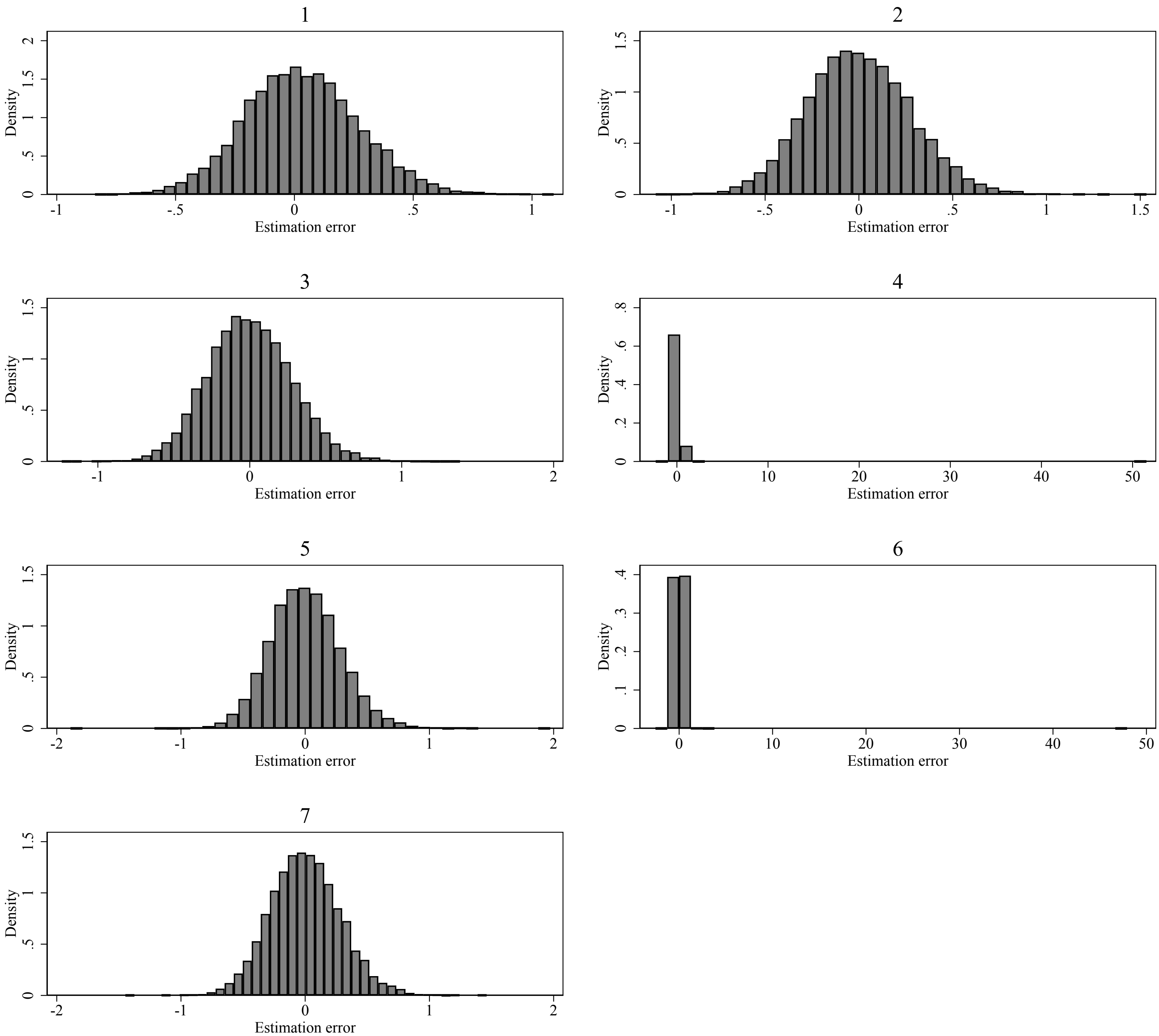
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.24: Simulation Results for Design B, $\delta = 0.02$, $N = 5{,}000$



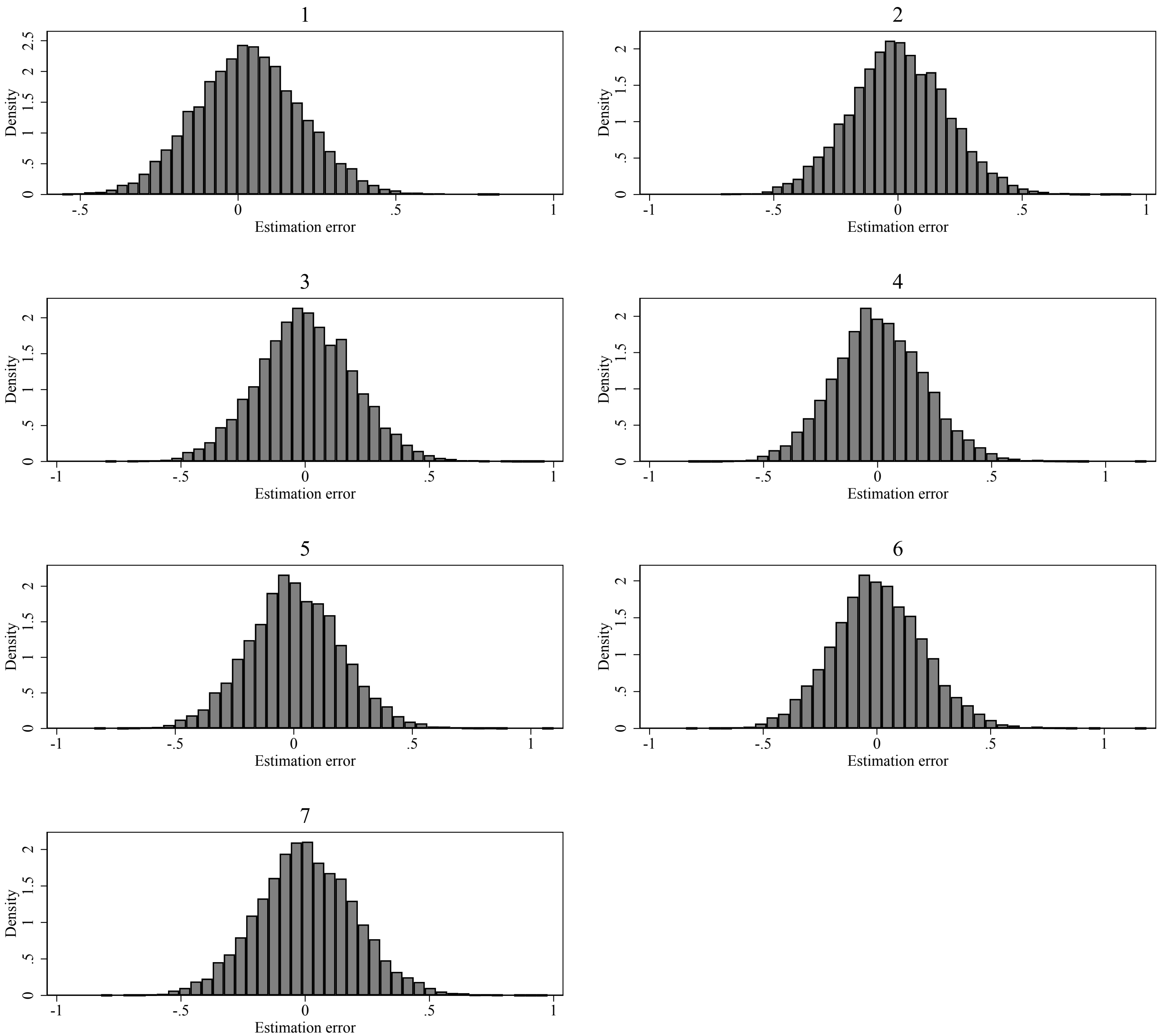
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.25: Simulation Results for Design B, δ = 0.05, N = 500



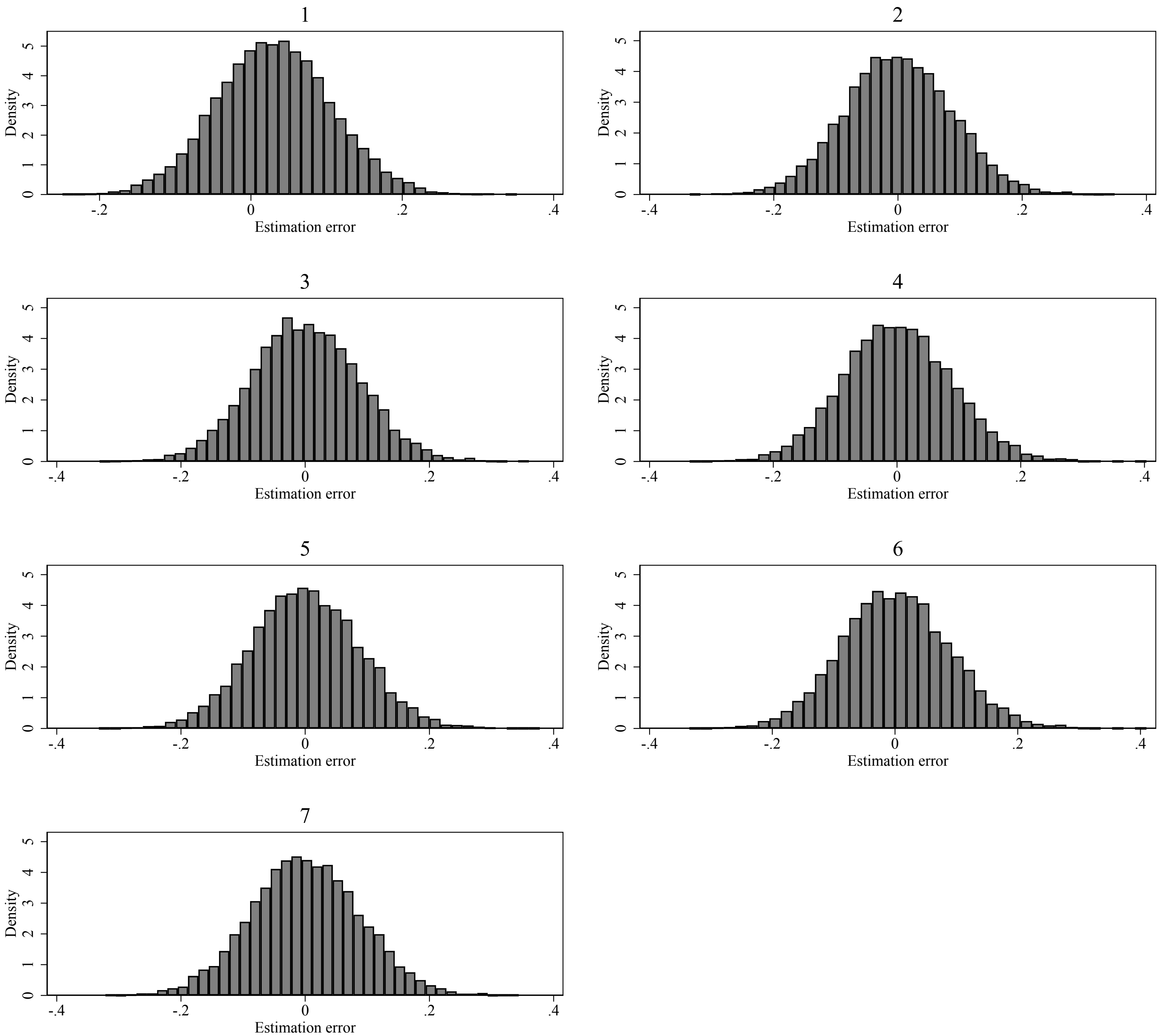
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ (= $\hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.26: Simulation Results for Design B, $\delta = 0.05$, $N = 1{,}000$



