Abadie's Kappa and Weighting Estimators of the Local Average Treatment Effect*

Tymon Słoczyński† S. Derya Uysal‡ Jeffrey M. Wooldridge§

Abstract

In this paper we study the finite sample and asymptotic properties of various weighting estimators of the local average treatment effect (LATE), several of which are based on Abadie's (2003) kappa theorem. Our framework presumes a binary treatment and a binary instrument, which may only be valid after conditioning on additional covariates. We argue that one of the Abadie estimators, which is weight normalized, is preferable in many contexts. Several other estimators, which are unnormalized, do not generally satisfy the properties of scale invariance with respect to the natural logarithm and translation invariance, thereby exhibiting sensitivity to the units of measurement when estimating the LATE in logs and the centering of the outcome variable more generally. On the other hand, when noncompliance is one-sided, certain unnormalized estimators have the advantage of being based on a denominator that is bounded away from zero. To reconcile these findings, we demonstrate that when the instrument propensity score is estimated using an appropriate covariate balancing approach, the resulting normalized estimator also shares this advantage. We use a simulation study and three empirical applications to illustrate our findings. In two cases, the unnormalized estimates are clearly unreasonable, with "incorrect" signs, magnitudes, or both.

*This version: March 24, 2023. For helpful comments, we thank Alberto Abadie, Josh Angrist, Bryan Graham, Phillip Heiler, Toru Kitagawa, Chris Muris, Tomasz Olma, Pedro Sant'Anna, Liyang Sun, seminar participants at Brandeis University, and conference participants at CFE, EEA, ESEM, MEG, NY Camp Econometrics, SEA, and the World Congress of the Econometric Society. Słoczyński acknowledges financial support from the Theodore and Jane Norman Fund. Uysal acknowledges financial support from the German Research Foundation through CRC TRR 190.

†Brandeis University

‡Ludwig Maximilian University of Munich

§Michigan State University

1 Introduction

A large literature following Imbens and Angrist (1994) focuses on identification and estimation of the local average treatment effect (LATE), that is, the average effect of treatment for "compliers," whose treatment status is affected by a binary instrument. In an important contribution to this literature, Abadie (2003) demonstrates how to identify any parameter that is defined in terms of moments of the joint distribution of the data for compliers. The result is based on "kappa weighting," with weights that depend on the instrument propensity score. Abadie's (2003) theorem has been highly influential in applied work, and it is now routinely used to estimate mean covariate values for compliers (e.g., Angrist et al., 2013; Dahl et al., 2014; Bisbee et al., 2017) and to approximate the conditional mean of an outcome variable in this subpopulation (e.g., Angrist, 2001; Cruces and Galiani, 2007; Angrist et al., 2013; Goda et al., 2017). At the same time, Abadie's (2003) result has stimulated a vibrant theoretical literature in econometrics, which focuses on estimating the LATE and its quantile counterparts (e.g., Frölich and Melly, 2013; Abadie and Cattaneo, 2018; Sant'Anna et al., 2022; Singh and Sun, 2022).

There is also an alternative way to construct weighting estimators of the LATE, which follows from the identification result in Frölich (2007). This result implies that the ratio of any consistent estimator of the average treatment effect (ATE) of the instrument on the outcome and any consistent estimator of its ATE on the treatment is consistent for the LATE. A simple approach is to estimate the LATE as the ratio of two particular weighting estimators. Although the recent literature in econometrics and statistics has adopted this approach, it focuses primarily on the ratio of two *unnormalized* estimators (Tan, 2006; Frölich, 2007; MaCurdy et al., 2011; Donald et al., 2014a,b; Abdulkadiroğlu et al., 2017), despite the fact that weighting estimators of the ATE are known to exhibit poor properties in finite samples when they are not normalized, i.e. when their weights do not sum to unity (Imbens, 2004; Millimet and Tchernis, 2009; Busso et al., 2014).

In this paper we provide a comprehensive treatment of both approaches to constructing weighting estimators of the LATE. We begin with an observation that Abadie's (2003) theorem lends itself to constructing a number of consistent estimators of the LATE, only one of which is normalized.

We argue that this estimator, which is different from the normalized version of Tan's (2006) estimator, is likely to dominate the other kappa weighting estimators in most cases. Unlike many other papers that stress the importance of normalization, we also provide an objective and intuitively appealing criterion that differentiates the normalized from the unnormalized estimators. Indeed, we demonstrate that the former, unlike the latter, satisfy the properties of *(i)* translation invariance and *(ii)* scale invariance with respect to the natural logarithm. This ensures that the normalized estimators are not sensitive to the centering of the outcome variable or, when estimating the LATE in logs, to the units of measurement of the untransformed outcome (cf. Chen and Roth, 2022).

Perhaps surprisingly, we also identify an important context, namely settings with one-sided noncompliance, in which certain unnormalized estimators have a major advantage over their normalized counterparts. Indeed, we demonstrate that a particular unnormalized estimator is based on a denominator that is bounded away from zero whenever there are no always-takers, that is, individuals who participate in the treatment regardless of the value of the instrument. Such boundedness is an important property for a ratio estimator (cf. Andrews et al., 2019). Interestingly, we also show that this particular unnormalized estimator is, in fact, identical to Tan's (2006) original weighting estimator. There is also another unnormalized estimator, which has not been studied before and whose denominator is bounded away from zero whenever there are no never-takers, that is, individuals who never participate in the treatment.

Our observations about translation and scale invariance as well as settings with one-sided noncompliance apply equally when the instrument propensity score is known and when it is estimated using maximum likelihood or nonparametrically. These observations make estimator choice potentially difficult, as none of the estimators discussed so far is simultaneously free from both of the problems we identify. To reconcile these findings, we demonstrate that when the instrument propensity score is estimated using an appropriate covariate balancing approach, as in Imai and Ratkovic (2014) and Heiler (2022), among others, the resulting normalized estimator avoids near-zero denominators when there are no always-takers *and also* when there are instead no never-takers. Given that this estimator is normalized, it is also translation invariant and scale invariant

with respect to the natural logarithm. We recommend this estimator for wider use in practice.

Aside from the finite sample properties of weighting estimators of the LATE, we also study their asymptotic properties. In a unified framework of M-estimation, under standard regularity conditions, our weighting estimators are asymptotically normal, and we derive their asymptotic variances. To illustrate our findings, we also use a simulation study and three empirical applications. The simulations confirm the stability of the appropriate unnormalized estimators in settings with one-sided noncompliance. In general, however, it seems advisable to use our preferred normalized estimator based on covariate balancing or at least, if the instrument propensity score is estimated using maximum likelihood, the normalized version of Tan's (2006) estimator.

Our empirical applications focus on causal effects of military service (Angrist, 1990), college education (Card, 1995), and childbearing (Angrist and Evans, 1998). In each of these cases, we document what we regard as superiority of normalized weighting. First, in our replication of Angrist (1990), the unnormalized estimates are highly variable across specifications, which is not the case for the instrumental variables (IV) estimates or normalized weighting. Second, in our replication of Card (1995), the IV estimates are excessively large, which is not the case for the normalized weighting estimates; the unnormalized estimates, on the other hand, are either even larger than the IV estimates or negative, which is unreasonable for causal effects of college education. Finally, in our replication of Angrist and Evans (1998), some of the unnormalized estimates of the effect of childbearing on log wages of mothers are positive, which is also not believable.

2 Framework

Our framework broadly follows Abadie (2003). Let Y denote the outcome variable of interest, D the binary treatment, and Z the binary instrument for D. We also introduce a vector of observed covariates, X, that predict Z. The instrument propensity score is written as $p(X) = \mathrm{P}(Z = 1 \mid X)$.

There are two potential outcomes, Y_1 and Y_0, only one of which is observed for a given individual, $Y = D \cdot Y_1 + (1 - D) \cdot Y_0$. Similarly, there are two potential treatments, D_1 and D_0, and it

is Z that determines which of them is observed, $D = Z \cdot D_1 + (1 - Z) \cdot D_0$. It will also be useful to include Z in the definition of potential outcomes, letting Y_{zd} denote the potential outcome that a given individual would obtain if $Z = z$ and $D = d$.

Angrist et al. (1996) divide the population into four mutually exclusive subgroups based on the latent values of D_1 and D_0. Individuals with $D_1 = D_0 = 1$ are referred to as *always-takers*, as they get treatment regardless of whether they are encouraged to do so or not; similarly, individuals with $D_1 = D_0 = 0$ are referred to as *never-takers*. Individuals with $D_1 = 1$ and $D_0 = 0$ are referred to as *compliers*, as they comply with their instrument assignment; they get treatment if they are encouraged to do so but not otherwise. Analogously, individuals with $D_1 = 0$ and $D_0 = 1$ are referred to as *defiers*, as they defy their instrument assignment.

As usual, we define the treatment effect as the difference in the outcomes with and without treatment, $Y_1 - Y_0$. Following Imbens and Angrist (1994), a large literature has focused on identification and estimation of the local average treatment effect (LATE), defined as

$$\tau_{\text{LATE}} = \mathrm{E}\left(Y_1 - Y_0 \mid D_1 > D_0\right),$$

i.e. as the average treatment effect for compliers or, in other words, for those individuals who would be induced to get treatment by the change in Z from zero to one.

Next, we review a general identification result due to Abadie (2003), which we will use, in turn, to discuss identification of τ_{LATE}. We begin by restating Abadie's (2003) assumptions.

Assumption IV. *(i) Independence of the instrument:* $(Y_{00}, Y_{01}, Y_{10}, Y_{11}, D_0, D_1) \perp Z \mid X$.
(ii) Exclusion of the instrument: $\mathrm{P}(Y_{1d} = Y_{0d} \mid X) = 1$ for $d \in \{0, 1\}$ a.s.
(iii) First stage: $0 < \mathrm{P}(Z = 1 \mid X) < 1$ and $\mathrm{P}(D_1 = 1 \mid X) > \mathrm{P}(D_0 = 1 \mid X)$ a.s.
(iv) Monotonicity: $\mathrm{P}(D_1 \geq D_0 \mid X) = 1$ a.s.

These assumptions are standard in the recent literature. Assumption IV(i) states that, conditional on covariates, the instrument is "as good as randomly assigned." Assumption IV(ii) implies that the instrument only affects the outcome through its effect on treatment status; it follows that $Y_0 = Y_{10} = Y_{00}$ and $Y_1 = Y_{11} = Y_{01}$. Assumption IV(iii) combines an overlap condition with a requirement that

the instrument affects the conditional probability of treatment. Finally, Assumption IV(iv) rules out the existence of defiers, and implies that the population consists of always-takers, never-takers, and compliers. Under Assumption IV, as demonstrated by Abadie (2003), any feature of the joint distribution of (Y, D, X), (Y_0, X), or (Y_1, X) is identified for compliers.

Lemma 2.1 (Abadie 2003, pp. 236–237)**.** *Let $g(\cdot)$ be any measurable real function of (Y, D, X) such that $\mathrm{E}|g(Y, D, X)| < \infty$. Define*

$$\begin{aligned} \kappa_0 &= (1 - D)\,\frac{(1 - Z) - (1 - p(X))}{p(X)\,(1 - p(X))}, \\ \kappa_1 &= D\frac{Z - p(X)}{p(X)\,(1 - p(X))}, \\ \kappa = \kappa_0\,(1 - p(X)) + \kappa_1 p(X) &= 1 - \frac{D\,(1 - Z)}{1 - p(X)} - \frac{(1 - D)\,Z}{p(X)}. \end{aligned}$$

Under Assumption IV,

(a) $\mathrm{E}\left[g(Y, D, X) \mid D_1 > D_0\right] = \frac{1}{\mathrm{P}(D_1 > D_0)}\mathrm{E}\left[\kappa\; g(Y, D, X)\right]$. *Also,*

(b) $\mathrm{E}\left[g(Y_0, X) \mid D_1 > D_0\right] = \frac{1}{\mathrm{P}(D_1 > D_0)}\mathrm{E}\left[\kappa_0\; g(Y, X)\right]$, *and*

(c) $\mathrm{E}\left[g(Y_1, X) \mid D_1 > D_0\right] = \frac{1}{\mathrm{P}(D_1 > D_0)}\mathrm{E}\left[\kappa_1\; g(Y, X)\right]$.

Moreover, (a–c) *also hold conditional on X.*

Both Abadie (2003) and the subsequent applied literature have focused on the implications of Lemma 2.1(a). Indeed, numerous papers have used this result to estimate mean covariate values for compliers (e.g., Angrist et al., 2013; Dahl et al., 2014; Bisbee et al., 2017) and to approximate the conditional mean of Y in this subpopulation (e.g., Angrist, 2001; Cruces and Galiani, 2007; Angrist et al., 2013; Goda et al., 2017). On the other hand, Lemma 2.1(b) and (c) have been used to identify and estimate τ_{LATE} and quantile treatment effects (e.g., Frölich and Melly, 2013; Abadie and Cattaneo, 2018; Sant'Anna et al., 2022; Singh and Sun, 2022).

To see how Lemma 2.1(b) and (c) identifies τ_{LATE}, take $g(Y_0, X) = Y_0$ and $g(Y_1, X) = Y_1$, and write:

$$\tau_{\mathrm{LATE}} = \frac{1}{\mathrm{P}(D_1 > D_0)}\mathrm{E}\left(\kappa_1 Y\right) - \frac{1}{\mathrm{P}(D_1 > D_0)}\mathrm{E}\left(\kappa_0 Y\right). \tag{1}$$

We can also rewrite equation (1) to obtain the following expression for τ_{LATE}:

$$\tau_{\text{LATE}} = \frac{1}{\text{P}(D_1 > D_0)}\text{E}\left[(\kappa_1 - \kappa_0)\, Y\right] = \frac{1}{\text{P}(D_1 > D_0)}\text{E}\left[Y\frac{Z - p(X)}{p(X)\,(1 - p(X))}\right]. \tag{2}$$

As we will see later, it is useful to treat equations (1) and (2) as distinct. In any case, it is clear that τ_{LATE} is identified as long as $\text{P}(D_1 > D_0)$ is identified. As noted by Abadie (2003), Lemma 2.1(a) implies that $\text{P}(D_1 > D_0) = \text{E}(\kappa)$, which follows from taking $g(Y, D, X) = 1$. Similarly, however, we can use Lemma 2.1(b) and (c) to obtain $\text{P}(D_1 > D_0) = \text{E}(\kappa_1)$ and $\text{P}(D_1 > D_0) = \text{E}(\kappa_0)$. This is not a novel observation but we will provide a more comprehensive discussion of its consequences than has been done in previous work. We conclude this section with the following remark.

Remark 2.2. $\text{E}(\kappa) = \text{E}(\kappa_1) - \text{E}\left[\frac{Z-p(X)}{p(X)}\right] = \text{E}(\kappa_1) = \text{E}(\kappa_1) - \text{E}\left[\frac{Z-p(X)}{p(X)(1-p(X))}\right] = \text{E}(\kappa_0)$.

The proof of Remark 2.2 follows from simple algebra and is omitted. The facts that $\text{E}\left[\frac{Z-p(X)}{p(X)}\right] = 0$ and $\text{E}\left[\frac{Z-p(X)}{p(X)(1-p(X))}\right] = 0$ hold by iterated expectations. It turns out that $\text{E}(\kappa) = \text{E}(\kappa_1) = \text{E}(\kappa_0)$. Additionally, Lemma 2.1 implies that each of these objects identifies $\text{P}(D_1 > D_0)$.

3 Estimation and Inference

In this section we study estimation and inference for τ_{LATE}. We begin with the case where $p(X)$ is known. While this is often not true in practice, our observations in Sections 3.2 and 3.3 apply equally in that case and when $p(X)$ is estimated using maximum likelihood or nonparametrically.