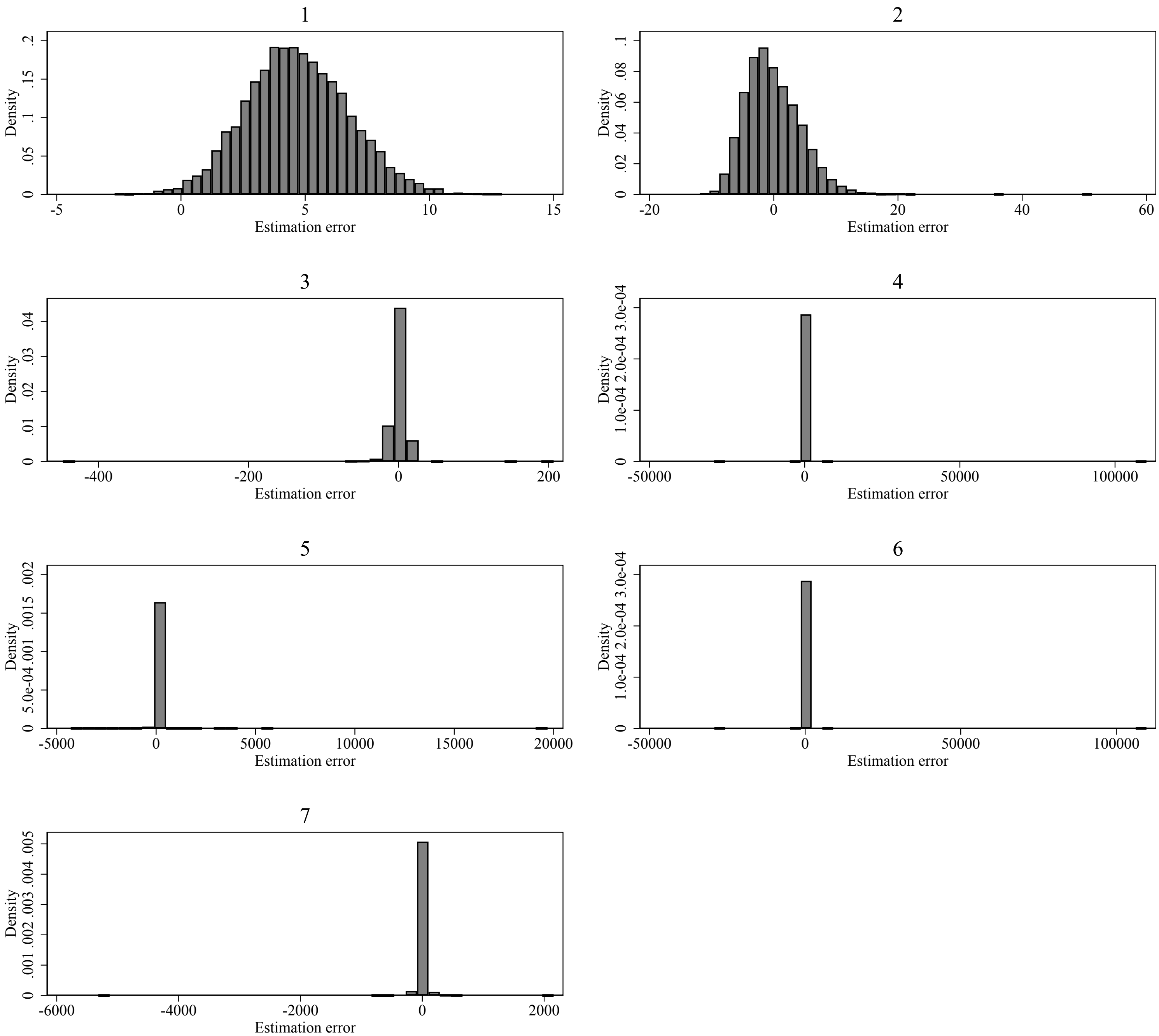
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ (= $\hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.27: Simulation Results for Design B, $\delta = 0.05$, $N = 5{,}000$



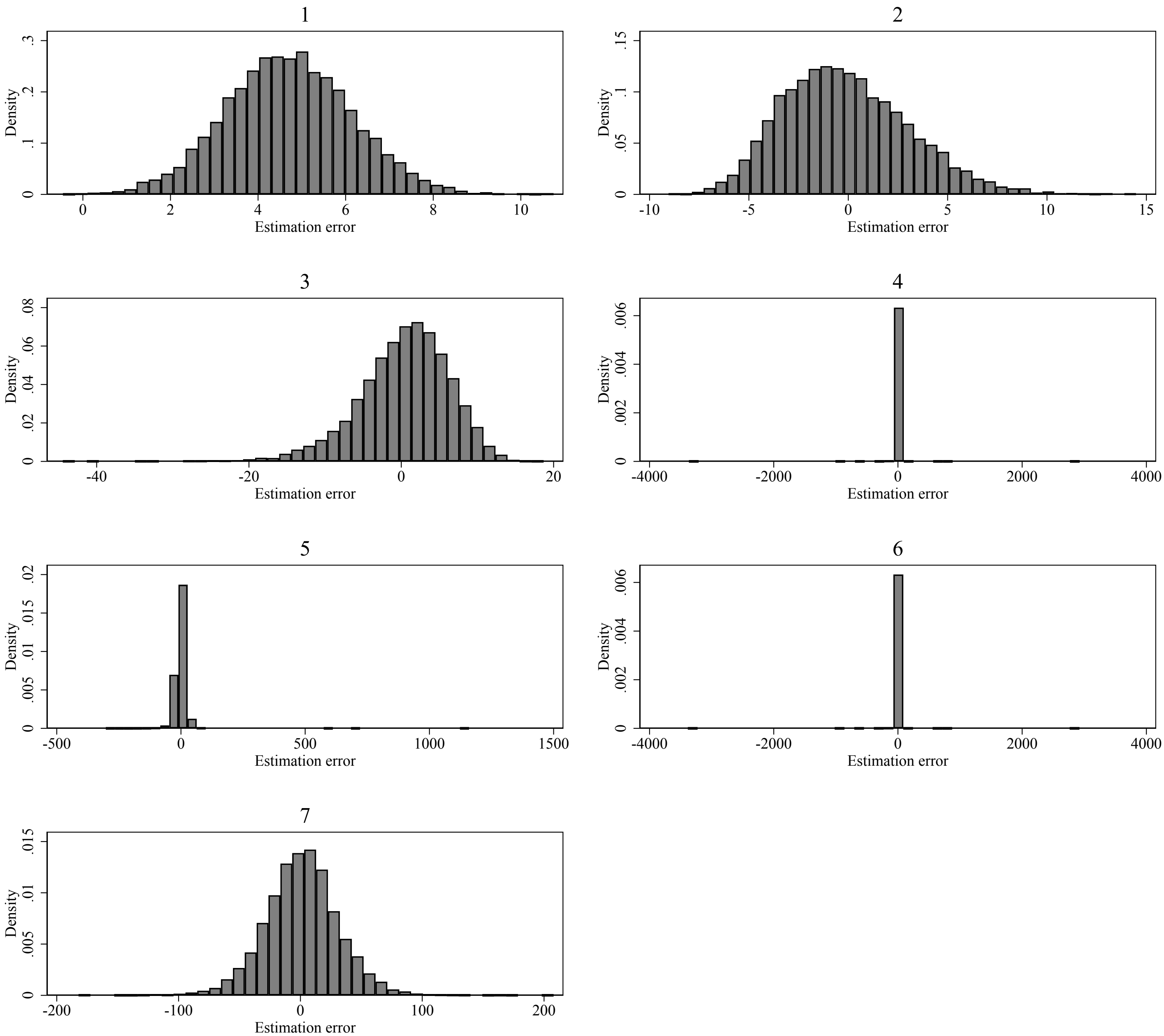
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.28: Simulation Results for Design C, $\delta = 0.01$, $N = 500$



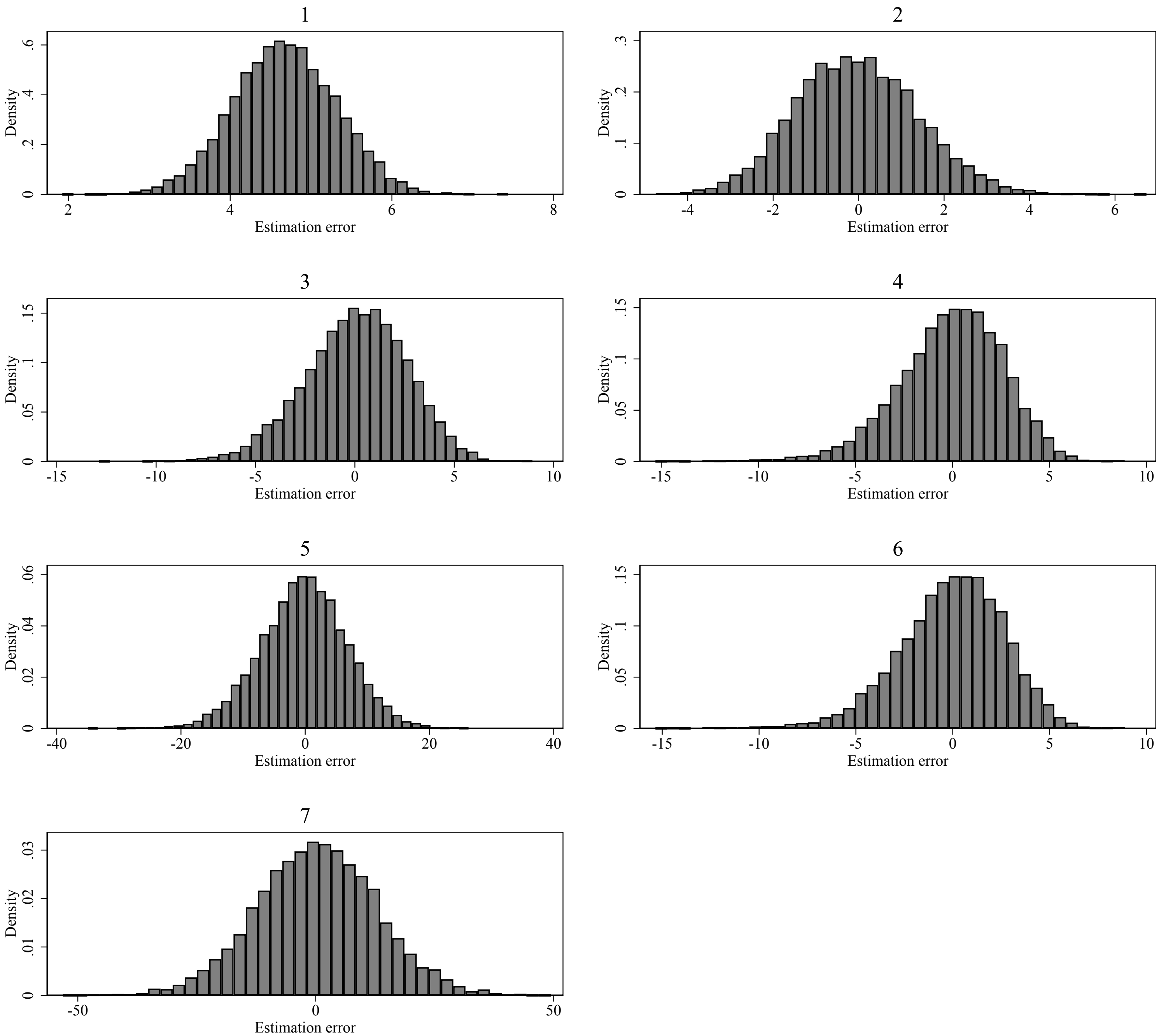
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.29: Simulation Results for Design C, $\delta = 0.01$, $N = 1{,}000$



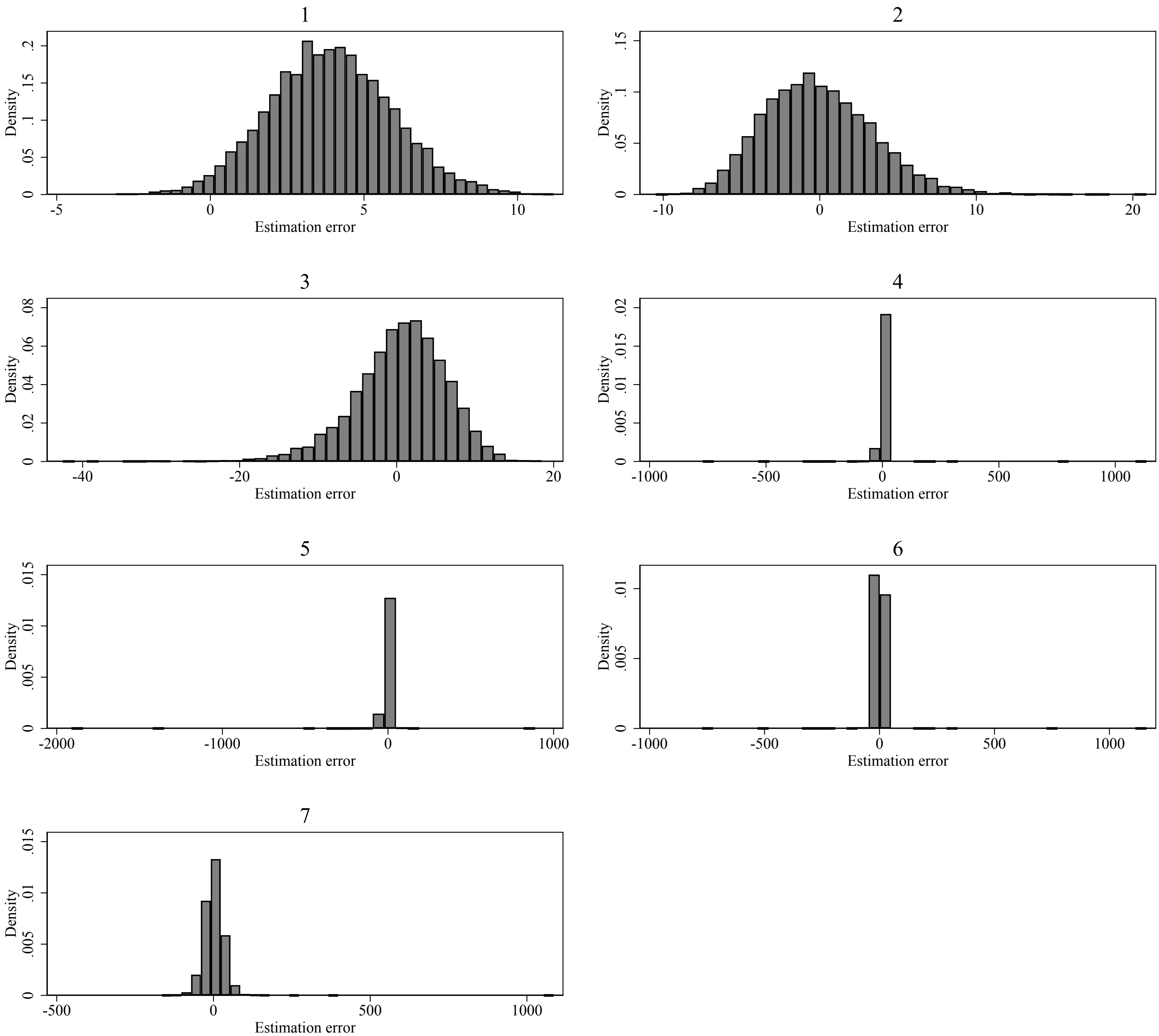
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.30: Simulation Results for Design C, $\delta = 0.01$, $N = 5{,}000$



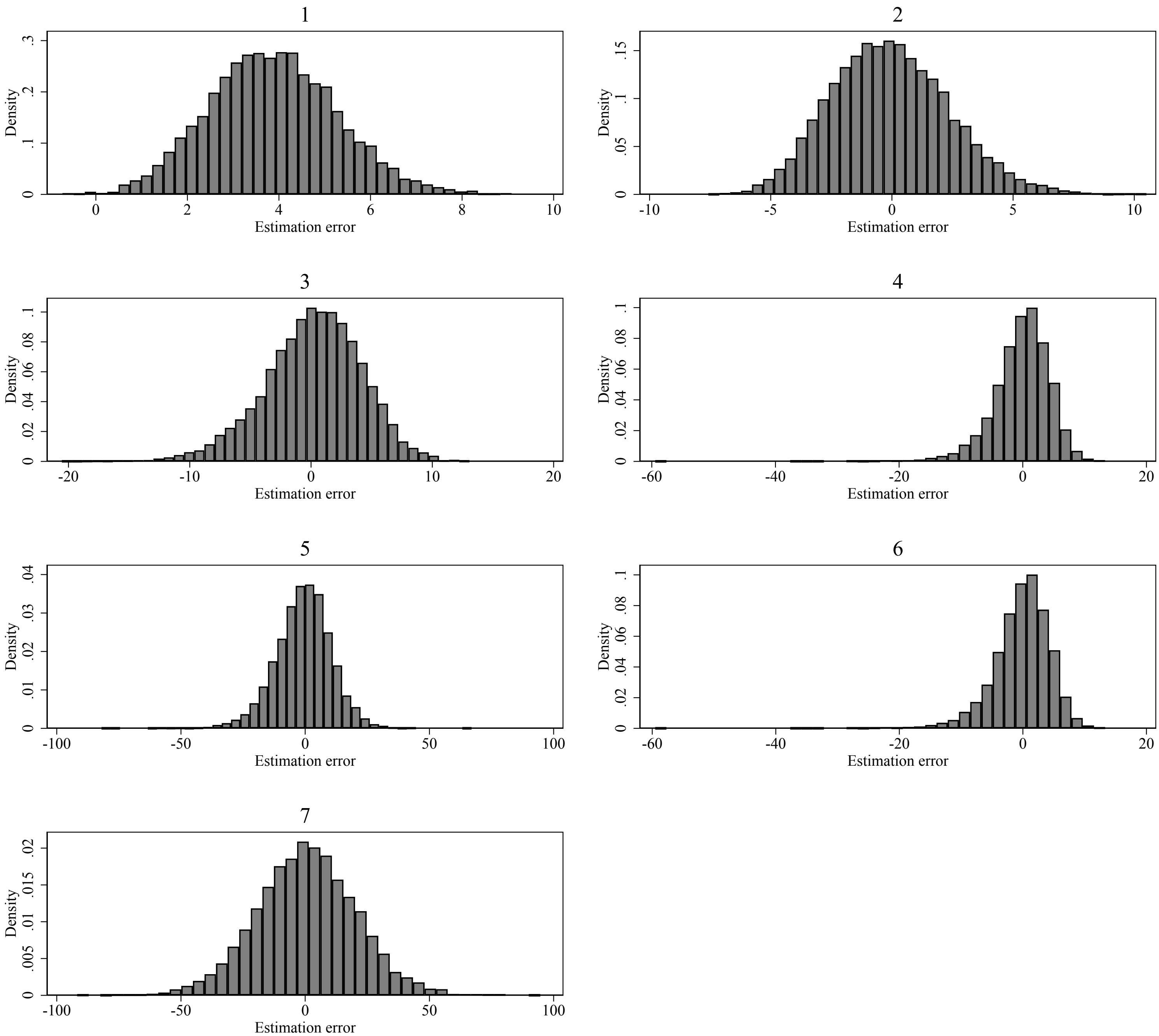
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ (= $\hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.31: Simulation Results for Design C, $\delta = 0.02$, $N = 500$