3.1 Estimation When the Instrument Propensity Score Is Known

Given a random sample $\{(D_i, Z_i, X_i, Y_i) : i = 1, \ldots, N\}$, and assuming that the instrument propensity score is known, equation (2) suggests that we can consistently estimate τ_{LATE} as follows:

$$\hat{\tau}_{\text{LATE}} = \frac{1}{\hat{\text{P}}(D_1 > D_0)}\left[N^{-1}\sum_{i=1}^{N} Y_i\frac{Z_i - p(X_i)}{p(X_i)\,(1 - p(X_i))}\right],$$

where $\hat{\mathrm{P}}(D_1 > D_0) \xrightarrow{p} \mathrm{P}(D_1 > D_0) > 0$. Our discussion in Section 2 also implies that there are at least three candidate estimators for $\mathrm{P}(D_1 > D_0)$, namely $N^{-1}\sum_{i=1}^{N}\kappa_i$, $N^{-1}\sum_{i=1}^{N}\kappa_{i1}$, and $N^{-1}\sum_{i=1}^{N}\kappa_{i0}$, where $\kappa_i = 1 - \frac{D_i(1-Z_i)}{1-p(X_i)} - \frac{(1-D_i)Z_i}{p(X_i)}$, $\kappa_{i1} = D_i\frac{Z_i - p(X_i)}{p(X_i)(1-p(X_i))}$, and $\kappa_{i0} = (1-D_i)\frac{(1-Z_i)-(1-p(X_i))}{p(X_i)(1-p(X_i))}$. Consequently, we have the following consistent estimators of τ_{LATE}:

$$\hat{\tau}_a = \left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N} Y_i\frac{Z_i - p(X_i)}{p(X_i)\,(1-p(X_i))}\right], \tag{3}$$

$$\hat{\tau}_{a,1} = \left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N} Y_i\frac{Z_i - p(X_i)}{p(X_i)\,(1-p(X_i))}\right], \tag{4}$$

$$\hat{\tau}_{a,0} = \left[\sum_{i=1}^{N}\kappa_{i0}\right]^{-1}\left[\sum_{i=1}^{N} Y_i\frac{Z_i - p(X_i)}{p(X_i)\,(1-p(X_i))}\right]. \tag{5}$$

One might (mistakenly, as it turns out) expect that the choice of the estimator for $\mathrm{P}(D_1 > D_0)$ is largely inconsequential. We discuss this issue extensively in what follows. For now, it should suffice to note that $N^{-1}\sum_{i=1}^{N}\frac{Z_i - p(X_i)}{p(X_i)}$ and $N^{-1}\sum_{i=1}^{N}\frac{Z_i - p(X_i)}{p(X_i)(1-p(X_i))}$ are not generally equal to zero or to each other, and hence $N^{-1}\sum_{i=1}^{N}\kappa_i$, $N^{-1}\sum_{i=1}^{N}\kappa_{i1}$, and $N^{-1}\sum_{i=1}^{N}\kappa_{i0}$ will also generally be different, unlike their population counterparts (cf. Remark 2.2).

Lemma 2.1 is not the only identification result that allows us to construct consistent estimators of the LATE. An alternative result is provided by Frölich (2007, Theorem 1). An implication of this result is that the ratio of any consistent estimator of the average treatment effect (ATE) of Z on Y and any consistent estimator of the ATE of Z on D is consistent for the LATE. Given our interest in weighting estimators, a natural candidate estimator is

$$\hat{\tau}_t = \left[\sum_{i=1}^{N}\frac{D_iZ_i}{p(X_i)} - \sum_{i=1}^{N}\frac{D_i\,(1-Z_i)}{1-p(X_i)}\right]^{-1}\left[\sum_{i=1}^{N}\frac{Y_iZ_i}{p(X_i)} - \sum_{i=1}^{N}\frac{Y_i\,(1-Z_i)}{1-p(X_i)}\right], \tag{6}$$

which was first suggested by Tan (2006). This estimator is equal to the ratio of two weighting estimators of the ATE of Z (on Y and D) under unconfoundedness (see, e.g., Hirano et al., 2003). The following remark, which has not been precisely stated in previous work, clarifies the relationship between $\hat{\tau}_t$ and the Abadie estimators introduced above.

Remark 3.1. $\hat{\tau}_t = \hat{\tau}_{a,1}$.

Remark 3.1 states that $\hat{\tau}_t$ and $\hat{\tau}_{a,1}$ are numerically identical, which can be seen by plugging in the expression for κ_{i1} into equation (4):

$$\hat{\tau}_{a,1} = \left[\sum_{i=1}^{N} D_i \frac{Z_i - p(X_i)}{p(X_i)\,(1 - p(X_i))}\right]^{-1} \left[\sum_{i=1}^{N} Y_i \frac{Z_i - p(X_i)}{p(X_i)\,(1 - p(X_i))}\right]. \tag{7}$$

As is easy to see, expressions (6) and (7) are equivalent. It is also important to note that $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$), or at least its variant where $p(X)$ is estimated, is by far the most popular weighting estimator of the LATE in the econometrics literature. It has been considered by Tan (2006), Frölich (2007), MaCurdy et al. (2011), Donald et al. (2014a,b), and Abdulkadiroğlu et al. (2017), among others. As we will see in the next section, however, this estimator has a major drawback in practice.

3.2 Unnormalized and Normalized Weights

Following Imbens (2004), Millimet and Tchernis (2009), and Busso et al. (2014), it is widely understood that weighting estimators of the ATE under unconfoundedness should be normalized, i.e. their weights should sum to unity, an idea that is often attributed to Hájek (1971). More recently, Khan and Ugander (2021) have provided a general treatment of normalization under unconfoundedness while Sant'Anna and Zhao (2020) and Callaway and Sant'Anna (2021) have stressed the importance of normalization in difference-in-differences methods. It is natural to expect that normalization will also be important when estimating the LATE (cf. Heiler, 2022).

It follows immediately that $\hat{\tau}_t$ is likely inferior to the ratio of two normalized, Hájek-type estimators of the ATE of Z under unconfoundedness:

$$\hat{\tau}_{t,norm} = \frac{\left[\sum_{i=1}^{N} \frac{Z_i}{p(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{Y_i Z_i}{p(X_i)} - \left[\sum_{i=1}^{N} \frac{1-Z_i}{1-p(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{Y_i(1-Z_i)}{1-p(X_i)}}{\left[\sum_{i=1}^{N} \frac{Z_i}{p(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{D_i Z_i}{p(X_i)} - \left[\sum_{i=1}^{N} \frac{1-Z_i}{1-p(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{D_i(1-Z_i)}{1-p(X_i)}}.$$

This estimator was proposed by Uysal (2011) and later applied by Bodory and Huber (2018) and Heiler (2022). It might not be immediately obvious how the importance of normalization affects our understanding of the Abadie estimators. To see this, note that $\hat{\tau}_a$, $\hat{\tau}_{a,1}$, and $\hat{\tau}_{a,0}$ can equivalently

be represented as sample analogues of equation (1):

$$\begin{aligned}
\hat{\tau}_a &= \left[\sum_{i=1}^{N} \kappa_i\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} Y_i\right] - \left[\sum_{i=1}^{N} \kappa_i\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} Y_i\right], \\
\hat{\tau}_{a,1} &= \left[\sum_{i=1}^{N} \kappa_{i1}\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} Y_i\right] - \left[\sum_{i=1}^{N} \kappa_{i1}\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} Y_i\right], \\
\hat{\tau}_{a,0} &= \left[\sum_{i=1}^{N} \kappa_{i0}\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} Y_i\right] - \left[\sum_{i=1}^{N} \kappa_{i0}\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} Y_i\right].
\end{aligned}$$

It turns out that none of these estimators is normalized. First, $\hat{\tau}_a$ uses weights of $\left[\sum_{i=1}^{N} \kappa_i\right]^{-1} \kappa_{i1}$ and $\left[\sum_{i=1}^{N} \kappa_i\right]^{-1} \kappa_{i0}$, which do not necessarily sum to unity across i. Second, $\hat{\tau}_{a,1}$ is based on weights of $\left[\sum_{i=1}^{N} \kappa_{i1}\right]^{-1} \kappa_{i1}$, which are properly normalized, and $\left[\sum_{i=1}^{N} \kappa_{i1}\right]^{-1} \kappa_{i0}$, which are not. Finally, $\hat{\tau}_{a,0}$ uses weights of $\left[\sum_{i=1}^{N} \kappa_{i0}\right]^{-1} \kappa_{i1}$, which do not necessarily sum to unity across i, and $\left[\sum_{i=1}^{N} \kappa_{i0}\right]^{-1} \kappa_{i0}$, which are properly normalized.

It is straightforward to construct a normalized Abadie estimator of the LATE. It turns out that the two denominators in equation (1) need to be estimated separately, using different estimators of $\mathrm{P}(D_1 > D_0)$, $N^{-1} \sum_{i=1}^{N} \kappa_{i1}$ and $N^{-1} \sum_{i=1}^{N} \kappa_{i0}$. The resulting estimator becomes

$$\hat{\tau}_{a,10} = \left[\sum_{i=1}^{N} \kappa_{i1}\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} Y_i\right] - \left[\sum_{i=1}^{N} \kappa_{i0}\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} Y_i\right],$$

where both sets of weights, $\left[\sum_{i=1}^{N} \kappa_{i1}\right]^{-1} \kappa_{i1}$ and $\left[\sum_{i=1}^{N} \kappa_{i0}\right]^{-1} \kappa_{i0}$, necessarily sum to unity across i. The normalized Abadie estimator has also been considered by Abadie and Cattaneo (2018) and Sant'Anna et al. (2022). While the literature on quantile treatment effects studies normalized Abadie estimators somewhat more often (see, e.g., Frölich and Melly, 2013), the importance of normalization is not explicitly recognized. Interestingly, if the goal is to estimate $\mathrm{E}(X \mid D_1 > D_0)$ rather than τ_{LATE} or quantile treatment effects, as in Angrist et al. (2013), Dahl et al. (2014), and Bisbee et al. (2017), among others, then three normalized estimators of this object can readily be constructed: $\left[\sum_{i=1}^{N} \kappa_i\right]^{-1} \sum_{i=1}^{N} \kappa_i X_i$, $\left[\sum_{i=1}^{N} \kappa_{i0}\right]^{-1} \sum_{i=1}^{N} \kappa_{i0} X_i$, and $\left[\sum_{i=1}^{N} \kappa_{i1}\right]^{-1} \sum_{i=1}^{N} \kappa_{i1} X_i$.

So far, we have made it seem obvious that weighting estimators should be normalized. Yet, it is natural to ask: *Why* is it so important that weights sum to unity? Many of the recommendations

to date are based on simulation results (e.g., Millimet and Tchernis, 2009; Busso et al., 2014), and it is not clear to what extent such evidence should guide estimator choice (cf. Advani et al., 2019). In what follows, we provide an objective and intuitively appealing criterion that differentiates the normalized from the unnormalized estimators.

To present our criterion, we need to introduce some additional notation. Let $\mathbf{Y}$ be a vector of observed data on outcomes and $\mathbf{W} = (\mathbf{D}\ \mathbf{Z}\ \mathbf{X})$ be a matrix of observed data on the remaining variables. We postulate that any reasonable estimator of τ_{LATE} must be translation invariant.

Definition TI (Translation Invariance)**.** We say that an estimator $\hat{\tau} = \hat{\tau}(\mathbf{Y}, \mathbf{W})$ is translation invariant if $\hat{\tau}(\mathbf{Y}, \mathbf{W}) = \hat{\tau}(\mathbf{Y} + k, \mathbf{W})$ for all $\mathbf{Y}$, $\mathbf{W}$, and k.

The property of translation invariance is defined as the invariance of an estimator to an additive change of the outcome values for all units by a fixed amount.[1] Put differently, estimators that are not translation invariant will generally depend on how the outcome variable is centered. If this variable is binary, the estimate may change when we relabel the zeros and ones, on top of the obvious sign change that is due to relabeling. If the outcome is a logarithm of some other variable, the estimator is also not invariant to scale transformations of that variable.

Definition SI (Scale Invariance)**.** We say that an estimator $\hat{\tau} = \hat{\tau}(\mathbf{Y}, \mathbf{W})$ is scale invariant with respect to g if $\hat{\tau}(f(\mathbf{Y}), \mathbf{W}) = \hat{\tau}(f(a\mathbf{Y}), \mathbf{W})$, $f(\mathbf{Y}) = (g(Y_1), \ldots, g(Y_N))$, for all $\mathbf{Y} > 0$, $\mathbf{W}$, and $a > 0$.

The property of scale invariance is defined as the invariance of an estimator that uses transformed outcome data to a multiplicative change of the outcome values for all units by a fixed amount. This property is tied to the transformation g, with the leading case of the natural logarithm, as in Chen and Roth (2022). To be clear, the idea here is as follows: the researcher transforms the outcome data prior to analysis, perhaps because they want to interpret the estimates as percentages, in which case they would use $g(Y) = \log(Y)$; however, if their estimator is not scale invariant with respect

[1] This property is also referred to as location invariance or shift invariance. It has been considered in several subfields of econometrics. Foster and Shorrocks (1991) and Zheng (1994) advocate for poverty indices that are translation invariant. Aronow and Middleton (2013) note that, under unconfoundedness, the usual (unnormalized) weighting estimator is not translation invariant. Olma (2021) discusses translation invariance in the context of nonparametric estimation of truncated conditional expectation functions. Del Bono et al. (2022) analyze a model of latent skill formation and endorse estimators that are translation invariant.

to the natural logarithm, the resulting estimates will depend on the units of Y, which directly contradicts the idea of interpreting them as percentages.

The following result demonstrates that the unnormalized weighting estimators discussed so far are not translation invariant and not scale invariant with respect to $g(Y) = \log(Y)$. On the other hand, the normalized estimators, $\hat{\tau}_{t,norm}$ and $\hat{\tau}_{a,10}$, satisfy both properties. (In practice, because $\log(ab) = \log(a) + \log(b)$, we expect the two properties to be equivalent.)

Proposition 3.2. *$\hat{\tau}_{t,norm}$ and $\hat{\tau}_{a,10}$ are translation invariant and scale invariant with respect to the natural logarithm. $\hat{\tau}_a$, $\hat{\tau}_t$ ($= \hat{\tau}_{a,1}$), and $\hat{\tau}_{a,0}$ are not translation invariant and not scale invariant with respect to the natural logarithm.*

Proof. See Appendix. □

The properties of translation and scale invariance are very appealing, and it makes intuitive sense to only use estimators that satisfy them. To conclude this section, we make two final observations. First, the point of Proposition 3.2 is similar but distinct from that of Chen and Roth (2022), who focus on the sensitivity to scaling of $\log(1 + Y)$ and similar transformations, and do not restrict their attention to any specific estimators (including weighting). Unlike in Chen and Roth (2022), the problem we describe disappears in large samples. On the other hand, the problem described by Chen and Roth (2022) disappears when the outcome only assumes strictly positive values, which is not the case in Proposition 3.2. Second, it is important to note that scale invariance is not always a valuable property with respect to other transformations. For example, when g is the identity function, we would *not* require that $\hat{\tau}(\mathbf{Y}, \mathbf{W}) = \hat{\tau}(a\mathbf{Y}, \mathbf{W})$. It would instead be desirable to have $\hat{\tau}(\mathbf{Y}, \mathbf{W}) = a^{-1}\hat{\tau}(a\mathbf{Y}, \mathbf{W})$, and this property is satisfied by every estimator that we consider, including the unnormalized estimators.

3.3 Near-Zero Denominators

Weighting estimators of the LATE, like two-stage least squares and many other IV methods, are an example of ratio estimators. A common problem with such estimators is that they behave badly

if their denominator is close to zero (cf. Andrews et al., 2019).

Even though Section 3.2 clearly justifies the preference for normalized weighting estimators, in this section we identify two situations under which certain *unnormalized* estimators have an important advantage: they are based on a denominator that is nonnegative by construction and bounded away from zero in all practically relevant situations. To see this, note that Table 1 provides simplified formulas for κ, κ_1, and κ_0 in each of the four subpopulations defined by their values of Z and D. For example, $\kappa = 1$ if $Z = 1$ and $D = 1$ or $Z = 0$ and $D = 0$; moreover, $\kappa = -\frac{1-p(X)}{p(X)}$ if $Z = 1$ and $D = 0$, and $\kappa = -\frac{p(X)}{1-p(X)}$ if $Z = 0$ and $D = 1$. It follows that $N^{-1}\sum_{i=1}^{N}\kappa_i$ is the mean of a collection of positive and negative values, and hence it can be positive, negative, or zero. This is despite the fact that $N^{-1}\sum_{i=1}^{N}\kappa_i$ is also a consistent estimator of the proportion of compliers, which is strictly positive under Assumption IV. Similarly, $N^{-1}\sum_{i=1}^{N}\kappa_{i1}$ and $N^{-1}\sum_{i=1}^{N}\kappa_{i0}$ are also not guaranteed to be positive or bounded away from zero.

The situation turns out to be different in settings with one-sided noncompliance, i.e. when individuals with $Z = 1$ or individuals with $Z = 0$ fully comply with their instrument assignment. If all individuals with $Z = 1$ get treatment or, equivalently, there are no never-takers, then the second row of Table 1 is empty and $\mathrm{P}(\kappa_0 \geq 0) = 1$. This is the case, for example, in studies that use twin births as an instrument for fertility (e.g., Angrist and Evans, 1998). Similarly, if there are no always-takers, then $\mathrm{P}(\kappa_1 \geq 0) = 1$. This is the case, for example, in randomized trials with noncompliance that make it impossible to access treatment if not offered. An implication of these observations is that in settings with one-sided noncompliance there exist estimators of $\mathrm{P}(D_1 > D_0)$, and perhaps also the LATE, that have some desirable properties in finite samples.

Remark 3.3. If there are no always-takers, $N^{-1}\sum_{i=1}^{N}\kappa_{i1} > \hat{\mathrm{P}}(D = 1) > 0$.

Remark 3.4. If there are no never-takers, $N^{-1}\sum_{i=1}^{N}\kappa_{i0} > \hat{\mathrm{P}}(D = 0) > 0$.

Proof. To prove Remark 3.3, note that $\frac{1}{p(X)} > 1$ by Assumption IV(iii). If there are no always-takers, then $\mathrm{P}(Z = 0, D = 1) = 0$. Thus, $N^{-1}\sum_{i=1}^{N}\kappa_{i1} > N^{-1}\left(\underbrace{1+1+\cdots+1}_{N\cdot\hat{\mathrm{P}}(D=1)} + \underbrace{0+0+\cdots+0}_{N\cdot\hat{\mathrm{P}}(D=0)}\right) = \hat{\mathrm{P}}(D = 1)$. The proof of Remark 3.4 is analogous. □

Remarks 3.3 and 3.4 demonstrate that settings with one-sided noncompliance offer a choice of estimators of $\mathrm{P}(D_1 > D_0)$ that are bounded from below by the sample proportion of treated or untreated units. Note that this property preserves a particular logical consistency of these estimators. If there are no always-takers and no defiers, every treated individual must be a complier. Similarly, every untreated individual must be a complier if there are no never-takers and no defiers.

An implication of Remarks 3.3 and 3.4 is that certain unnormalized estimators have the advantage of avoiding near-zero denominators in settings with one-sided noncompliance. If there are no always-takers or never-takers, we expect $\hat{\tau}_{a,1}$ and $\hat{\tau}_{a,0}$, respectively, to perform relatively well in finite samples. Whether or not this dominates the disadvantage that these estimators are unnormalized is an empirical issue. Note, however, that if $N^{-1}\sum_{i=1}^{N}\kappa_{i1}$ is away from zero but $N^{-1}\sum_{i=1}^{N}\kappa_{i0}$ is not, then this will negatively affect the performance of not only $\hat{\tau}_{a,0}$ but also $\hat{\tau}_{a,10}$. Likewise, if $N^{-1}\sum_{i=1}^{N}\kappa_{i1}$ is close to zero, then both $\hat{\tau}_{a,1}$ and $\hat{\tau}_{a,10}$ will be affected.

Additionally, if the goal is to estimate $\mathrm{E}\left(X \mid D_1 > D_0\right)$ and noncompliance is one-sided, it becomes easier to choose between the three *normalized* estimators in Section 3.2, $\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\sum_{i=1}^{N}\kappa_i X_i$, $\left[\sum_{i=1}^{N}\kappa_{i0}\right]^{-1}\sum_{i=1}^{N}\kappa_{i0}X_i$, and $\left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\sum_{i=1}^{N}\kappa_{i1}X_i$. Indeed, whenever there are no always-takers or never-takers, the denominator of $\left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\sum_{i=1}^{N}\kappa_{i1}X_i$ and $\left[\sum_{i=1}^{N}\kappa_{i0}\right]^{-1}\sum_{i=1}^{N}\kappa_{i0}X_i$, respectively, is nonnegative by construction and bounded away from zero.

3.4 Maximum Likelihood Estimation

Our discussion so far assumes that the instrument propensity score is known, which is often unrealistic. In practice, researchers typically adopt a parametric model for $p(X)$, say $F(X, \alpha)$, and estimate the unknown parameters by maximum likelihood (cf. Sant'Anna et al., 2022). It turns out that our observations in Sections 3.2 and 3.3 apply equally in this case. Indeed, the normalized estimators are translation invariant and scale invariant with respect to $g(Y) = \log(Y)$ while the unnormalized estimators are not. At the same time, two specific unnormalized estimators, analogous to those in Section 3.3, avoid near-zero denominators in settings with one-sided noncompliance.

From now on, we will reuse our notation and let $\hat{\tau}_{t,norm}$, $\hat{\tau}_{a,10}$, $\hat{\tau}_a$, $\hat{\tau}_t$ $(= \hat{\tau}_{a,1})$, and $\hat{\tau}_{a,0}$ denote the

analogues of the previously introduced estimators, with $\hat{p}_{ml}(X) = F(X, \hat{\alpha}_{ml})$ replacing $p(X)$ and $\hat{\alpha}_{ml}$ denoting the maximum likelihood estimator of α. So, for example,

$$\hat{\tau}_{t,norm} = \frac{\left[\sum_{i=1}^{N} \frac{Z_i}{\hat{p}_{ml}(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{Y_i Z_i}{\hat{p}_{ml}(X_i)} - \left[\sum_{i=1}^{N} \frac{1-Z_i}{1-\hat{p}_{ml}(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{Y_i(1-Z_i)}{1-\hat{p}_{ml}(X_i)}}{\left[\sum_{i=1}^{N} \frac{Z_i}{\hat{p}_{ml}(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{D_i Z_i}{\hat{p}_{ml}(X_i)} - \left[\sum_{i=1}^{N} \frac{1-Z_i}{1-\hat{p}_{ml}(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{D_i(1-Z_i)}{1-\hat{p}_{ml}(X_i)}}, \tag{8}$$

and similarly for the remaining estimators.

3.5 Covariate Balancing Estimation

Our conclusions so far are somewhat perplexing: $\hat{\tau}_a$, $\hat{\tau}_t$ $(= \hat{\tau}_{a,1})$, and $\hat{\tau}_{a,0}$ are potentially problematic in practice, as they are not translation invariant or scale invariant with respect to $g(Y) = \log(Y)$; yet, on the other hand, $\hat{\tau}_t$ $(= \hat{\tau}_{a,1})$ and $\hat{\tau}_{a,0}$ avoid near-zero denominators when there are no always-takers or never-takers, respectively. In this section we provide a solution to this conundrum.

Our solution is to consider an estimator analogous to equation (8), albeit using a different estimate of the instrument propensity score, $\hat{p}_{cb}(X) = F(X, \hat{\alpha}_{cb})$, where $\hat{\alpha}_{cb}$ denotes the just-identified variant of the covariate balancing estimator of α proposed by Imai and Ratkovic (2014). This approach is best understood as using a different set of moment conditions than maximum likelihood. Indeed, the population moment conditions in Imai and Ratkovic (2014) are

$$\mathrm{E}\left[X \frac{Z - p(X)}{p(X)\,(1 - p(X))}\right] = 0. \tag{9}$$

There are also other approaches to covariate balancing, many of which have been studied by Graham et al. (2012), Heiler (2022), and Sant'Anna et al. (2022), among others. In this paper, however, we focus on the approach of Imai and Ratkovic (2014), which amounts to using the moment conditions in equation (9) to estimate α. Consequently, we have the following estimator of τ_{LATE}:

$$\hat{\tau}_{cb} = \frac{\left[\sum_{i=1}^{N} \frac{Z_i}{\hat{p}_{cb}(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{Y_i Z_i}{\hat{p}_{cb}(X_i)} - \left[\sum_{i=1}^{N} \frac{1-Z_i}{1-\hat{p}_{cb}(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{Y_i(1-Z_i)}{1-\hat{p}_{cb}(X_i)}}{\left[\sum_{i=1}^{N} \frac{Z_i}{\hat{p}_{cb}(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{D_i Z_i}{\hat{p}_{cb}(X_i)} - \left[\sum_{i=1}^{N} \frac{1-Z_i}{1-\hat{p}_{cb}(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{D_i(1-Z_i)}{1-\hat{p}_{cb}(X_i)}}. \tag{10}$$

This estimator is also considered by Heiler (2022), who shows that it is numerically identical to the analogue of $\hat{\tau}_t$ $(= \hat{\tau}_{a,1})$ that uses $\hat{p}_{cb}(X)$ rather than $p(X)$ or $\hat{p}_{ml}(X)$, as long as X includes a constant.

We build on this observation and determine that, when X includes a constant, $\hat{\tau}_{cb}$ is also identical to the analogues of $\hat{\tau}_{a,10}$ and $\hat{\tau}_{a,0}$ that use $\hat{p}_{cb}(X)$.

Proposition 3.5. *If X includes a constant, the analogues of $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$), $\hat{\tau}_{a,0}$, and $\hat{\tau}_{a,10}$ that use $\hat{p}_{cb}(X)$ are numerically identical and equal to $\hat{\tau}_{cb}$.*

Proof. See Appendix. □

Proposition 3.5 demonstrates the existence of a weighting estimator of τ_{LATE}, $\hat{\tau}_{cb}$, which shares all the advantages of other estimators that we outlined in Sections 3.2 and 3.3. Indeed, because $\hat{\tau}_{cb}$ is normalized, it is translation invariant and scale invariant with respect to $g(Y) = \log(Y)$. At the same time, because it shares the structure of $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$) and $\hat{\tau}_{a,0}$, it avoids near-zero denominators when there are no always-takers *and also* when there are instead no never-takers. This estimator is also recommended by Heiler (2022) but we are the first to determine its advantages listed above.

3.6 Asymptotic Theory

So far, we have focused on the finite sample properties of several weighting estimators of the LATE. In this section we move on to the asymptotic properties of these estimators, which we study in a unified framework of M-estimation. The M-estimator, $\hat{\theta}$, of 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 $\mathrm{E}[\psi(O, \theta)] = 0$.[2] Under standard regularity conditions[3] 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 $\mathrm{E}[\psi(O, \theta)\psi(O, \theta)']$ exists and is finite, the asymptotic distribution of an M-estimator is given by

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, A^{-1}VA^{-1\prime}) \tag{11}$$

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

[3] 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).

with

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

Since all the weighting estimators considered in this paper can be represented as an M-estimator, we can apply these general results to obtain the asymptotic distribution of each estimator.

Weighting estimators are all functions of the instrument propensity score (IPS), $p(X)$. In this section, as in Sections 3.4 and 3.5, we assume a parametric model, $F(X,\alpha)$, for $p(X)$. Thus, the LATE can be estimated by a two-step procedure where the parameters of the instrument propensity score are estimated in the first step and the unknown IPS is replaced with its estimate in the second step to estimate the LATE. Alternatively, one could jointly estimate α and τ_{LATE} within an M-estimation framework using both moment functions related to α and τ_{LATE}. The moment function related to the estimation of α is either the score from the maximum likelihood estimation or the covariate balancing condition from Imai and Ratkovic (2014). The moment functions related to τ_{LATE} are derived from the identification results of the LATE. All moment functions are summarized in Table 2. For different weighting estimators, different combinations of moment functions will be necessary. For example, if τ_{LATE} is estimated by $\hat{\tau}_a$ using ML-based propensity scores, then

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

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 (11) 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 previous sections.

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 $\hat{\tau}_a$, $\hat{\tau}_{a,1}$ $(= \hat{\tau}_t)$, $\hat{\tau}_{a,0}$, $\hat{\tau}_{t,norm}$, and $\hat{\tau}_{cb}$ by Δ, i.e.,

$$\Delta \equiv \mathrm{E}\left[Y\frac{Z-p(X)}{p(X)\,(1-p(X))}\right]. \tag{12}$$

Recall that the expectation on the right hand side is equal to $\mathrm{E}\left[(\kappa_1 - \kappa_0)\,Y\right]$; see equation (2). 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 (12) 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 the normalized weights are used to estimate μ_z and m_z for $z = 0, 1$, the resulting ratio estimator corresponds to $\hat{\tau}_{t,norm}$ or $\hat{\tau}_{cb}$, depending on whether the propensity score is estimated using maximum likelihood or covariate balancing, respectively.

In what follows, we first consider ML-based estimation of the instrument propensity score. For the estimator $\hat{\tau}_a$, we use the moment functions $\psi_{\alpha,ml}$, ψ_Δ, and ψ_Γ. Based on the result given in equation (11), the asymptotic distribution of $\hat{\tau}_a$ can be derived as follows:

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

where

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

with

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

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

The estimators $\hat{\tau}_{a,1}$ $(= \hat{\tau}_t)$ and $\hat{\tau}_{a,0}$ use the same moment functions for α and Δ as $\hat{\tau}_a$. 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 $\hat{\tau}_{a,1}$ and $\hat{\tau}_{a,0}$ have the same form as $\hat{\tau}_a$, where Γ is replaced with Γ_1 and Γ_0. Thus, the asymptotic distributions of $\hat{\tau}_{a,1}$ and $\hat{\tau}_{a,0}$ can be summarized as follows:

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

where

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

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

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

where

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

with

$$\begin{aligned}
\psi_{\Gamma_0} &= \frac{Z_i(D_i-1)}{F(X_i,\alpha)} - \frac{(1-Z_i)(D_i-1)}{1-F(X_i,\alpha)} - \Gamma_0, \\
E_{\Gamma_0,\alpha} &= \text{E}\left[\frac{\partial\psi_{\Gamma_0}}{\partial\alpha}\right] = \text{E}\left[-\left(\frac{(D-1)Z}{F(X,\alpha)^2} + \frac{(D-1)(1-Z)}{(1-F(X,\alpha))^2}\right)\nabla_\alpha F(X,\alpha)\right].
\end{aligned}$$

The estimator $\hat{\tau}_{a,10}$ 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

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

where

$$\begin{aligned}
V_{\tau_{a,10}} &= -\left(\frac{E_{\Delta_1,\alpha}}{\Gamma_1} - \frac{E_{\Delta_0,\alpha}}{\Gamma_0} - \frac{\Delta_1 E_{\Gamma_1,\alpha}}{\Gamma_1^2} + \frac{\Delta_0 E_{\Gamma_0,\alpha}}{\Gamma_0^2}\right)(-E_H^{-1})\left(\frac{E_{\Delta_1,\alpha}}{\Gamma_1} - \frac{E_{\Delta_0,\alpha}}{\Gamma_0} - \frac{\Delta_1 E_{\Gamma_1,\alpha}}{\Gamma_1^2} + \frac{\Delta_0 E_{\Gamma_0,\alpha}}{\Gamma_0^2}\right)' \\
&+ \text{E}\left(\frac{1}{\Gamma_1}\psi_{\Delta_1} - \frac{\Delta_1}{\Gamma_1^2}\psi_{\Gamma_1}\right)^2 + \text{E}\left(\frac{1}{\Gamma_0}\psi_{\Delta_0} - \frac{\Delta_0}{\Gamma_0^2}\psi_{\Gamma_0}\right)^2
\end{aligned}$$

with

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

Finally, we examine the estimators $\hat{\tau}_{t,norm}$ and $\hat{\tau}_{cb}$, which are both ratio estimators with the same

structure. 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 $\hat{\tau}_{t,norm}$, while it is estimated using covariate balancing for $\hat{\tau}_{cb}$. As a result, the former employs $\psi_{\alpha,ml}$ whereas the latter uses $\psi_{\alpha,cb}$ 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:

$$\sqrt{N}\left(\hat{\tau}_{t,norm} - \tau_{\text{LATE}}\right) \xrightarrow{d} \mathcal{N}(0, V_{\tau_{t,norm}}),$$

where

$$\begin{aligned} V_{\tau_{t,norm}} &= -\left(\frac{1}{\Gamma}(E_{\mu_1,\alpha} - E_{\mu_0,\alpha}) - \frac{\Delta}{\Gamma^2}(E_{m_1,\alpha} - E_{m_0,\alpha})\right)(-E_H^{-1})\left(\frac{1}{\Gamma}(E_{\mu_1,\alpha} - E_{\mu_0,\alpha}) - \frac{\Delta}{\Gamma^2}(E_{m_1,\alpha} - E_{m_0,\alpha})\right)' \\ &+ \mathrm{E}\left(\frac{1}{\Gamma}\psi_{\mu_1} - \frac{\Delta}{\Gamma^2}\psi_{m_1}\right)^2 + \mathrm{E}\left(\frac{1}{\Gamma}\psi_{\mu_0} - \frac{\Delta}{\Gamma^2}\psi_{m_0}\right)^2 \end{aligned}$$

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

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

where

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

with

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

In fact, $V_{\tau_{t,norm}}$ has the same structure as $V_{\tau_{cb}}$, but it enjoys some additional simplifications when the ML-based moment condition is used to estimate $p(X)$. Namely, $\mathrm{E}\left[\frac{\partial\psi_{\alpha,ml}(\cdot)}{\partial\alpha'}\right] = -\mathrm{E}\left[\psi_{\alpha,ml}(\cdot)\psi_{\alpha,ml}(\cdot)'\right]$, $\mathrm{E}\left[\frac{\partial\psi_{\mu_z}}{\partial\alpha}\right] = -\mathrm{E}\left[\psi_{\mu_z}(\cdot)\psi_{\alpha,ml}(\cdot)'\right]$, and $\mathrm{E}\left[\frac{\partial\psi_{m_z}}{\partial\alpha}\right] = -\mathrm{E}\left[\psi_{m_z}(\cdot)\psi_{\alpha,ml}(\cdot)'\right]$ for $z = 0, 1$.