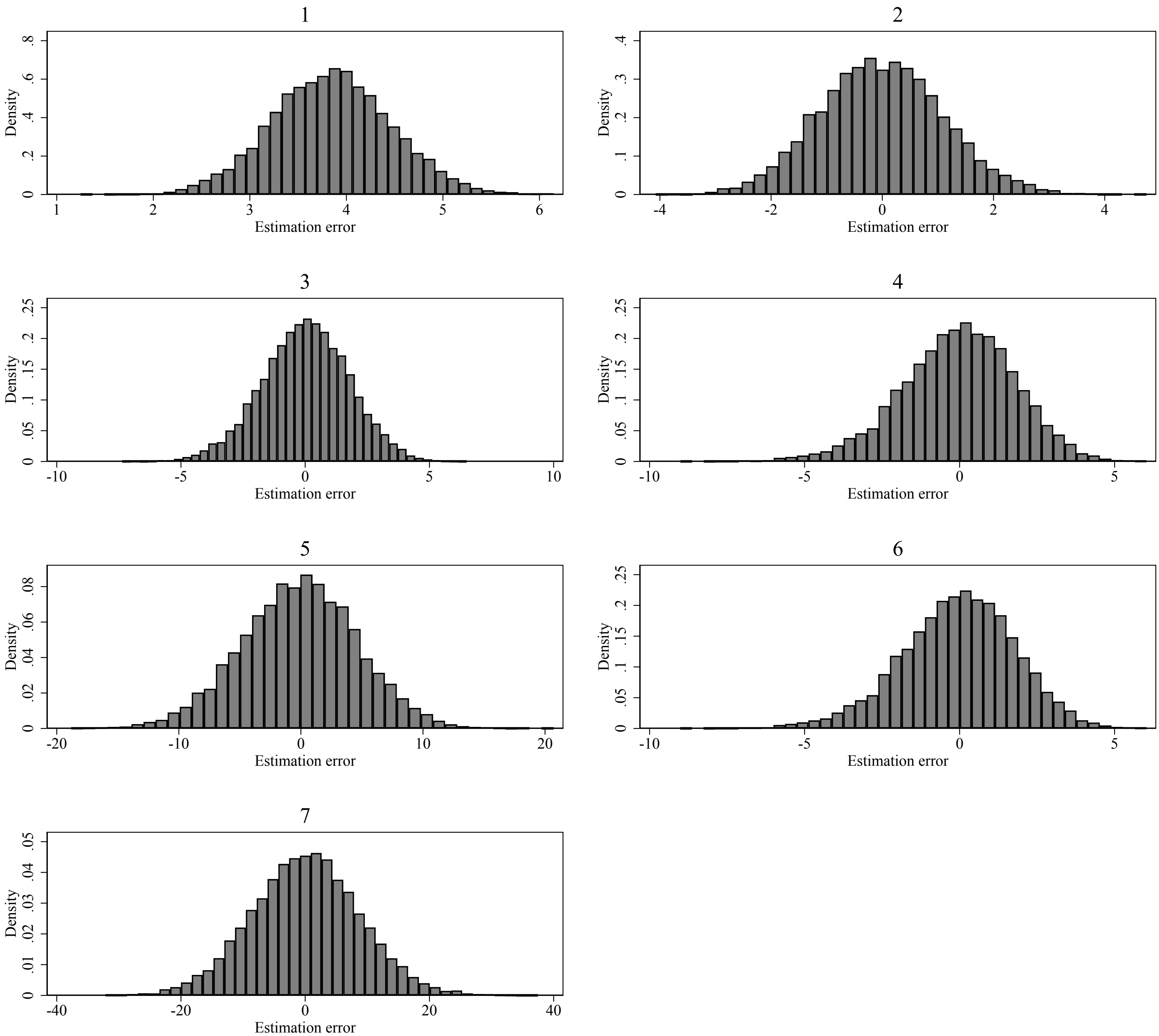
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.32: Simulation Results for Design C, $\delta = 0.02$, $N = 1{,}000$



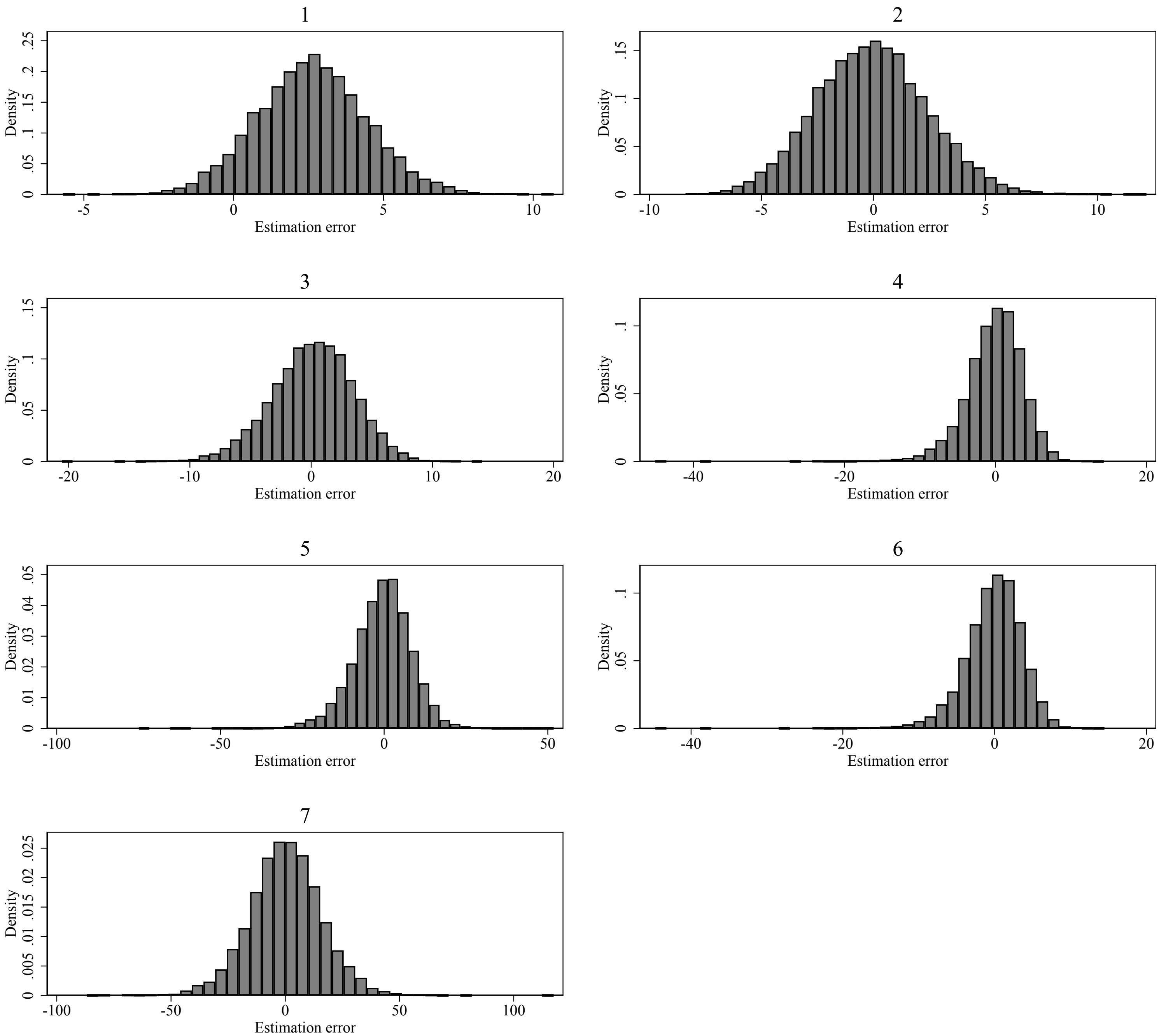
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ (= $\hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.33: Simulation Results for Design C, $\delta = 0.02$, $N = 5{,}000$



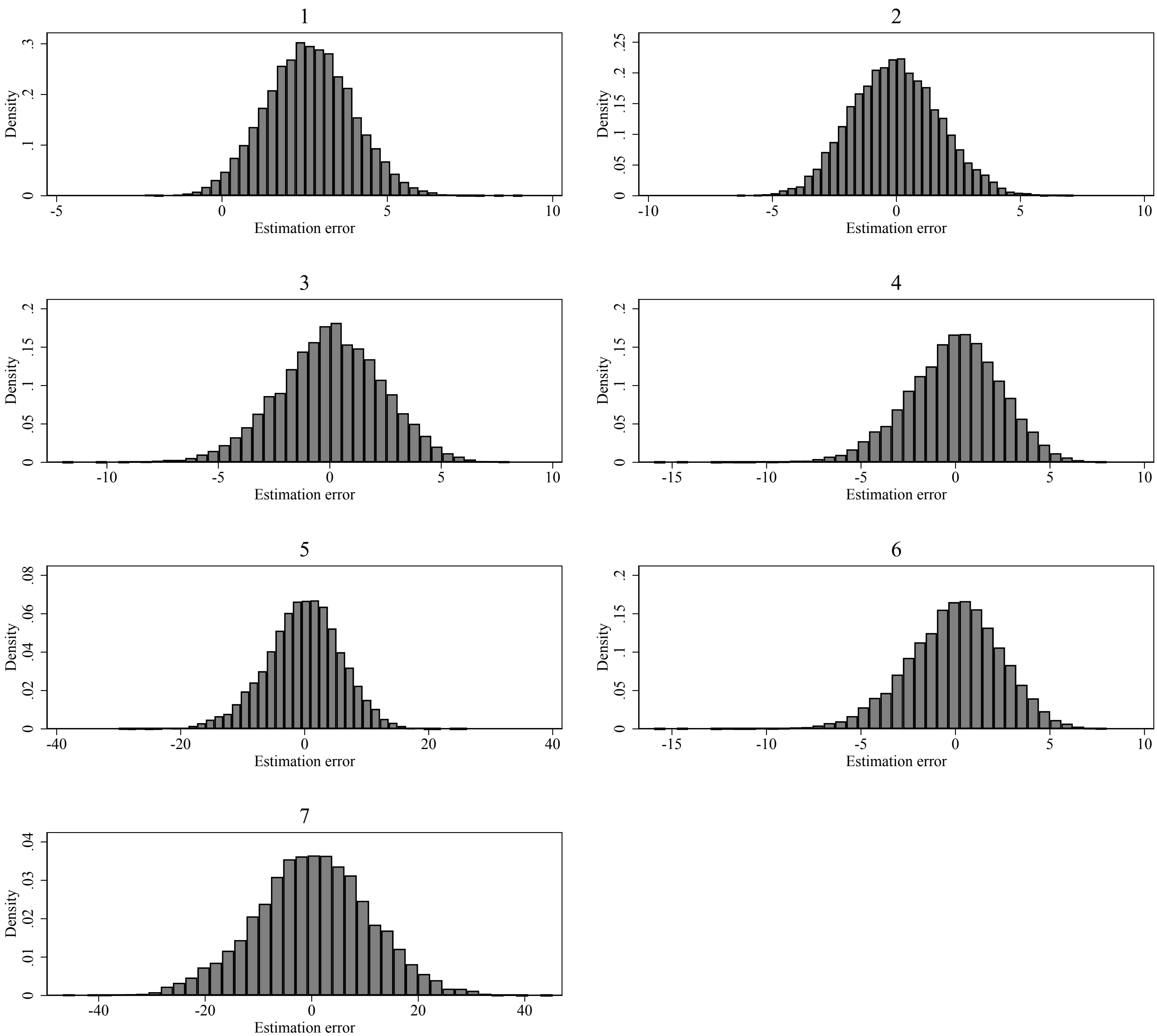
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.34: Simulation Results for Design C, $\delta = 0.05$, $N = 500$