Although it would be interesting to compare the asymptotic variances of the different weighting estimators of τ_{LATE}, we leave this task to future research, given the very involved expressions above. At this time, we instead make three additional points. First, we conjecture that, as in Kitagawa and Muris (2016) and Khan and Ugander (2021), normalization may help reduce the asymptotic variance of an estimator, in which case $\hat{\tau}_{t,norm}$ would be more efficient than $\hat{\tau}_t$ $(= \hat{\tau}_{a,1})$. Second, we note that our preferred estimator, $\hat{\tau}_{cb}$, attains the semiparametric efficiency bound derived by Frölich (2007) and Hong and Nekipelov (2010) as long as the number of balancing constraints grows appropriately with the sample size (see Heiler, 2022). Third, we recognize that our asymptotic analysis implicitly requires a restriction stronger than Assumption IV(iii), namely the "strong overlap" assumption of Khan and Tamer (2010) and Heiler and Kazak (2021).

4 Simulation Study

In this section we use a simulation study to illustrate our findings on the properties of weighting estimators of the LATE. To reduce the number of researcher degrees of freedom, we focus on data-generating processes from Heiler (2022), which leads to the following system of equations:

$$\begin{aligned} Z &= 1[u < \pi(X)], \\ \pi(X) &= 1/\left(1 + \exp\left(-\mu_z(X) \cdot \theta_0\right)\right), \\ D_z &= 1[\mu_d(X, z) > v], \\ Y_1 &= \mu_{y_1}(X) + \varepsilon_1, \\ Y_0 &= \varepsilon_0, \end{aligned}$$

where u and X are i.i.d. standard uniform, $\begin{pmatrix} \varepsilon_1 \\ \varepsilon_0 \\ v \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0.5 & 0 & 1 \end{bmatrix}\right)$, $\theta_0 = \ln((1-\delta)/\delta)$, and $\delta \in \{0.01, 0.02, 0.05\}$. What remains to be specified is three functions, namely $\mu_d(x, z)$, $\mu_{y_1}(x)$, and $\mu_z(x)$. Our choices for these functions are listed in Table 3. It is useful to note that, given these choices and the fact that X has a standard uniform distribution, δ is equal to the lowest possible value of the instrument propensity score and (symmetrically) one minus the instrument propensity score, that is, $\delta \leq \mathrm{P}(Z = 1 \mid X) \leq 1 - \delta$. Thus, δ controls the degree of overlap in the data.

Note that Designs A.1, B, C, and D in Table 3 are identical to Designs A, B, C, and D, respectively, in Heiler (2022). It is easy to see that Design A.1 corresponds to a setting with (near) one-sided noncompliance, as $\mathrm{P}(D = 1 \mid Z = 1) = \Phi(4) = 0.99997$, where $\Phi(\cdot)$ is the standard normal cdf. It follows that there are essentially no never-takers in Design A.1. To illustrate our findings from Section 3.3 on near-zero denominators, we are also interested in a design with (nearly) no always-takers. This is accomplished by Design A.2, which is identical to Design A.1 except for a small change to $\mu_d(x, z)$ that reverses the direction of noncompliance. Indeed, in Design A.2, $\mathrm{P}(D = 1 \mid Z = 0) = \Phi(-4) = 0.00003$, which means that there are essentially no always-takers.

It is also useful to note that Designs A.1 and A.2 correspond to the case of a fully independent instrument while in the remaining designs the instrument is conditionally independent. Additionally, in Designs A.1, A.2, and B, treatment effect heterogeneity is only due to the correlation between ε_1 and v; in Designs C and D, on the other hand, the dependence of $\mu_{y_1}(X)$ on X constitutes another source of heterogeneity. In the end, the linear IV estimator that controls for X is expected to perform very well in Designs A.1, A.2, and B but not necessarily elsewhere (cf. Heiler, 2022).

In our simulations, similar to Heiler (2022), we thus use the linear IV estimator as a benchmark that the weighting estimators will not be able to outperform in Designs A.1, A.2, and B while almost certainly being able to do so in Designs C and D. We also consider $\hat{\tau}_{cb}$, $\hat{\tau}_{t,norm}$, $\hat{\tau}_{a,10}$, $\hat{\tau}_a$, $\hat{\tau}_{a,1}$ $(= \hat{\tau}_t)$, and $\hat{\tau}_{a,0}$, also controlling for X. This leads to a misspecification in Design D, where $\mu_z(X)$ is quadratic in X but we mistakenly omit the quadratic term. We consider three sample sizes, $N = 500$, $N = 1{,}000$, and $N = 5{,}000$, and 10,000 replications for each combination of a design, a value of δ, and a sample size.

Our main results are reported in Tables A.1 to A.5. For each estimator, we report the mean squared error (MSE), normalized by the MSE of the linear IV estimator, the absolute bias, and the coverage rate for a nominal 95% confidence interval.

In Design A.1, as expected, the linear IV estimator outperforms all weighting estimators of the LATE, with MSEs of these estimators always at least 31% larger, and sometimes orders of magnitude larger, than that of linear IV. With better overlap and larger sample sizes, all estimators have small biases. When overlap is poor and/or samples small, linear IV is better than the weighting estimators in terms of bias, too. Coverage rates are close to the nominal coverage rate for all estimators in all cases. At the same time, in a comparison of different weighting estimators, it turns out that three of them, $\hat{\tau}_t$, $\hat{\tau}_a$, and $\hat{\tau}_{a,10}$, are very unstable when overlap is sufficiently poor, $\delta \in \{0.01, 0.02\}$, and samples are small, $N = 500$. This is documented by very large MSEs in these cases. As predicted by Section 3.3, however, $\hat{\tau}_{a,0}$ does not suffer from instability, even in the most challenging case with $\delta = 0.01$ and $N = 500$. This is because there are (nearly) no never-takers in Design A.1. This stability is also shared by $\hat{\tau}_{cb}$ and $\hat{\tau}_{t,norm}$, which overall perform better than $\hat{\tau}_{a,0}$.

Our results for Design A.2 are generally similar, except for the relative performance of linear IV in terms of bias and, especially, the exact list of weighting estimators that suffer from instability. Unlike in Design A.1, when overlap is poor and/or samples small, the bias of linear IV is not clearly smaller than that of (most of) the weighting estimators. Also, it is $\hat{\tau}_{a,0}$, $\hat{\tau}_{a,10}$, and perhaps $\hat{\tau}_a$ that suffer from instability in such cases—but clearly not $\hat{\tau}_t$. As discussed in Section 3.3, this is because there are (nearly) no always-takers in Design A.2. As before, $\hat{\tau}_{cb}$ and $\hat{\tau}_{t,norm}$ perform marginally better than the best unnormalized estimator (in this case, $\hat{\tau}_t$).

In Design B, the instrument is no longer fully independent and noncompliance is no longer one-sided. While linear IV remains dominant in terms of MSE, it is always outperformed by most of the weighting estimators in terms of bias, often substantially and sometimes by all of them. In a comparison of different weighting estimators, $\hat{\tau}_{cb}$ and $\hat{\tau}_{t,norm}$ remain best overall while $\hat{\tau}_t$, $\hat{\tau}_a$, and $\hat{\tau}_{a,10}$ clearly suffer from instability when overlap is sufficiently poor and samples sufficiently small. The case of $\hat{\tau}_{a,0}$ is borderline, which is perhaps due to the fact that there are many more always-takers than never-takers in this design (although both groups clearly exist, unlike before).

Next, in Design C, we introduce another source of treatment effect heterogeneity through the dependence of $\mu_{y_1}(X)$ on X. The linear IV estimator is no longer consistent for the LATE, which is illustrated by its large bias in all cases, including the least challenging case with $\delta = 0.05$ and $N = 5{,}000$. Given that we define the coverage rate as the fraction of replications in which the LATE is contained in a nominal 95% confidence interval, we also obtain very low coverage rates for linear IV, never exceeding 66% and approaching 0% when the sample size is sufficiently large. Coverage rates for all the weighting estimators are close to the nominal level when overlap is good and samples large enough. The only weighting estimators that never suffer from instability are $\hat{\tau}_{cb}$ and $\hat{\tau}_{t,norm}$, although $\hat{\tau}_{cb}$ is now dominant, with substantial improvements in MSE in all cases.

Finally, in Design D, the instrument propensity score is misspecified, as we mistakenly omit the quadratic in X. The linear IV estimator remains inconsistent, too, and its coverage rates are close to 0% in all cases. While the weighting estimators clearly differ in performance, sometimes in unexpected ways, the most striking feature of the simulation results for Design D is the dominance

of $\hat{\tau}_{cb}$, in terms of MSE, bias, and coverage. In fact, despite misspecification of the instrument propensity score, the coverage rate for $\hat{\tau}_{cb}$ approaches the nominal level when overlap is sufficiently good and samples sufficiently large, which is not the case for any other estimator.

It seems natural to interpret the instability of different weighting estimators of the LATE as a consequence of near-zero denominators, as we have done so far. To corroborate this interpretation, in Figures A.1 to A.5, we present box plots with simulation evidence on all estimators of the proportion of compliers that we consider: the first-stage coefficient on Z in linear IV; the denominator of $\hat{\tau}_{t,norm}$; $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1}$, $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, and $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i}$, with the logit instrument propensity score; the denominator of $\hat{\tau}_{cb}$; and $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1} = N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, with the covariate-balancing instrument propensity score.[4] A straightforward comparison of Tables A.1 to A.5 with Figures A.1 to A.5 reveals that instability of weighting estimators of the LATE is indeed associated with situations in which the supports of their denominators, the estimators of the proportion of compliers, are crossing zero. In fact, it is not negative estimates of this proportion that are particularly problematic, even if they make no logical sense, but rather those estimates that are very close to zero, as this results in dividing by "near zero" to construct an estimate of the LATE, which leads to instability. Additional simulation evidence is also provided in Figures B.1 to B.45, which present histograms for each combination of an estimator, a design, a value of δ, and a sample size. In cases with instability, the normal approximation to the sampling distribution is clearly inappropriate.

5 Empirical Applications

In this section we use three empirical applications to illustrate our findings from Section 3 and qualify some of our simulation results from Section 4. Our conclusions so far can be summarized as follows. It is natural to regard $\hat{\tau}_{cb}$, and perhaps also $\hat{\tau}_{t,norm}$ and $\hat{\tau}_{a,10}$, as the weighting estimators of choice, as these estimators, unlike others, are translation invariant and scale invariant with respect

[4]Even though, as shown in Proposition 3.5, the analogues of $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$), $\hat{\tau}_{a,0}$, and $\hat{\tau}_{a,10}$ that use $\hat{p}_{cb}(X)$ are numerically identical and equal to $\hat{\tau}_{cb}$, it is not the case that the denominator of $\hat{\tau}_{cb}$ and $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1} = N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, with the covariate-balancing instrument propensity score, are identical. Thus, for completeness, we consider both of these estimators of the proportion of compliers in Figures A.1 to A.5.

to the natural logarithm (cf. Proposition 3.2). On the other hand, whenever there are no always-takers or no never-takers, respectively, $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$) and $\hat{\tau}_{a,0}$ have the advantage of being based on a denominator that is bounded away from zero, a property that is also shared by $\hat{\tau}_{cb}$ in both scenarios. In simulations, this property clearly translates to numerical stability of these estimators in settings with one-sided noncompliance. While $\hat{\tau}_{t,norm}$ does not seem to suffer from instability anyway, this is not generally true about $\hat{\tau}_{a,10}$. Based on our simulation results alone, we should perhaps use $\hat{\tau}_{cb}$ exclusively in all applications.

At the same time, it is not clear whether the potential instability of some of the weighting estimators will translate to practical problems in most cases. After all, dividing by "near zero" is still a relatively infrequent phenomenon across 10,000 replications in our simulation study, and instability problems usually disappear altogether in larger samples, with N = 1,000 and N = 5,000. Given that in modern applications samples are usually much larger than 1,000 observations, it is possible that such problems will usually be irrelevant in practice, in which case normalization (and translation and scale invariance) could again play a central role, with $\hat{\tau}_{cb}$, $\hat{\tau}_{t,norm}$, and $\hat{\tau}_{a,10}$ preferable to $\hat{\tau}_t$, $\hat{\tau}_a$, and $\hat{\tau}_{a,0}$. Indeed, this is what our empirical applications seem to suggest.

5.1 Causal Effects of Military Service (Angrist, 1990)

In our first empirical application, we revisit Angrist's (1990) study of causal effects of military service using the draft eligibility instrument. In the early 1970s, during the Vietnam War period, priority for induction was determined in a sequence of televised draft lotteries, in which an integer from 1–365 was randomly assigned (without replacement) to each date of birth in a given cohort. Subsequently, only men with lottery numbers below a ceiling determined by the Defense Department could have been drafted. Thus, the draft eligibility instrument in Angrist (1990) takes the value 1 for individuals with lottery numbers below the ceiling and 0 otherwise. Because the ceilings were cohort specific, it is essential to control for age in subsequent analysis.

This study has been revisited by Kitagawa (2015) and Mourifié and Wan (2017), among others. In what follows, we use a sample of 3,027 individuals from the 1984 Survey of Income and

Program Participation (SIPP), which is also considered by Mourifié and Wan (2017). Our outcome of interest is log wage. We also consider five sets of covariates: race and years of schooling, as in Mourifié and Wan (2017); age; a cubic in age; race, years of schooling, and age; and race, years of schooling, and a cubic in age. Summary statistics for these data are reported in Table 6 of Mourifié and Wan (2017).

Table 4 reports our estimates of causal effects of military service on log wages for each of the five specifications. Panels A and B, which report IV and normalized weighting estimates, respectively, suggest that these effects were positive and economically meaningful in the period under study, with a range of estimates from 15–34 log points. The differences between the IV and weighting estimates (as well as their standard errors) are always very minor. Although the estimated effects are all positive, they are not statistically different from zero in columns 2–5, that is, whenever we control for age, possibly among other covariates.