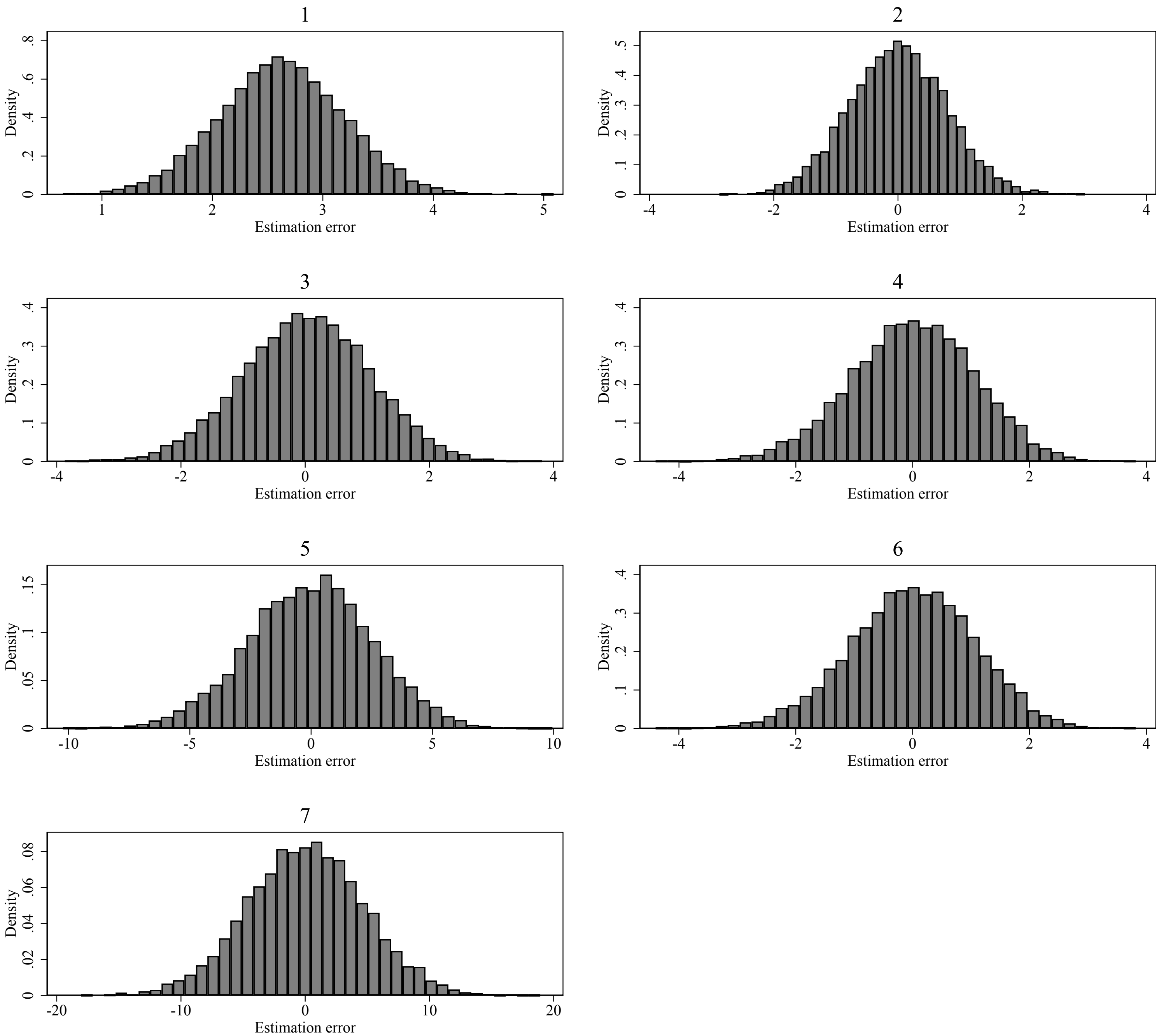
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ $(= \hat{\tau}_{a,1}^{ml})$. "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.35: Simulation Results for Design C, δ = 0.05, N = 1,000



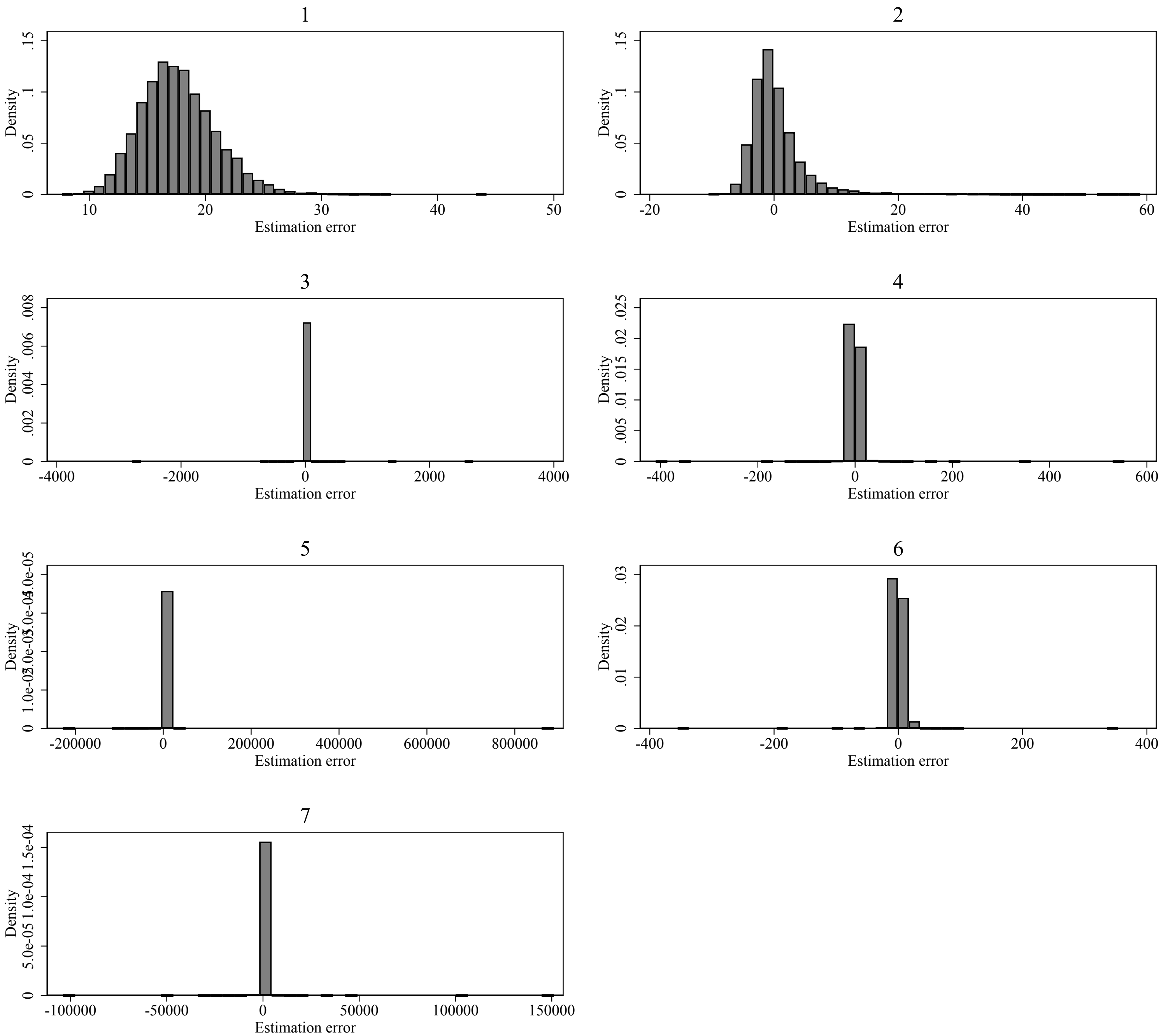
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ (= $\hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.36: Simulation Results for Design C, $\delta = 0.05$, $N = 5{,}000$



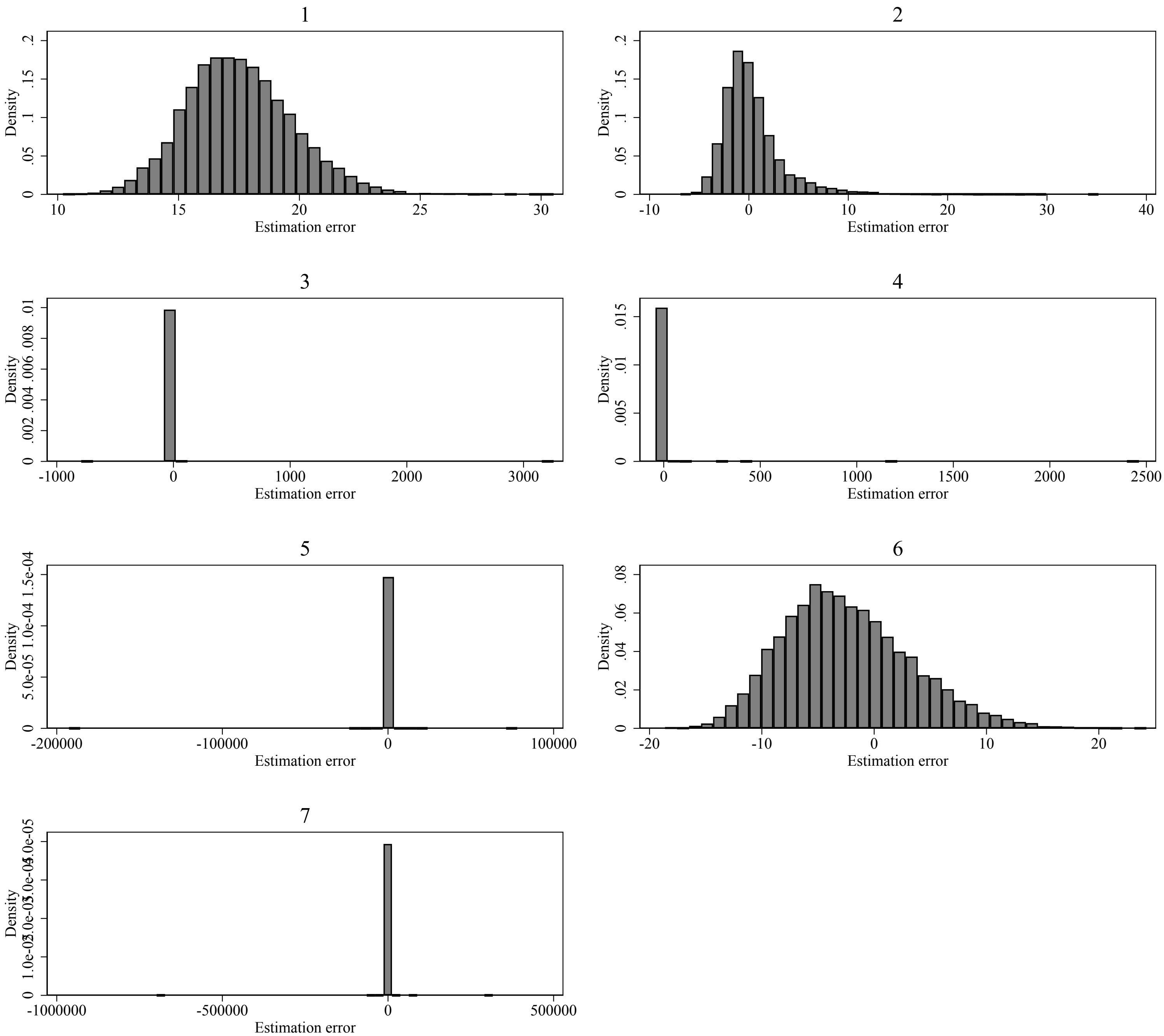
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ (= $\hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.37: Simulation Results for Design D, $\delta = 0.01$, $N = 500$



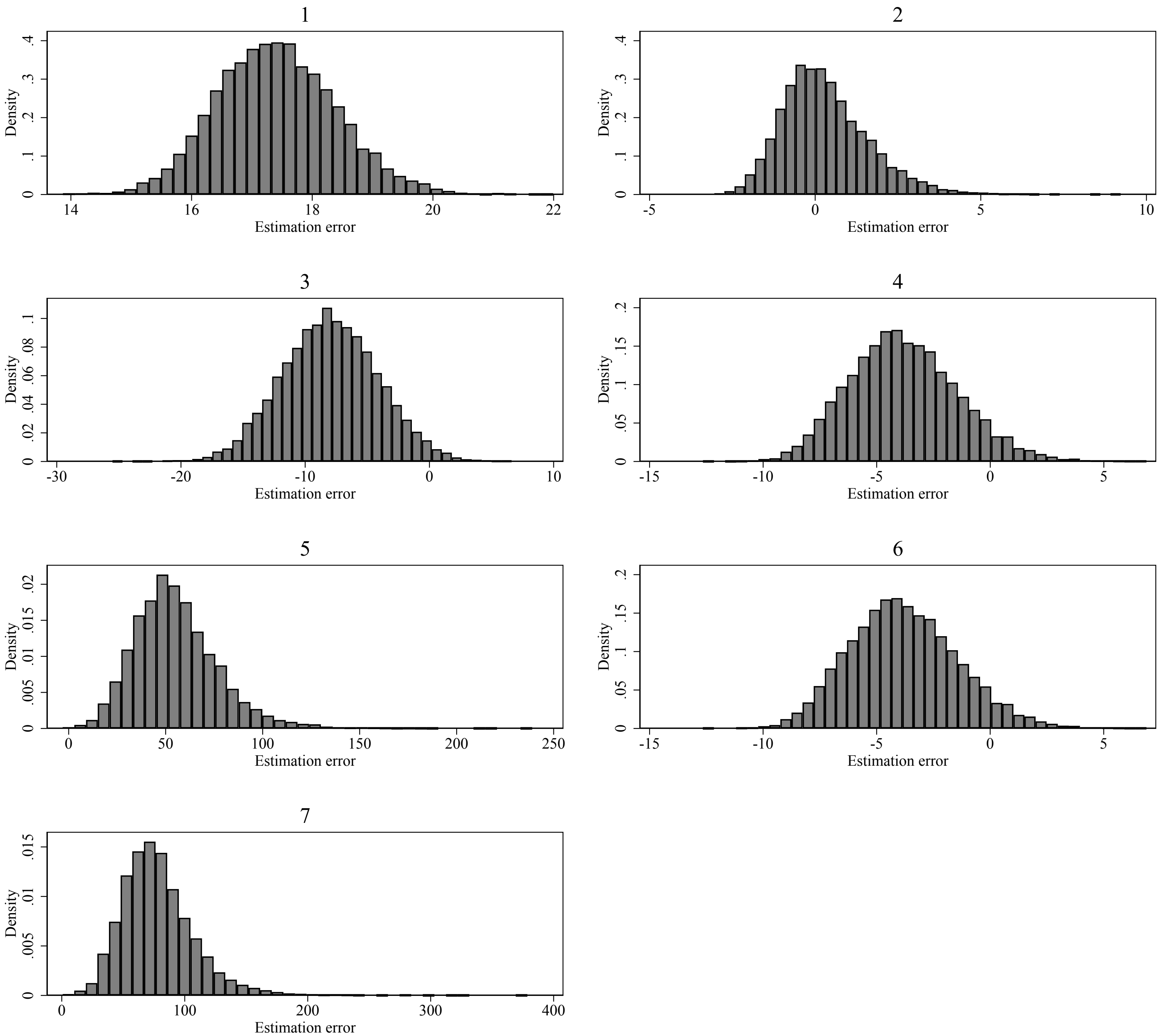
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.38: Simulation Results for Design D, $\delta = 0.01$, $N = 1{,}000$



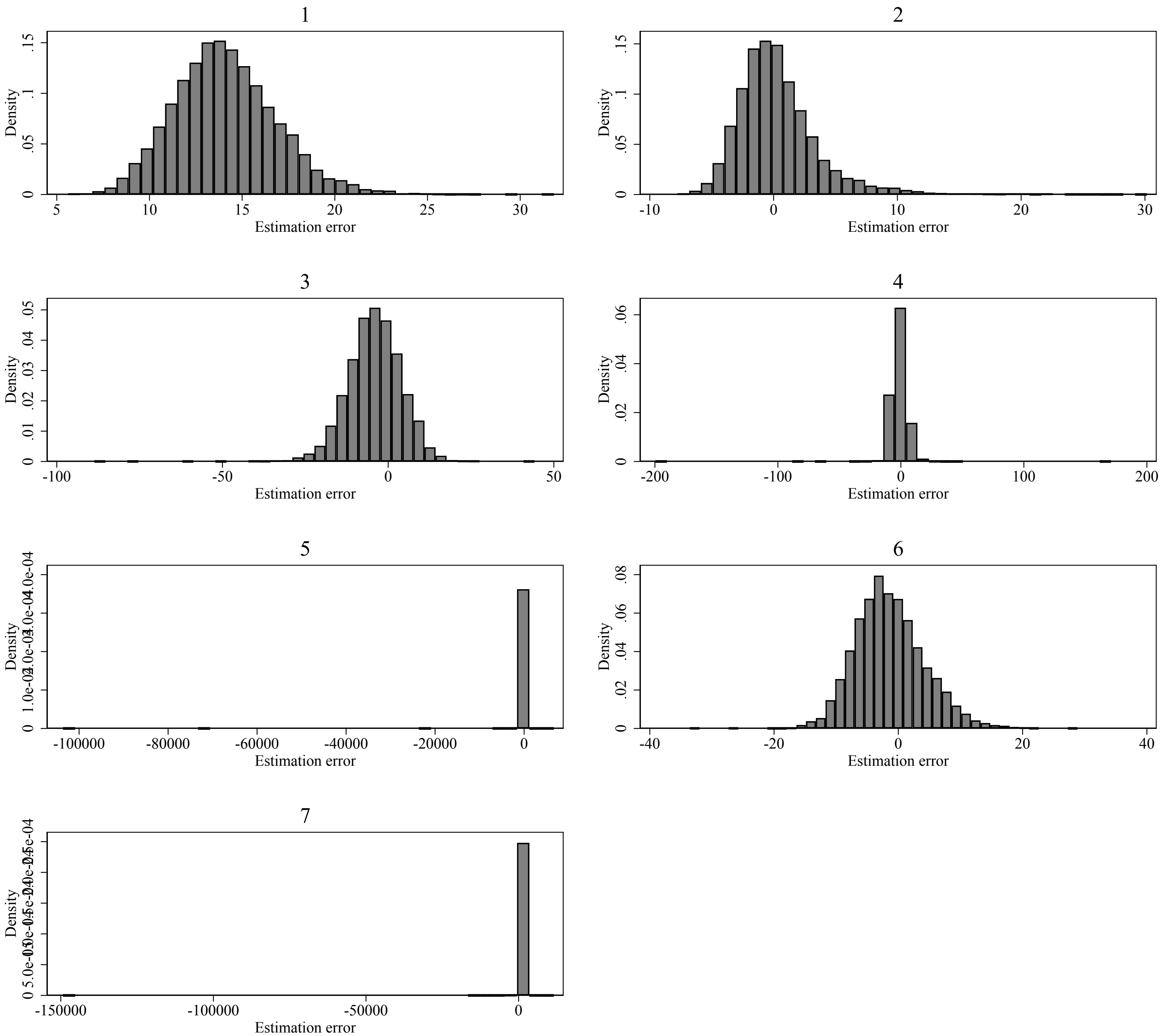
Notes: The details of this simulation design are provided in Section 5 in the main text. “1” corresponds to the 2SLS estimator that additively controls for X. “2” corresponds to $\hat{\tau}_u^{cb}$. “3” corresponds to $\hat{\tau}_u^{ml}$. “4” corresponds to $\hat{\tau}_{a,10}^{ml}$. “5” corresponds to $\hat{\tau}_a^{ml}$. “6” corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). “7” corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.39: Simulation Results for Design D, δ = 0.01, N = 5,000