Panel C of Table 4 reports unnormalized weighting estimates for the same specifications. Unlike in panels A and B, these estimates are heavily dependent on the set of covariates that we use. When we control for race and years of schooling (column 1), the estimates and standard errors are practically identical to the IV and normalized weighting estimates. Controlling for age substantially reduces the estimates, which are very small but remain positive when race and years of schooling are not additionally controlled for (column 2) while becoming slightly negative when they are (column 4). However, when age is replaced with a cubic in age, the estimates again become positive and large in magnitude while remaining insignificant (columns 3 and 5). Importantly, the apparent fragility of the unnormalized weighting estimates is not shared by the IV and normalized estimates in panels A and B, as discussed above.

5.2 Causal Effects of College Education (Card, 1995)

In our second empirical application, we revisit Card's (1995) study of causal effects of education using the college proximity instrument. Card (1995) uses data from the National Longitudinal Survey of Young Men (NLSYM) and restricts his attention to a subsample of 3,010 individuals who

were interviewed in 1976 and reported valid information on wage and education. His endogenous variable of interest is years of schooling, which is instrumented by an indicator for the presence of a four-year college in the respondent's local labor market in 1966.

This study has been revisited by many papers, including Tan (2006), Huber and Mellace (2015), Kitagawa (2015), Mourifié and Wan (2017), Andresen and Huber (2021), Słoczyński (2021), and Blandhol et al. (2022). Most of these papers focus on binarized versions of Card's (1995) main endogenous explanatory variable of interest. Specifically, Tan (2006) and Słoczyński (2021) study the effects of having at least thirteen years of schooling ("some college attendance") while Huber and Mellace (2015), Kitagawa (2015), Mourifié and Wan (2017), and Andresen and Huber (2021) focus on having at least sixteen years of schooling ("four-year college degree"). In what follows, we consider both binarizations as well as an additional treatment, which we define as having at least fourteen years of schooling ("two-year college degree"). Our outcome of interest is log wage. We also consider two sets of covariates: a quadratic in experience, nine regional indicators, and indicators for whether Black, whether lived in an SMSA in 1966 and 1976, and whether lived in the South in 1976, as in Card (1995); and indicators for whether Black, whether lived in an SMSA in 1966 and 1976, and whether lived in the South in 1966 and 1976, as in Kitagawa (2015). Summary statistics for these data are reported in Table 1 of Card (1995).

Table 5 reports our estimates of causal effects of college education on log wages. As previously noted by Słoczyński (2021), the IV estimates, as reported in panel A, are "too large," in the sense that it is implausible and inconsistent with the recent applied literature that some college attendance could increase wages by 58–66 log points, with estimated effects of two- and four-year degrees that are even larger. Słoczyński (2021) argues that this is driven by a failure of Assumption IV(iv). At the same time, Andresen and Huber (2021) argue that the "four-year college degree" treatment violates Assumption IV(ii). Importantly, however, Andresen and Huber's (2021) test would not reject the null of no violation at least for the "some college attendance" treatment.

In this paper we ignore these possible violations of Assumption IV and instead observe that the estimated effects are no longer "too large" in panel B, which reports the normalized weighting

estimates. The substantial decrease in the magnitude of the estimated effects leads to a lack of statistical significance of these estimates. Taken at face value, however, the estimates suggest that some college attendance increases wages by 29–38 log points while two- and four-year degrees would increase wages by 34–45 and 59–85 log points, respectively. This is much more plausible than the IV estimates in panel A.

Panel C of Table 5 reports the corresponding values of $\hat{\tau}_a$, $\hat{\tau}_t$, and $\hat{\tau}_{a,0}$. These unnormalized estimates are all over the place. Whenever we use the set of covariates from Card (1995), the estimated effects of college education are negative, which is not believable. When instead we use the specification from Kitagawa (2015), the estimates are again positive but become extremely large in magnitude, well in excess of the IV estimates that already seemed "too large." As in our replication of Angrist (1990), the normalized estimates do not share this evident fragility of unnormalized weighting.

5.3 Causal Effects of Childbearing (Angrist and Evans, 1998)

In our third empirical application, we revisit Angrist and Evans's (1998) study of causal effects of childbearing using the sibling sex composition and twin birth instruments. Given that fertility is clearly endogenous in standard models of labor market outcomes, many papers have tried to identify exogenous sources of its variation. Rosenzweig and Wolpin (1980) argue that the incidence of a twin birth provides such exogenous variation. Angrist and Evans (1998) use twinning as an instrument for having at least three children in a sample of women with two or more children, while considering the sex composition of the first two children as an alternative instrument, with two boys or two girls shown to substantially increase the likelihood of having another child.

This study has been revisited by Frölich and Melly (2013), Bisbee et al. (2017), Mourifié and Wan (2017), and Farbmacher et al. (2018), among many others. Some papers use the incidence of a same-sex twin birth as an alternative to any twin birth. Farbmacher et al. (2018) argue that both the twin instrument and the same-sex twin instrument are invalid, as dizygotic twinning is known to be correlated with maternal characteristics. As an alternative, Farbmacher et al. (2018)

assume that monozygotic twinning is exogenous, and construct new instruments on the basis of this assumption. In this paper we ignore these alternative instruments, as they are not binary, but we acknowledge the possible concerns about independence of twinning.

In what follows, we use Farbmacher et al.'s (2018) subsample of the 1980 US Census that consists of all women aged 21–35 with at least two children. The number of observations is 394,840, which is nearly identical to the sample size in Angrist and Evans (1998). Summary statistics for these data are reported in Table 2 of Angrist and Evans (1998). Our outcomes of interest are log income and an indicator for labor force participation. The treatment is having more than two children. The set of covariates consists of age, age at first birth, sex of the first and second children, and indicators for whether Black, whether Hispanic, and whether another race. The instruments are indicators for whether the mother gave birth to twins at second birth, whether the mother gave birth to same-sex twins at second birth, and whether the first two children are of the same sex. Clearly, both twin birth instruments only allow for one-sided noncompliance, and it is impossible to be a never-taker. (If a woman gives birth to twins at second birth, she will necessarily have more than two children.) Unlike in previous applications, we do not focus on a comparison across different sets of covariates, as this appears to be largely inconsequential here.

Table 6 reports our estimates of causal effects of childbearing on labor market outcomes. Panels A and B, which report IV and normalized weighting estimates, respectively, suggest that these effects are negative and economically meaningful, although some of the effects on log income are not statistically different from zero. As in our replication of Angrist (1990), the differences between the IV and weighting estimates (as well as their standard errors) are always very minor.

Panel C of Table 6 reports the unnormalized estimates. Interestingly, in columns 1–5, these estimates and their standard errors are also very similar to the estimates and standard errors in panels A and B. These cases correspond to the effects on labor force participation using any instrument and the effects on log income using the twin birth instruments. When instead we focus on causal effects of childbearing on log income using the sibling sex composition instrument (column 6), it turns out that the unnormalized estimates become positive and similar in magnitude to the (nega-

tive) IV and normalized estimates. However, it is clearly not believable that childbearing improves female labor market outcomes, which again illustrates the fragility of unnormalized weighting.

6 Conclusion

In this paper we study the properties of several weighting estimators of the local average treatment effect (LATE), which are based on the identification results of Abadie (2003) and Frölich (2007). We make several novel observations. First, we show that some of the most popular estimators of the LATE are not scale invariant with respect to the natural logarithm or translation invariant, which translates to their sensitivity to the units of measurement when estimating the LATE in logs and the centering of the outcome variable more generally. At the same time, we discuss normalized weighting estimators that possess these important properties. Second, we demonstrate that, perhaps counterintuitively, two unnormalized weighting estimators of the LATE have an advantage of being based on a denominator that is bounded away from zero in settings with one-sided noncompliance. Finally, we study an alternative estimator that has all the desirable properties described so far; this estimator is based on an appropriate covariate balancing approach to estimate the instrument propensity score (see also Imai and Ratkovic, 2014; Heiler, 2022; Sant'Anna et al., 2022).

We illustrate our findings with a simulation study and three empirical applications. In simulations, the covariate balancing estimator and the normalized version of Tan's (2006) estimator perform relatively well in every setting under consideration. In empirical applications, each of the unnormalized estimators appears to be unreliable in at least some cases, with high variability of estimates across specifications as well as several occurrences of "incorrect" signs, magnitudes, or both, including negative estimates of the effects of education on earnings and positive estimates of the effects of fertility on female wages. It is particularly interesting that these issues are present in three of the most influential applications of IV estimation in labor economics, namely, in studies of causal effects of military service using the draft eligibility instrument (Angrist, 1990), causal effects of education using the college proximity instrument (Card, 1995), and causal effects of

childbearing using the sibling sex composition instrument (Angrist and Evans, 1998).

References

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

_ **and Matias D. Cattaneo**, "Econometric Methods for Program Evaluation," *Annual Review of Economics*, 2018, *10*, 465–503.

Abdulkadiroğlu, Atila, Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak, "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation," *Econometrica*, 2017, *85* (5), 1373–1432.

Advani, Arun, Toru Kitagawa, and Tymon Słoczyński, "Mostly Harmless Simulations? Using Monte Carlo Studies for Estimator Selection," *Journal of Applied Econometrics*, 2019, *34* (6), 893–910.

Andresen, Martin E. and Martin Huber, "Instrument-Based Estimation with Binarised Treatments: Issues and Tests for the Exclusion Restriction," *Econometrics Journal*, 2021, *24* (3), 536–558.

Andrews, Isaiah, James H. Stock, and Liyang Sun, "Weak Instruments in Instrumental Variables Regression: Theory and Practice," *Annual Review of Economics*, 2019, *11*, 727–753.

Angrist, Joshua D., "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review*, 1990, *80* (3), 313–336.

_ , "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors," *Journal of Business & Economic Statistics*, 2001, *19* (1), 2–16.

_ **and William N. Evans**, "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review*, 1998, *88* (3), 450–477.

_ **, Guido W. Imbens, and Donald B. Rubin**, "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 1996, *91* (434), 444–455.

_ **, Parag A. Pathak, and Christopher R. Walters**, "Explaining Charter School Effectiveness," *American Economic Journal: Applied Economics*, 2013, *5* (4), 1–27.

Aronow, Peter M. and Joel A. Middleton, "A Class of Unbiased Estimators of the Average Treatment Effect in Randomized Experiments," *Journal of Causal Inference*, 2013, *1* (1), 135–154.

Bisbee, James, Rajeev Dehejia, Cristian Pop-Eleches, and Cyrus Samii, "Local Instruments, Global Extrapolation: External Validity of the Labor Supply–Fertility Local Average Treatment Effect," *Journal of Labor Economics*, 2017, *35* (S1), S99–S147.

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky, "When Is TSLS Actually LATE?," 2022.

Bodory, Hugo and Martin Huber, "The Causalweight Package for Causal Inference in R," 2018.

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

Busso, Matias, John DiNardo, and Justin McCrary, "New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators," *Review of Economics and Statistics*, 2014, *96* (5), 885–897.

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Card, David, "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in Louis N. Christofides, E. Kenneth Grant, and Robert Swidinsky, eds., *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, Toronto–Buffalo–London: University of Toronto Press, 1995, pp. 201–222.

Chen, Jiafeng and Jonathan Roth, "Log-Like? ATEs Defined with Zero Outcomes Are (Arbitrarily) Scale-Dependent," 2022.

Cruces, Guillermo and Sebastian Galiani, "Fertility and Female Labor Supply in Latin America: New Causal Evidence," *Labour Economics*, 2007, *14* (3), 565–573.

Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad, "Family Welfare Cultures," *Quarterly Journal of Economics*, 2014, *129* (4), 1711–1752.

Del Bono, Emilia, Josh Kinsler, and Ronni Pavan, "Identification of Dynamic Latent Factor Models of Skill Formation with Translog Production," *Journal of Applied Econometrics*, 2022, *37* (6), 1256–1265.

Donald, Stephen G., Yu-Chin Hsu, and Robert P. Lieli, "Inverse Probability Weighted Estimation of Local Average Treatment Effects: A Higher Order MSE Expansion," *Statistics and Probability Letters*, 2014, *95*, 132–138.

_ , _ , and _ , "Testing the Unconfoundedness Assumption via Inverse Probability Weighted Estimators of (L)ATT," *Journal of Business & Economic Statistics*, 2014, *32* (3), 395–415.

Farbmacher, Helmut, Raphael Guber, and Johan Vikström, "Increasing the Credibility of the Twin Birth Instrument," *Journal of Applied Econometrics*, 2018, *33* (3), 457–472.

Foster, James E. and Anthony F. Shorrocks, "Subgroup Consistent Poverty Indices," *Econometrica*, 1991, *59* (3), 687–709.

Frölich, Markus, "Nonparametric IV Estimation of Local Average Treatment Effects with Covariates," *Journal of Econometrics*, 2007, *139* (1), 35–75.

_ and Blaise Melly, "Unconditional Quantile Treatment Effects under Endogeneity," *Journal of Business & Economic Statistics*, 2013, *31* (3), 346–357.

Goda, Gopi Shah, Damon Jones, and Colleen Flaherty Manchester, "Retirement Plan Type and Employee Mobility: The Role of Selection," *Journal of Human Resources*, 2017, *52* (3), 654–679.

Graham, Bryan S., Cristine Campos de Xavier Pinto, and Daniel Egel, "Inverse Probability Tilting for Moment Condition Models with Missing Data," *Review of Economic Studies*, 2012, *79* (3), 1053–1079.

Hájek, Jaroslav, "Comment on "An Essay on the Logical Foundations of Survey Sampling, Part One" by D. Basu," in Vidyadhar P. Godambe and David A. Sprott, eds., *Foundations of Statistical Inference*, Toronto–Montreal: Holt, Rinehart and Winston, 1971, p. 236.

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

_ and Ekaterina Kazak, "Valid Inference for Treatment Effect Parameters under Irregular Identification and Many Extreme Propensity Scores," *Journal of Econometrics*, 2021, *222* (2), 1083–1108.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder, "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 2003, *71* (4), 1161–1189.

Hong, Han and Denis Nekipelov, "Semiparametric Efficiency in Nonlinear LATE Models,"

Quantitative Economics, 2010, *1* (2), 279–304.

Huber, Martin and Giovanni Mellace, "Testing Instrument Validity for LATE Identification Based on Inequality Moment Constraints," *Review of Economics and Statistics*, 2015, *97* (2), 398–411.

Imai, Kosuke and Marc Ratkovic, "Covariate Balancing Propensity Score," *Journal of the Royal Statistical Society, Series B*, 2014, *76* (1), 243–263.

Imbens, Guido W., "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *Review of Economics and Statistics*, 2004, *86* (1), 4–29.

_ **and Joshua D. Angrist**, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475.

Khan, Samir and Johan Ugander, "Adaptive Normalization for IPW Estimation," 2021.

Khan, Shakeeb and Elie Tamer, "Irregular Identification, Support Conditions, and Inverse Weight Estimation," *Econometrica*, 2010, *78* (6), 2021–2042.

Kitagawa, Toru, "A Test for Instrument Validity," *Econometrica*, 2015, *83* (5), 2043–2063.

_ **and Chris Muris**, "Model Averaging in Semiparametric Estimation of Treatment Effects," *Journal of Econometrics*, 2016, *193* (1), 271–289.

MaCurdy, Thomas, Xiaohong Chen, and Han Hong, "Flexible Estimation of Treatment Effect Parameters," *American Economic Review: Papers & Proceedings*, 2011, *101* (3), 544–551.

Millimet, Daniel L. and Rusty Tchernis, "On the Specification of Propensity Scores, with Applications to the Analysis of Trade Policies," *Journal of Business & Economic Statistics*, 2009, *27* (3), 397–415.

Mourifié, Ismael and Yuanyuan Wan, "Testing Local Average Treatment Effect Assumptions," *Review of Economics and Statistics*, 2017, *99* (2), 305–313.

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.

Olma, Tomasz, "Nonparametric Estimation of Truncated Conditional Expectation Functions," 2021.