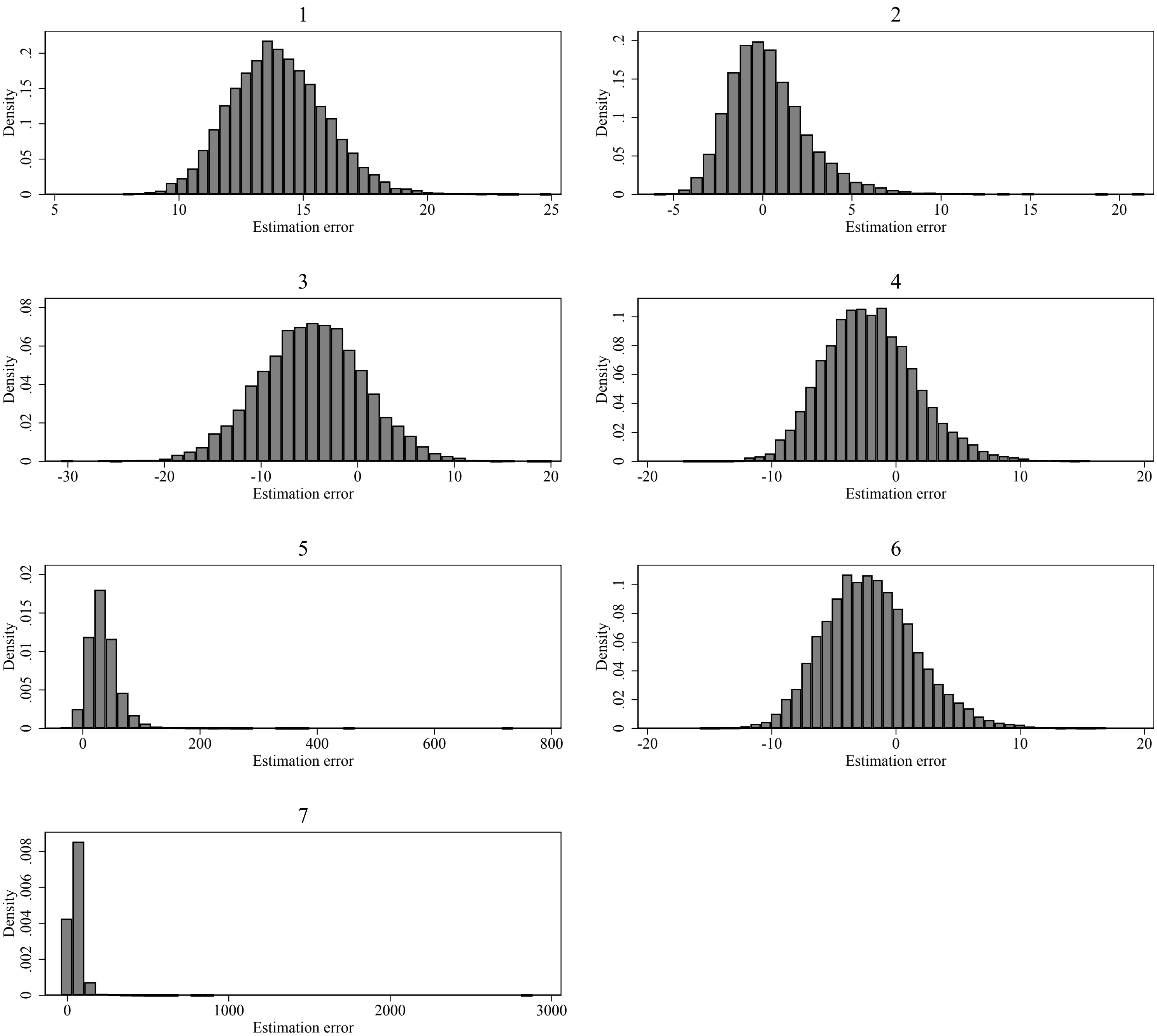
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ (= $\hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.40: Simulation Results for Design D, δ = 0.02, N = 500



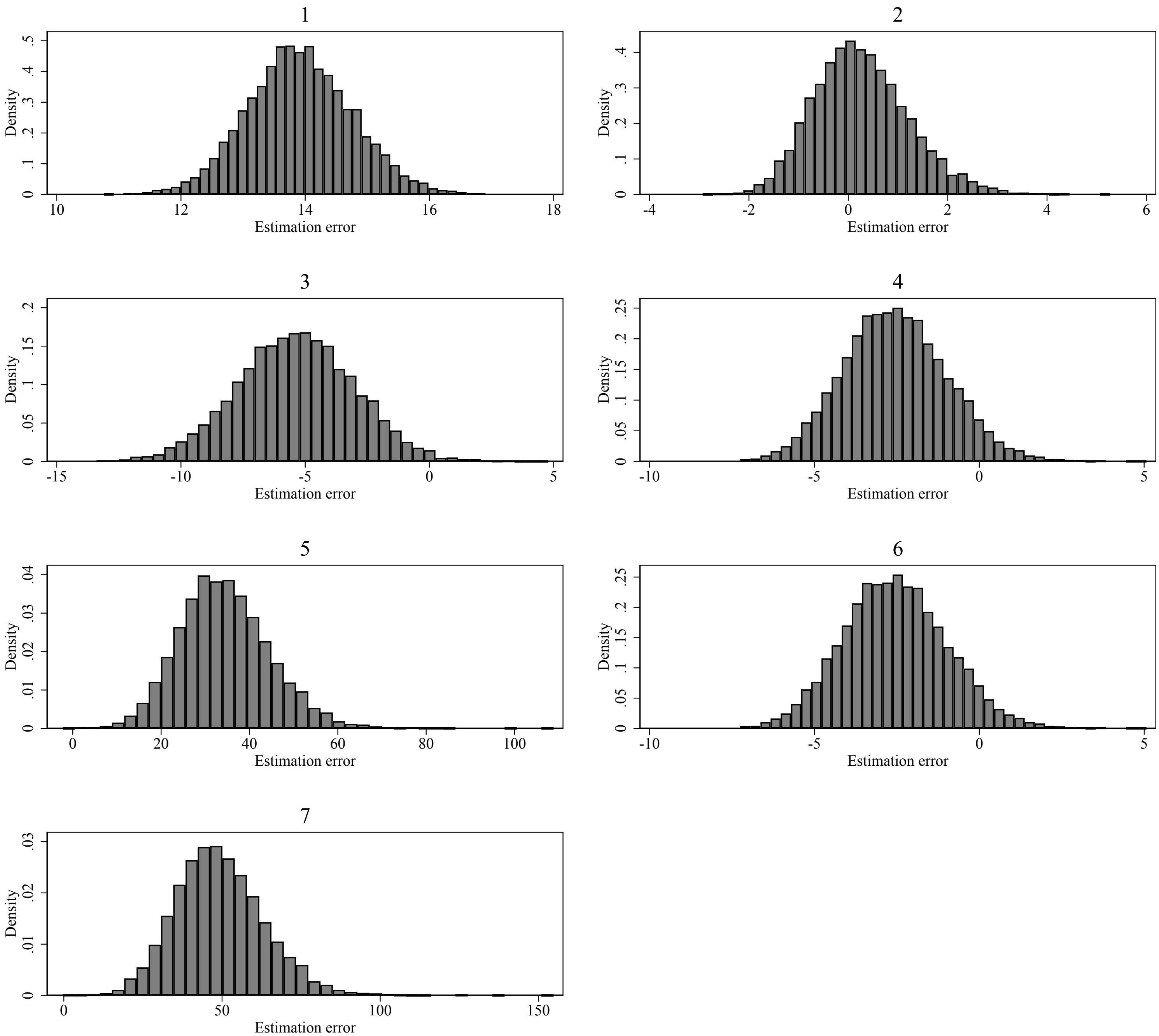
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ (= $\hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.41: Simulation Results for Design D, $\delta = 0.02$, $N = 1{,}000$



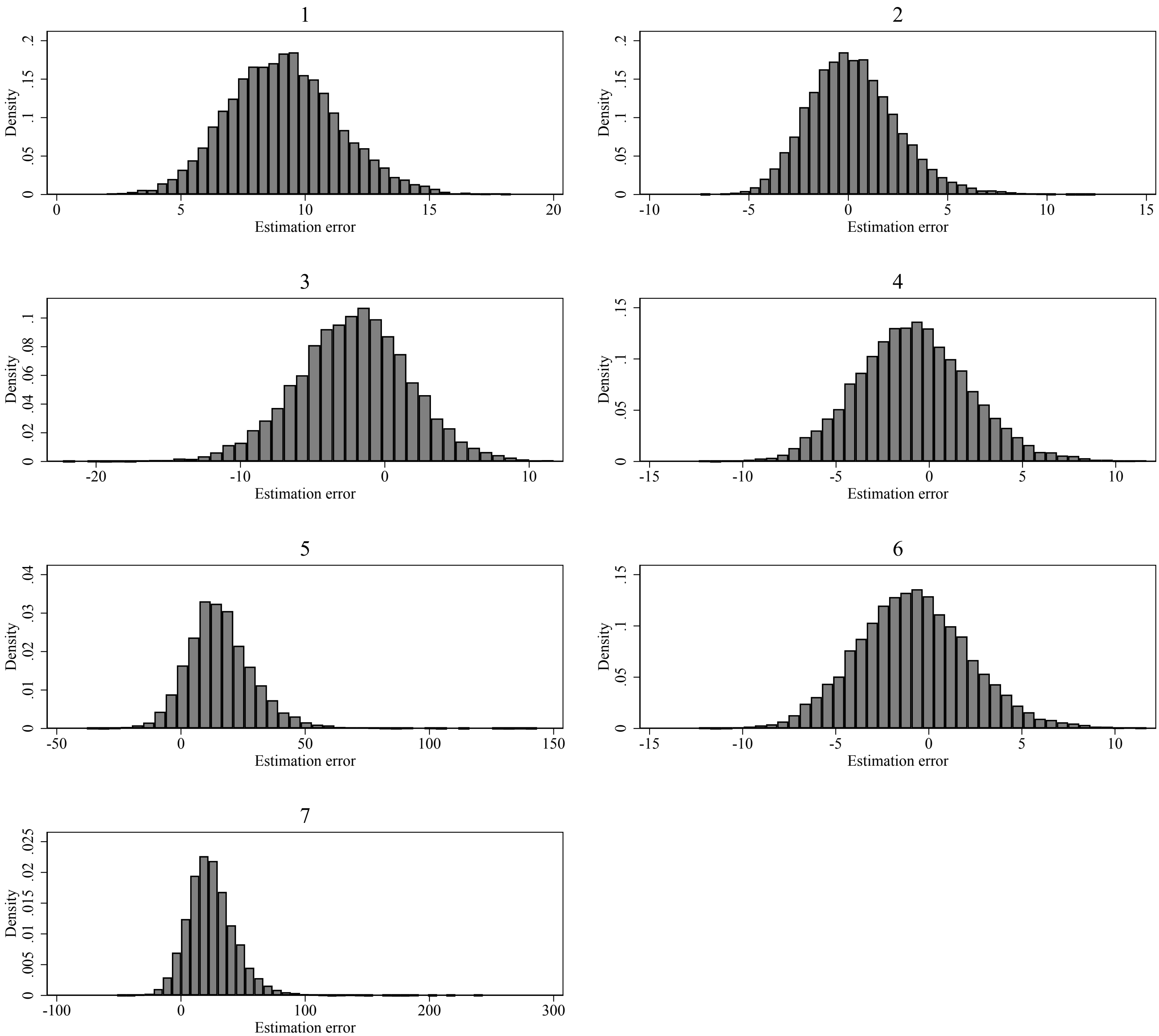
Notes: The details of this simulation design are provided in Section 5 in the main text. “1” corresponds to the 2SLS estimator that additively controls for X. “2” corresponds to $\hat{\tau}_u^{cb}$. “3” corresponds to $\hat{\tau}_u^{ml}$. “4” corresponds to $\hat{\tau}_{a,10}^{ml}$. “5” corresponds to $\hat{\tau}_a^{ml}$. “6” corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). “7” corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.42: Simulation Results for Design D, $\delta = 0.02$, $N = 5{,}000$