Rosenzweig, Mark R. and Kenneth I. Wolpin, "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, 1980, *48* (1), 227–240.

Sant'Anna, Pedro H. C. and Jun Zhao, "Doubly Robust Difference-in-Differences Estimators," *Journal of Econometrics*, 2020, *219* (1), 101–122.

_ **, Xiaojun Song, and Qi Xu**, "Covariate Distribution Balance via Propensity Scores," *Journal of Applied Econometrics*, 2022, *37* (6), 1093–1120.

Singh, Rahul and Liyang Sun, "Double Robustness for Complier Parameters," 2022.

Słoczyński, Tymon, "When Should We (Not) Interpret Linear IV Estimands as LATE?," 2021.

Tan, Zhiqiang, "Regression and Weighting Methods for Causal Inference Using Instrumental Variables," *Journal of the American Statistical Association*, 2006, *101* (476), 1607–1618.

Uysal, S. Derya, "Doubly Robust IV Estimation of Local Average Treatment Effects," 2011.

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

Zheng, Buhong, "Can a Poverty Index Be Both Relative and Absolute?," *Econometrica*, 1994, *62* (6), 1453–1458.

Proofs

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

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

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

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

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

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

and

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

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

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.

The fact that $\hat{\tau}_{t,norm}$ and $\hat{\tau}_{a,10}$ are scale invariant with respect to the natural logarithm while $\hat{\tau}_a$, $\hat{\tau}_t$ $(= \hat{\tau}_{a,1})$, and $\hat{\tau}_{a,0}$ do not have this property follows from the proofs above and the fact that $\log(aY_i) = \log(a) + \log(Y_i)$.

Proof of Proposition 3.5. The sample moment conditions that correspond to the population moment conditions in equation (9) are $N^{-1}\sum_{i=1}^{N} X_i \frac{Z_i - \hat{p}_{cb}(X_i)}{\hat{p}_{cb}(X_i)(1-\hat{p}_{cb}(X_i))} = 0$. 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 the analogues of $\hat{\tau}_t$ $(= \hat{\tau}_{a,1})$, $\hat{\tau}_{a,0}$, and $\hat{\tau}_{a,10}$ that use $\hat{p}_{cb}(X)$ are numerically identical to each other. They are also identical to $\hat{\tau}_{cb}$ following the result in Heiler (2022), which says that $\hat{\tau}_{cb}$ is identical to the analogue of $\hat{\tau}_t$ that uses $\hat{p}_{cb}(X)$.

Tables and Figures

Table 1: Simplified Formulas for κ, κ_1, and κ_0 in Subpopulations Defined by Z and D

	κ	$\text{sgn}(\kappa)$	κ_1	$\text{sgn}(\kappa_1)$	κ_0	$\text{sgn}(\kappa_0)$
$Z = 1, D = 1$	1	+	$\frac{1}{p(X)}$	+	0	0
$Z = 1, D = 0$	$-\frac{1-p(X)}{p(X)}$	−	0	0	$-\frac{1}{p(X)}$	−
$Z = 0, D = 1$	$-\frac{p(X)}{1-p(X)}$	−	$-\frac{1}{1-p(X)}$	−	0	0
$Z = 0, D = 0$	1	+	0	0	$\frac{1}{1-p(X)}$	+

Table 2: Parameters and Moment Functions

Parameter	Population Relation	Related Moment Condition
α	$\mathrm{P}(Z = 1 \mid X) = F(X, \alpha)$	$\psi_{\alpha,ml} = \frac{(Z_i - F(X_i,\alpha))}{F(X_i,\alpha)(1-F(X,\alpha))} \nabla_\alpha F(X, \alpha)$
		$\psi_{\alpha,cb} = \frac{(Z_i - F(X_i,\alpha))}{F(X_i,\alpha)(1-F(X,\alpha))} X_i$
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}	$\tau_{\mathrm{LATE}} = \frac{\Delta}{\Gamma} = \frac{\Delta}{\Gamma_1} = \frac{\Delta}{\Gamma_0} = \frac{\Delta_1}{\Gamma_1} - \frac{\Delta_0}{\Gamma_0} = \frac{\mu_1-\mu_0}{m_1-m_0}$	$\psi_{\tau_a} = \frac{\Delta}{\Gamma} - \tau_a$
		$\psi_{\tau_{a,1}} = \frac{\Delta}{\Gamma_1} - \tau_{a,1}$
		$\psi_{\tau_{a,0}} = \frac{\Delta}{\Gamma_0} - \tau_{a,0}$
		$\psi_{\tau_{a,10}} = \frac{\Delta_1}{\Gamma_1} - \frac{\Delta_0}{\Gamma_0} - \tau_{a,10}$
		$\psi_{\tau_{t,norm}} = \frac{\mu_1-\mu_0}{m_1-m_0} - \tau_{t,norm}$
		$\psi_{\tau_{cb}} = \frac{\mu_1-\mu_0}{m_1-m_0} - \tau_{cb}$

Table 3: Simulation Designs

	Design A.1	Design A.2	Design B	Design C	Design D
$\mu_d(x,z)$	$4z$	$4(z-1)$	$-1+2x+2.122z$	$-1+2x+2.122z$	$-1+2x+2.122z$
$\mu_{y_1}(x)$	0.3989	0.3989	0.3989	$9(x+3)^2$	$9(x+3)^2$
$\mu_z(x)$	$2x-1$	$2x-1$	$2x-1$	$2x-1$	$x+x^2-1$

Table 4: Causal Effects of Military Service on Log Wages

	(1)	(2)	(3)	(4)	(5)
A. IV	0.338	0.233	0.227	0.170	0.172
	(0.137)	(0.212)	(0.229)	(0.197)	(0.213)
B. Normalized estimates:					
$\hat{\tau}_{cb}$	0.338	0.229	0.210	0.170	0.171
	(0.137)	(0.213)	(0.233)	(0.198)	(0.218)
$\hat{\tau}_{t,norm}$	0.338	0.234	0.202	0.170	0.145
	(0.137)	(0.211)	(0.235)	(0.196)	(0.219)
$\hat{\tau}_{a,10}$	0.338	0.227	0.204	0.166	0.146
	(0.137)	(0.204)	(0.239)	(0.190)	(0.223)
C. Unnormalized estimates:					
$\hat{\tau}_{a}$	0.338	0.015	0.314	–0.037	0.268
	(0.137)	(0.207)	(0.252)	(0.195)	(0.238)
$\hat{\tau}_{t} = \hat{\tau}_{a,1}$	0.338	0.016	0.302	–0.039	0.256
	(0.137)	(0.219)	(0.240)	(0.206)	(0.225)
$\hat{\tau}_{a,0}$	0.338	0.014	0.317	–0.036	0.270
	(0.137)	(0.199)	(0.255)	(0.188)	(0.240)
Age		✓		✓	
Cubic in age			✓		✓
Race	✓			✓	✓
Years of schooling	✓			✓	✓
Observations	3,027	3,027	3,027	3,027	3,027

Notes: The data are Mourifié and Wan's (2017) subsample of the 1984 Survey of Income and Program Participation (SIPP), which is based on Angrist (1990). The outcome is log wages. The treatment is an indicator for whether an individual is a veteran. The instrument is an indicator for whether an individual had a lottery number below the draft eligibility ceiling. "IV" is the linear IV estimate with covariates reported in the table. The weighting estimators are defined in Section 3, with the approach of Imai and Ratkovic (2014) used to estimate the instrument propensity score for $\hat{\tau}_{cb}$ and a logit for the remaining estimators, also controlling for the covariates reported in the table in both cases. Standard errors are in parentheses. For IV, we use robust standard errors. For the remaining estimators, we calculate the standard errors using the asymptotic variance of the M-estimator in Section 3.6.

Table 5: Causal Effects of College Education on Log Wages

	Some college		Two-year degree		Four-year degree	
	(1)	(2)	(3)	(4)	(5)	(6)
A. IV	0.661	0.575	0.741	0.637	1.392	0.991
	(0.294)	(0.308)	(0.340)	(0.352)	(0.798)	(0.610)
B. Normalized estimates:						
$\hat{\tau}_{cb}$	0.376	0.331	0.451	0.375	0.853	0.588
	(0.223)	(0.236)	(0.274)	(0.270)	(0.549)	(0.433)
$\hat{\tau}_{t,norm}$	0.331	0.356	0.377	0.400	0.619	0.628
	(0.202)	(0.244)	(0.233)	(0.278)	(0.387)	(0.448)
$\hat{\tau}_{a,10}$	0.346	0.293	0.391	0.339	0.586	0.836
	(0.200)	(0.252)	(0.227)	(0.307)	(0.356)	(0.821)
C. Unnormalized estimates:						
$\hat{\tau}_a$	–0.319	2.248	–0.362	2.597	–0.594	4.317
	(1.182)	(0.971)	(1.337)	(1.198)	(2.184)	(2.485)
$\hat{\tau}_t = \hat{\tau}_{a,1}$	–0.321	2.053	–0.365	2.340	–0.601	3.651
	(1.201)	(0.813)	(1.362)	(0.976)	(2.251)	(1.780)
$\hat{\tau}_{a,0}$	–0.290	2.846	–0.325	3.430	–0.501	7.241
	(1.036)	(1.592)	(1.152)	(2.141)	(1.728)	(7.245)
Specification	Card	Kitagawa	Card	Kitagawa	Card	Kitagawa
Observations	3,010	3,010	3,010	3,010	3,010	3,010

Notes: The data are Card's (1995) subsample of the National Longitudinal Survey of Young Men (NLSYM). The outcome is log wages. The treatment is an indicator for whether an individual has at least thirteen ("some college"), fourteen ("two-year degree"), or sixteen years of schooling ("four-year degree"). The instrument is an indicator for whether an individual grew up in the vicinity of a four-year college. The first specification ("Card") follows Card (1995) and includes experience, experience squared, nine regional indicators, and indicators for whether Black, whether lived in an SMSA in 1966 and 1976, and whether lived in the South in 1976. The second specification ("Kitagawa") follows Kitagawa (2015) and includes indicators for whether Black, whether lived in an SMSA in 1966 and 1976, and whether lived in the South in 1966 and 1976. "IV" is the linear IV estimate with covariates listed above. The weighting estimators are defined in Section 3, with the approach of Imai and Ratkovic (2014) used to estimate the instrument propensity score for $\hat{\tau}_{cb}$ and a logit for the remaining estimators, also controlling for the covariates listed above in both cases. Standard errors are in parentheses. For IV, we use robust standard errors. For the remaining estimators, we calculate the standard errors using the asymptotic variance of the M-estimator in Section 3.6.

Table 6: Causal Effects of Childbearing on Labor Force Participation and Log Income

	Labor force participation			Log income		
	(1)	(2)	(3)	(4)	(5)	(6)
A. IV	–0.081	–0.082	–0.117	–0.072	–0.112	–0.135
	(0.014)	(0.017)	(0.025)	(0.045)	(0.054)	(0.092)
B. Normalized estimates:						
$\hat{\tau}_{cb}$	–0.085	–0.083	–0.117	–0.084	–0.120	–0.135
	(0.014)	(0.017)	(0.025)	(0.046)	(0.055)	(0.092)
$\hat{\tau}_{t,norm}$	–0.084	–0.083	–0.117	–0.079	–0.119	–0.135
	(0.014)	(0.017)	(0.025)	(0.045)	(0.055)	(0.092)
$\hat{\tau}_{a,10}$	–0.084	–0.083	–0.117	–0.079	–0.119	–0.132
	(0.014)	(0.017)	(0.025)	(0.045)	(0.055)	(0.093)
C. Unnormalized estimates:						
$\hat{\tau}_a$	–0.084	–0.083	–0.100	–0.087	–0.118	0.143
	(0.014)	(0.017)	(0.025)	(0.046)	(0.055)	(0.102)
$\hat{\tau}_t = \hat{\tau}_{a,1}$	–0.084	–0.083	–0.099	–0.087	–0.118	0.140
	(0.014)	(0.017)	(0.025)	(0.046)	(0.055)	(0.100)
$\hat{\tau}_{a,0}$	–0.084	–0.083	–0.102	–0.087	–0.118	0.145
	(0.014)	(0.017)	(0.026)	(0.046)	(0.055)	(0.104)
Instrument	Twins	Same-sex twins	Same-sex siblings	Twins	Same-sex twins	Same-sex siblings
Observations	394,840	394,840	394,840	220,502	220,502	220,502

Notes: The data are Farbmacher et al.'s (2018) subsample of the 1980 US Census, which is based on Angrist and Evans (1998). The outcome is an indicator for whether a woman worked for pay in the preceding year ("labor force participation") or log income. The treatment is an indicator for whether a woman has at least three children. The instrument is an indicator for whether a woman gave birth to twins at second birth (columns 1 and 4), whether she gave birth to same-sex twins at second birth (columns 2 and 5), and whether her first two children are either two boys or two girls (columns 3 and 6). The set of covariates consists of age, age at first birth, sex of the first and second children, and indicators for whether Black, whether Hispanic, and whether another race. "IV" is the linear IV estimate with covariates listed above. The weighting estimators are defined in Section 3, with the approach of Imai and Ratkovic (2014) used to estimate the instrument propensity score for $\hat{\tau}_{cb}$ and a logit for the remaining estimators, also controlling for the covariates listed above in both cases. Standard errors are in parentheses. For IV, we use robust standard errors. For the remaining estimators, we calculate the standard errors using the asymptotic variance of the M-estimator in Section 3.6.

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

			Normalized estimators			Unnormalized estimators		
		IV	$\hat{\tau}_{cb}$	$\hat{\tau}_{t,norm}$	$\hat{\tau}_{a,10}$	$\hat{\tau}_{a}$	$\hat{\tau}_{t} = \hat{\tau}_{a,1}$	$\hat{\tau}_{a,0}$
$\delta = 0.01$								
$N = 500$	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 4. "MSE" is the mean squared error of an estimator, normalized by the mean squared error of linear IV. "|B|" is the absolute bias. "Coverage rate" is the coverage rate for a nominal 95% confidence interval. "IV" is the linear IV estimator that controls for X. The weighting estimators are defined in Section 3, with the approach of Imai and Ratkovic (2014) used to estimate the instrument propensity score for $\hat{\tau}_{cb}$ and a logit for the remaining estimators, also controlling for X in both cases. Results are based on 10,000 replications.

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

			Normalized estimators			Unnormalized estimators		
		IV	$\hat{\tau}_{cb}$	$\hat{\tau}_{t,norm}$	$\hat{\tau}_{a,10}$	$\hat{\tau}_{a}$	$\hat{\tau}_{t} = \hat{\tau}_{a,1}$	$\hat{\tau}_{a,0}$
$\delta = 0.01$								
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 4. "MSE" is the mean squared error of an estimator, normalized by the mean squared error of linear IV. "|B|" is the absolute bias. "Coverage rate" is the coverage rate for a nominal 95% confidence interval. "IV" is the linear IV estimator that controls for X. The weighting estimators are defined in Section 3, with the approach of Imai and Ratkovic (2014) used to estimate the instrument propensity score for $\hat{\tau}_{cb}$ and a logit for the remaining estimators, also controlling for X in both cases. Results are based on 10,000 replications.

Table A.3: Simulation Results for Design B

			Normalized estimators			Unnormalized estimators		
		IV	$\hat{\tau}_{cb}$	$\hat{\tau}_{t,norm}$	$\hat{\tau}_{a,10}$	$\hat{\tau}_{a}$	$\hat{\tau}_{t} = \hat{\tau}_{a,1}$	$\hat{\tau}_{a,0}$
$\delta = 0.01$								
$N = 500$	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 4. "MSE" is the mean squared error of an estimator, normalized by the mean squared error of linear IV. "|B|" is the absolute bias. "Coverage rate" is the coverage rate for a nominal 95% confidence interval. "IV" is the linear IV estimator that controls for X. The weighting estimators are defined in Section 3, with the approach of Imai and Ratkovic (2014) used to estimate the instrument propensity score for $\hat{\tau}_{cb}$ and a logit for the remaining estimators, also controlling for X in both cases. Results are based on 10,000 replications.

Table A.4: Simulation Results for Design C

			Normalized estimators			Unnormalized estimators		
		IV	$\hat{\tau}_{cb}$	$\hat{\tau}_{t,norm}$	$\hat{\tau}_{a,10}$	$\hat{\tau}_{a}$	$\hat{\tau}_{t} = \hat{\tau}_{a,1}$	$\hat{\tau}_{a,0}$
$\delta = 0.01$								
$N = 500$	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

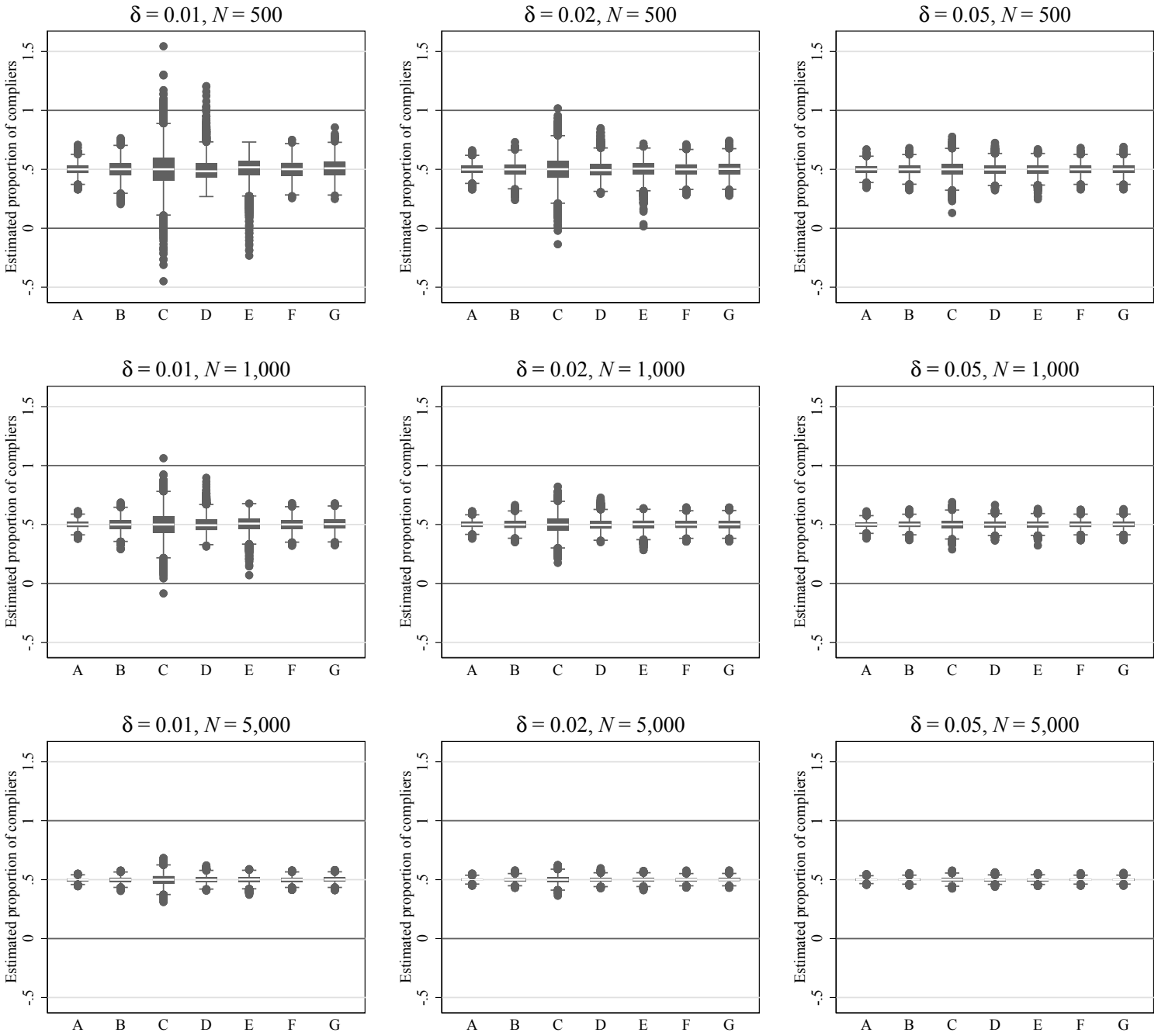
Notes: The details of this simulation design are provided in Section 4. "MSE" is the mean squared error of an estimator, normalized by the mean squared error of linear IV. "|B|" is the absolute bias. "Coverage rate" is the coverage rate for a nominal 95% confidence interval. "IV" is the linear IV estimator that controls for X. The weighting estimators are defined in Section 3, with the approach of Imai and Ratkovic (2014) used to estimate the instrument propensity score for $\hat{\tau}_{cb}$ and a logit for the remaining estimators, also controlling for X in both cases. Results are based on 10,000 replications.

Table A.5: Simulation Results for Design D

			Normalized estimators			Unnormalized estimators		
		IV	$\hat{\tau}_{cb}$	$\hat{\tau}_{t,norm}$	$\hat{\tau}_{a,10}$	$\hat{\tau}_{a}$	$\hat{\tau}_{t} = \hat{\tau}_{a,1}$	$\hat{\tau}_{a,0}$
$\delta = 0.01$								
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

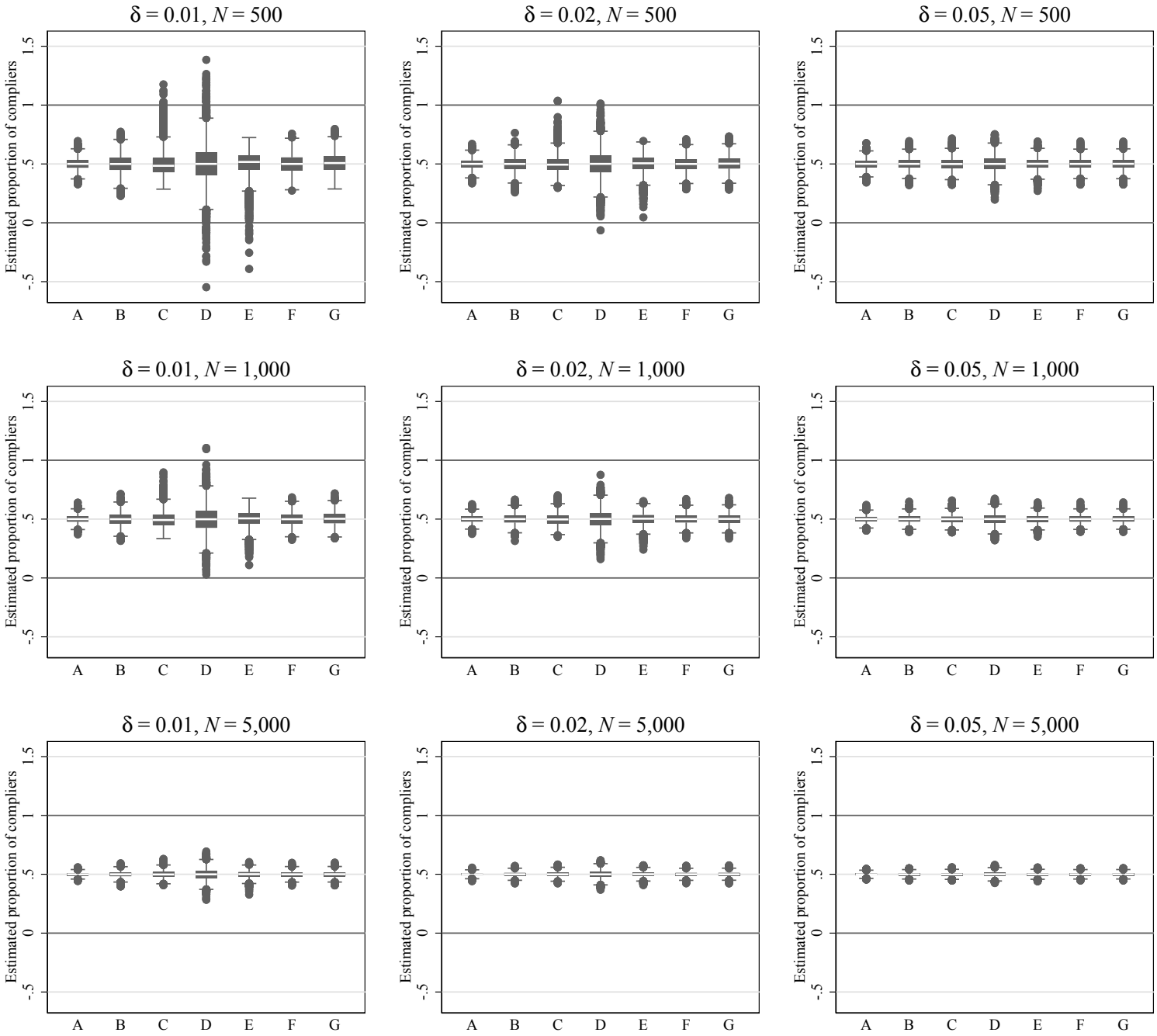
Notes: The details of this simulation design are provided in Section 4. "MSE" is the mean squared error of an estimator, normalized by the mean squared error of linear IV. "|B|" is the absolute bias. "Coverage rate" is the coverage rate for a nominal 95% confidence interval. "IV" is the linear IV estimator that controls for X. The weighting estimators are defined in Section 3, with the approach of Imai and Ratkovic (2014) used to estimate the instrument propensity score for $\hat{\tau}_{cb}$ and a logit for the remaining estimators, also controlling for X in both cases. Results are based on 10,000 replications.

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



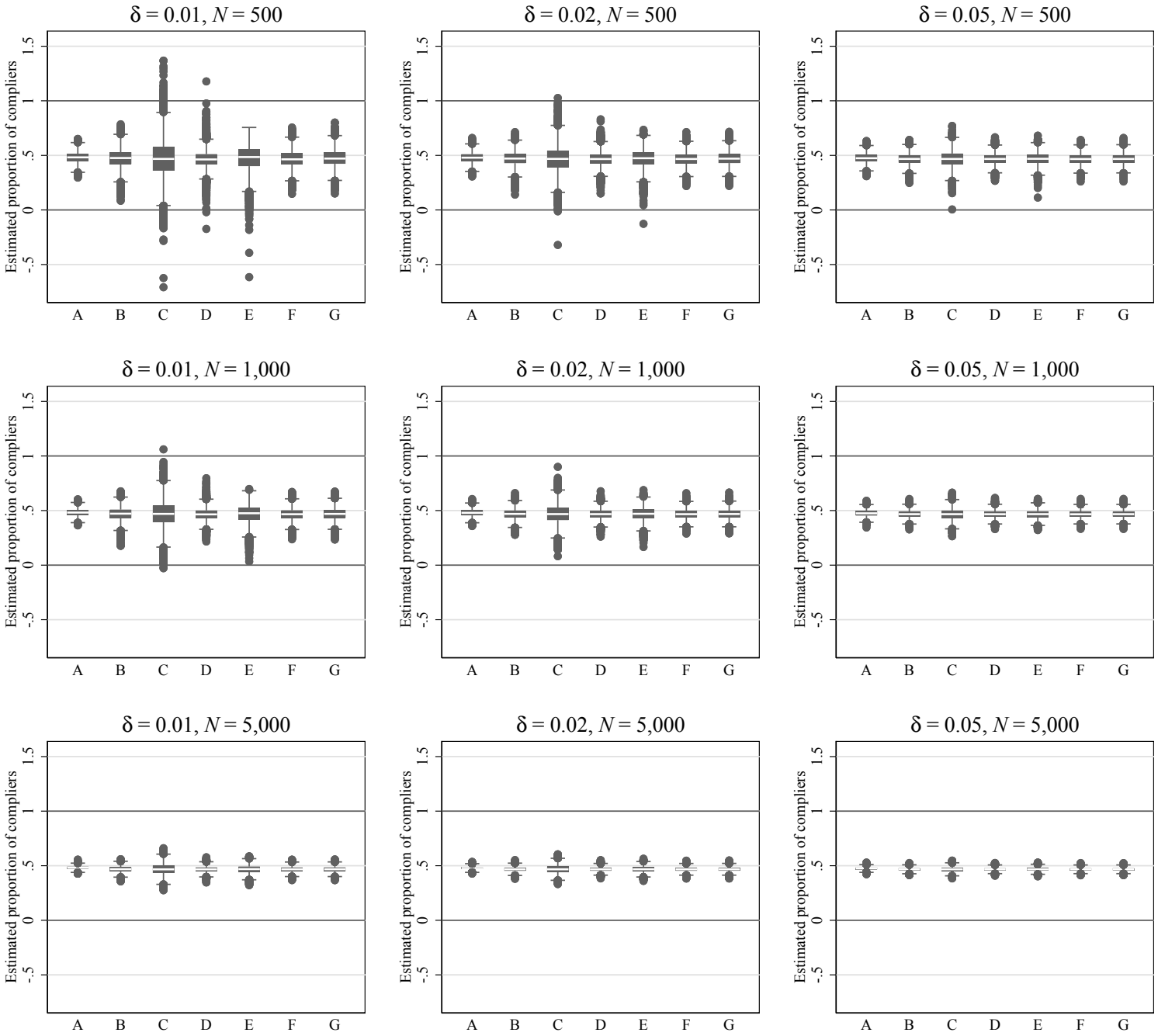
Notes: The details of this simulation design are provided in Section 4. "A" corresponds to the first-stage coefficient on Z in linear IV, controlling for X. "B" corresponds to the denominator of $\hat{\tau}_{t,norm}$. "C," "D," and "E" correspond to $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1}$, $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, and $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i}$, respectively. These estimators, as well as the denominator of $\hat{\tau}_{t,norm}$, are based on an instrument propensity score, which is estimated using a logit, also controlling for X. "F" corresponds to the denominator of $\hat{\tau}_{cb}$, as in equation (10). "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 approach of Imai and Ratkovic (2014), controlling for X, as in the case of the denominator of $\hat{\tau}_{cb}$. Results are based on 10,000 replications.