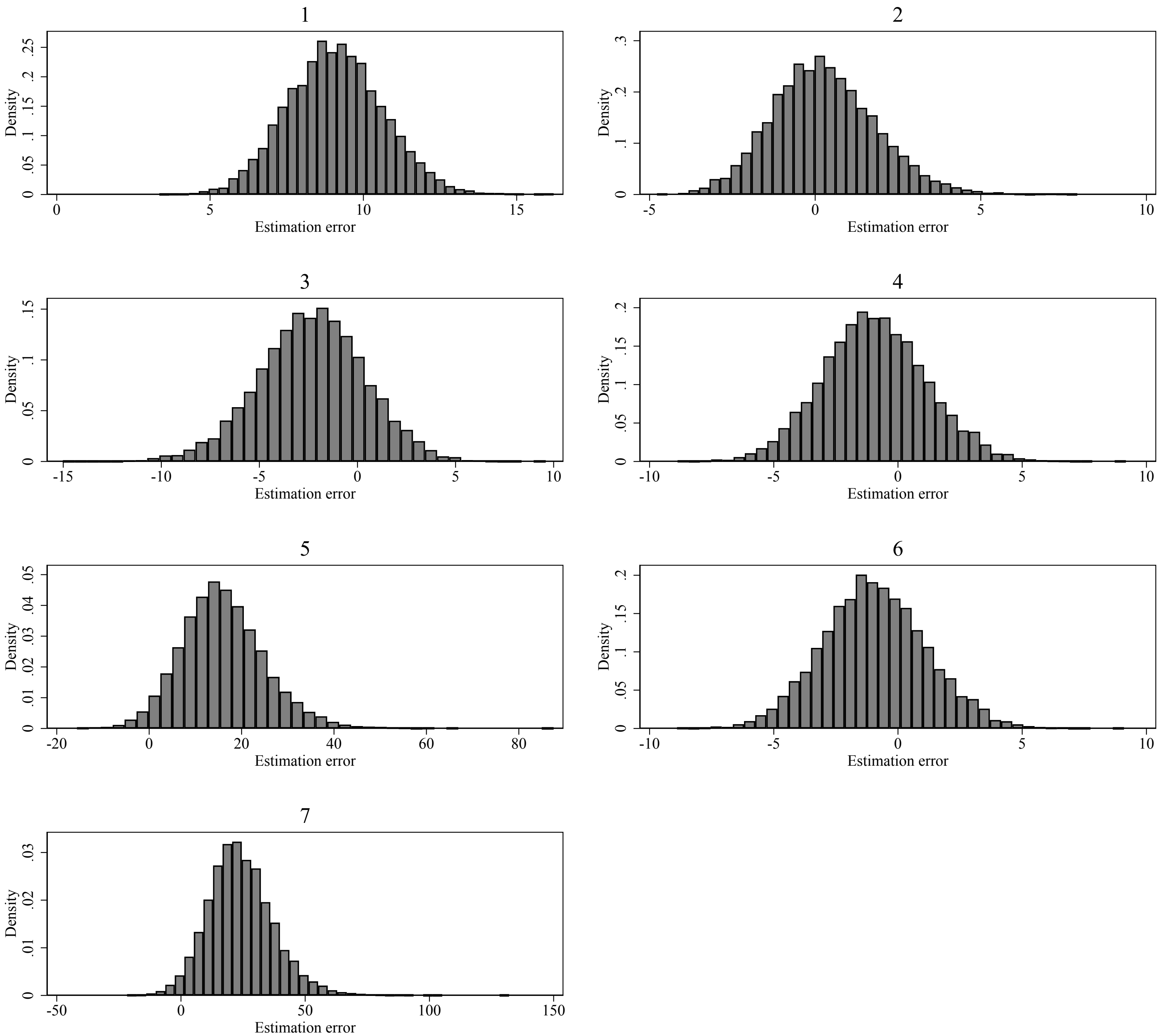
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.43: Simulation Results for Design D, $\delta = 0.05$, $N = 500$



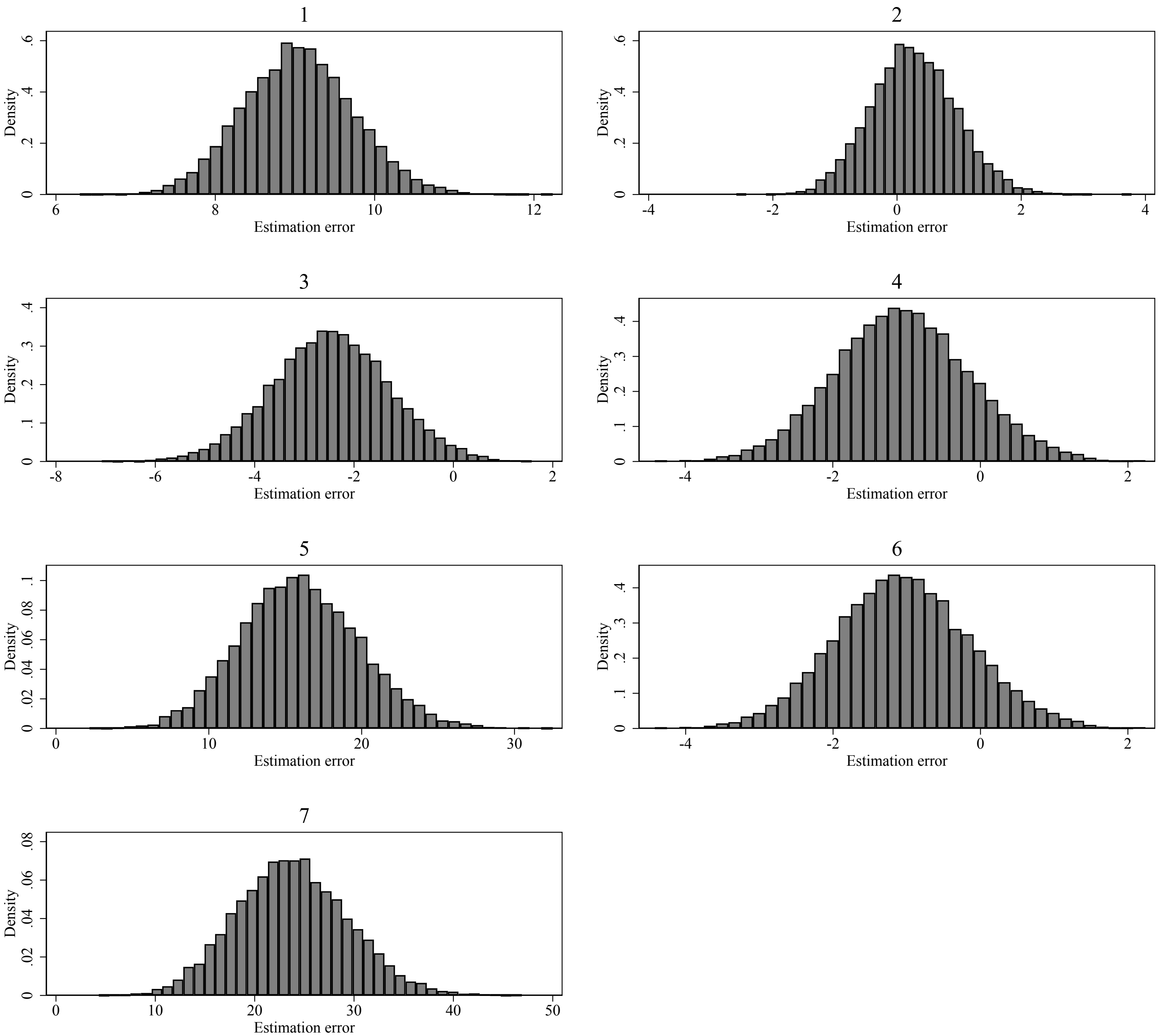
Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.44: Simulation Results for Design D, δ = 0.05, N = 1,000



Notes: The details of this simulation design are provided in Section 5 in the main text. “1” corresponds to the 2SLS estimator that additively controls for X. “2” corresponds to $\hat{\tau}_u^{cb}$. “3” corresponds to $\hat{\tau}_u^{ml}$. “4” corresponds to $\hat{\tau}_{a,10}^{ml}$. “5” corresponds to $\hat{\tau}_a^{ml}$. “6” corresponds to $\hat{\tau}_t^{ml}$ (= $\hat{\tau}_{a,1}^{ml}$). “7” corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

Figure C.45: Simulation Results for Design D, $\delta = 0.05$, $N = 5{,}000$



Notes: The details of this simulation design are provided in Section 5 in the main text. "1" corresponds to the 2SLS estimator that additively controls for X. "2" corresponds to $\hat{\tau}_u^{cb}$. "3" corresponds to $\hat{\tau}_u^{ml}$. "4" corresponds to $\hat{\tau}_{a,10}^{ml}$. "5" corresponds to $\hat{\tau}_a^{ml}$. "6" corresponds to $\hat{\tau}_t^{ml}$ ($= \hat{\tau}_{a,1}^{ml}$). "7" corresponds to $\hat{\tau}_{a,0}^{ml}$. All weighting estimators also control for X. Results are based on 10,000 replications.

References

Abadie, Alberto, "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 2003, *113* (2), 231–263.

Angrist, Joshua D. and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton–Oxford: Princeton University Press, 2009.

Boos, Dennis D. and Leonard A. Stefanski, *Essential Statistical Inference: Theory and Methods*, New York: Springer, 2013.

Cohodes, Sarah R., "The Long-Run Impacts of Specialized Programming for High-Achieving Students," *American Economic Journal: Economic Policy*, 2020, *12* (1), 127–166.

Goodman, Joshua, Oded Gurantz, and Jonathan Smith, "Take Two! SAT Retaking and College Enrollment Gaps," *American Economic Journal: Economic Policy*, 2020, *12* (2), 115–158.

Heiler, Phillip, "Efficient Covariate Balancing for the Local Average Treatment Effect," *Journal of Business & Economic Statistics*, 2022, *40* (4), 1569–1582.

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman, "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment," *Quarterly Journal of Economics*, 2001, *116* (2), 607–654.

Leung, Pauline and Christopher O'Leary, "Unemployment Insurance and Means-Tested Program Interactions: Evidence from Administrative Data," *American Economic Journal: Economic Policy*, 2020, *12* (2), 159–192.

Londoño-Vélez, Juliana, Catherine Rodríguez, and Fabio Sánchez, "Upstream and Downstream Impacts of College Merit-Based Financial Aid for Low-Income Students: Ser Pilo Paga in Colombia," *American Economic Journal: Economic Policy*, 2020, *12* (2), 193–227.

Marx, Benjamin M. and Lesley J. Turner, "Student Loan Nudges: Experimental Evidence on Borrowing and Educational Attainment," *American Economic Journal: Economic Policy*, 2019, *11* (2), 108–141.

Newey, Whitney K. and Daniel McFadden, "Large Sample Estimation and Hypothesis Testing," in Robert Engle and Daniel McFadden, eds., *Handbook of Econometrics*, Vol. 4, Amsterdam: North-Holland, 1994, pp. 2111–2245.

Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., Cambridge–London: MIT Press, 2010.

Young, Alwyn, "Consistency without Inference: Instrumental Variables in Practical Application," *European Economic Review*, 2022, *147*, 104112.