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



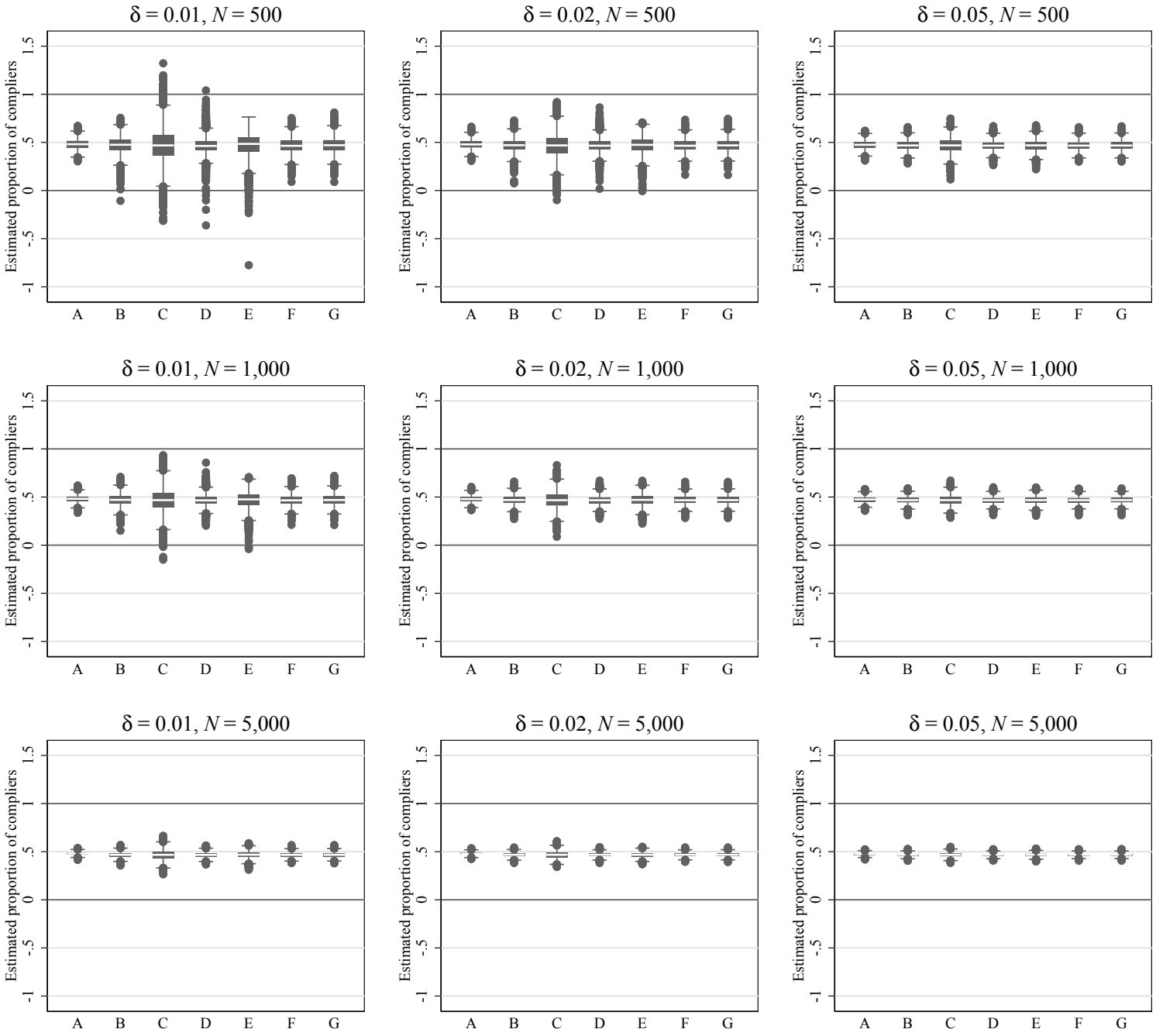
Notes: The details of this simulation design are provided in Section 4. "A" corresponds to the first-stage coefficient on Z in linear IV, controlling for X. "B" corresponds to the denominator of $\hat{\tau}_{t,norm}$. "C," "D," and "E" correspond to $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1}$, $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, and $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i}$, respectively. These estimators, as well as the denominator of $\hat{\tau}_{t,norm}$, are based on an instrument propensity score, which is estimated using a logit, also controlling for X. "F" corresponds to the denominator of $\hat{\tau}_{cb}$, as in equation (10). "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 approach of Imai and Ratkovic (2014), controlling for X, as in the case of the denominator of $\hat{\tau}_{cb}$. Results are based on 10,000 replications.

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



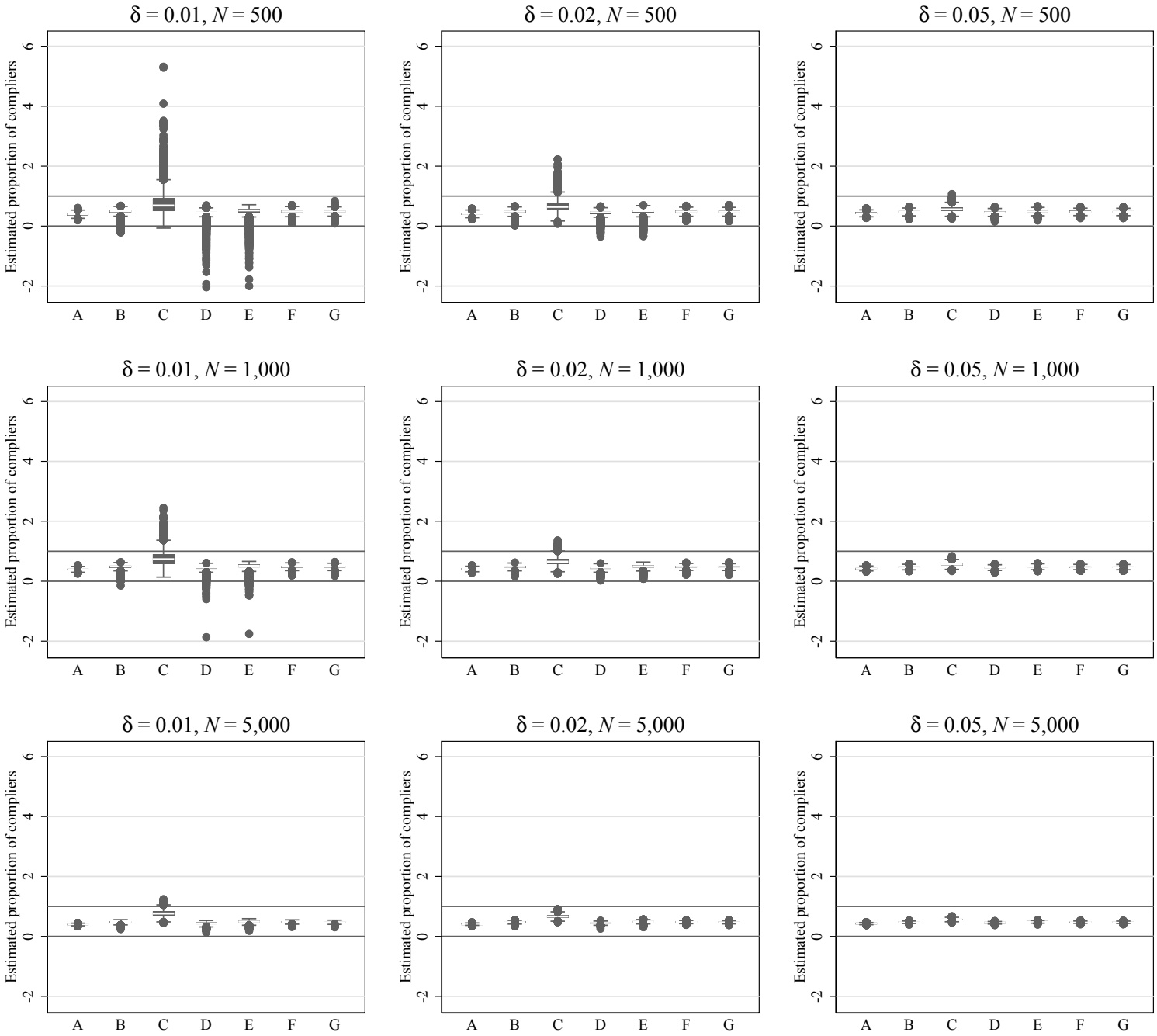
Notes: The details of this simulation design are provided in Section 4. "A" corresponds to the first-stage coefficient on Z in linear IV, controlling for X. "B" corresponds to the denominator of $\hat{\tau}_{t,norm}$. "C," "D," and "E" correspond to $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1}$, $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, and $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i}$, respectively. These estimators, as well as the denominator of $\hat{\tau}_{t,norm}$, are based on an instrument propensity score, which is estimated using a logit, also controlling for X. "F" corresponds to the denominator of $\hat{\tau}_{cb}$, as in equation (10). "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 approach of Imai and Ratkovic (2014), controlling for X, as in the case of the denominator of $\hat{\tau}_{cb}$. Results are based on 10,000 replications.

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



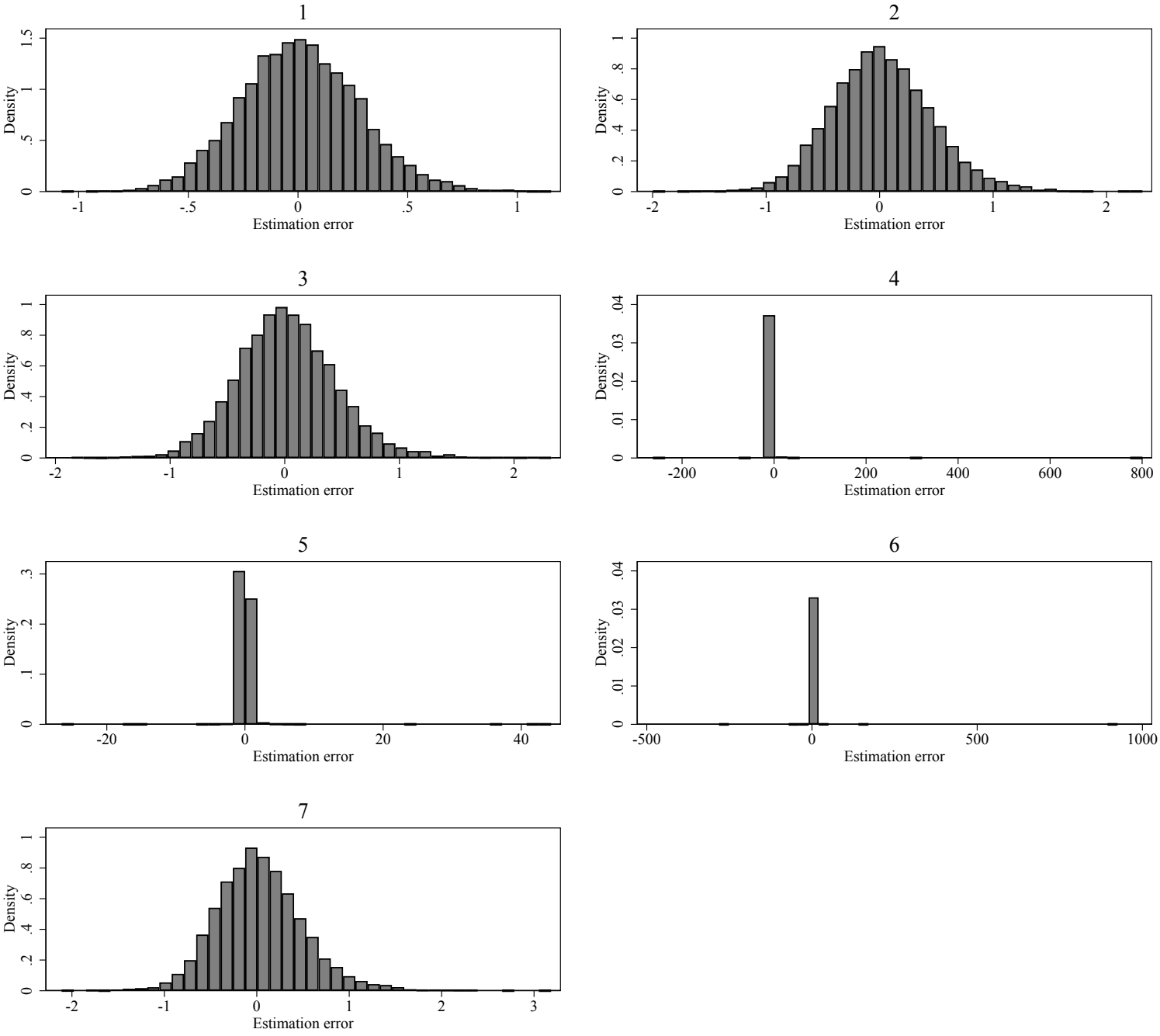
Notes: The details of this simulation design are provided in Section 4. "A" corresponds to the first-stage coefficient on Z in linear IV, controlling for X. "B" corresponds to the denominator of $\hat{\tau}_{t,norm}$. "C," "D," and "E" correspond to $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1}$, $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, and $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i}$, respectively. These estimators, as well as the denominator of $\hat{\tau}_{t,norm}$, are based on an instrument propensity score, which is estimated using a logit, also controlling for X. "F" corresponds to the denominator of $\hat{\tau}_{cb}$, as in equation (10). "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 approach of Imai and Ratkovic (2014), controlling for X, as in the case of the denominator of $\hat{\tau}_{cb}$. Results are based on 10,000 replications.

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



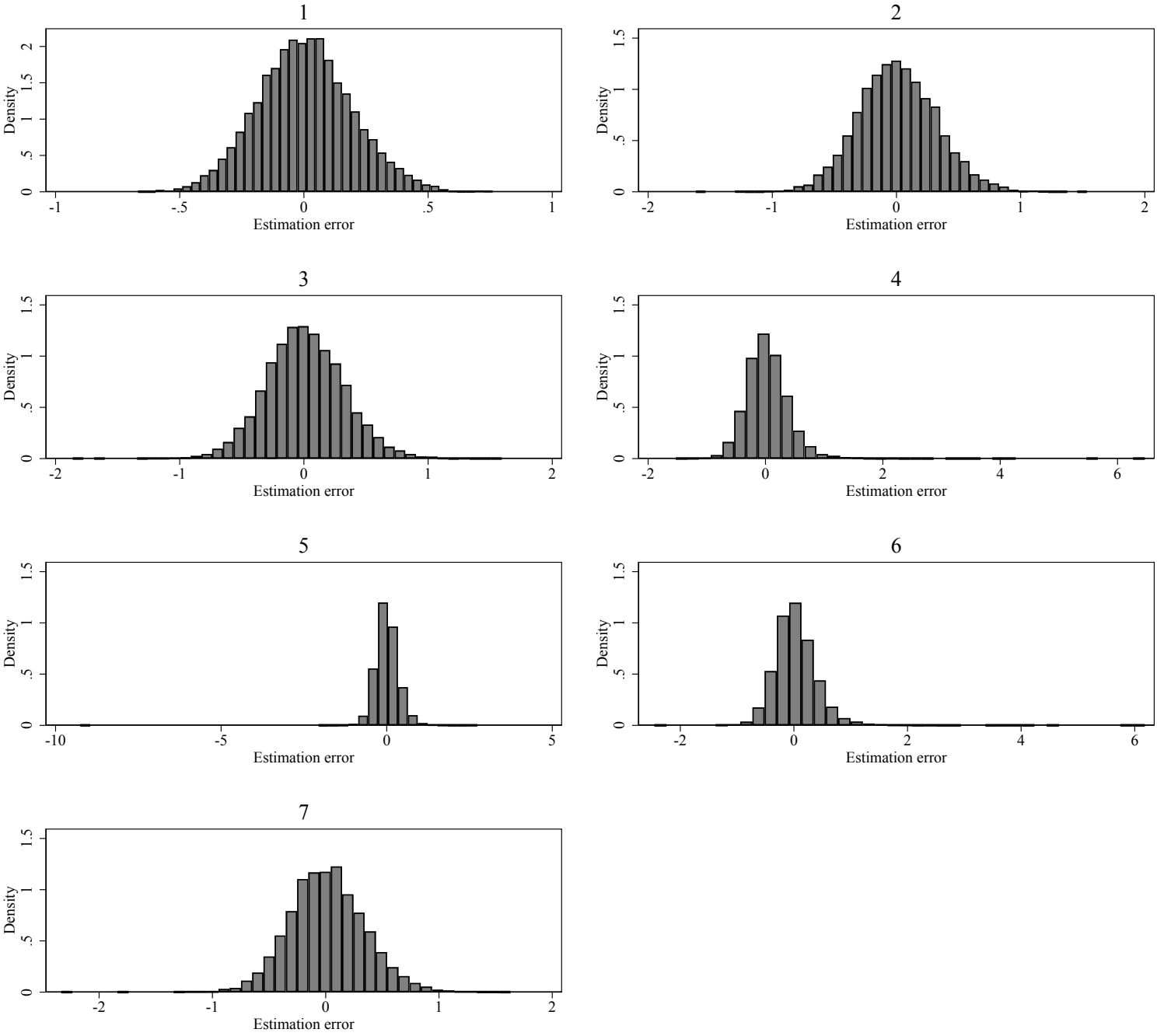
Notes: The details of this simulation design are provided in Section 4. "A" corresponds to the first-stage coefficient on Z in linear IV, controlling for X. "B" corresponds to the denominator of $\hat{\tau}_{t,norm}$. "C," "D," and "E" correspond to $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i1}$, $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i0}$, and $N^{-1}\sum_{i=1}^{N}\hat{\kappa}_{i}$, respectively. These estimators, as well as the denominator of $\hat{\tau}_{t,norm}$, are based on an instrument propensity score, which is estimated using a logit, also controlling for X. "F" corresponds to the denominator of $\hat{\tau}_{cb}$, as in equation (10). "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 approach of Imai and Ratkovic (2014), controlling for X, as in the case of the denominator of $\hat{\tau}_{cb}$. Results are based on 10,000 replications.

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



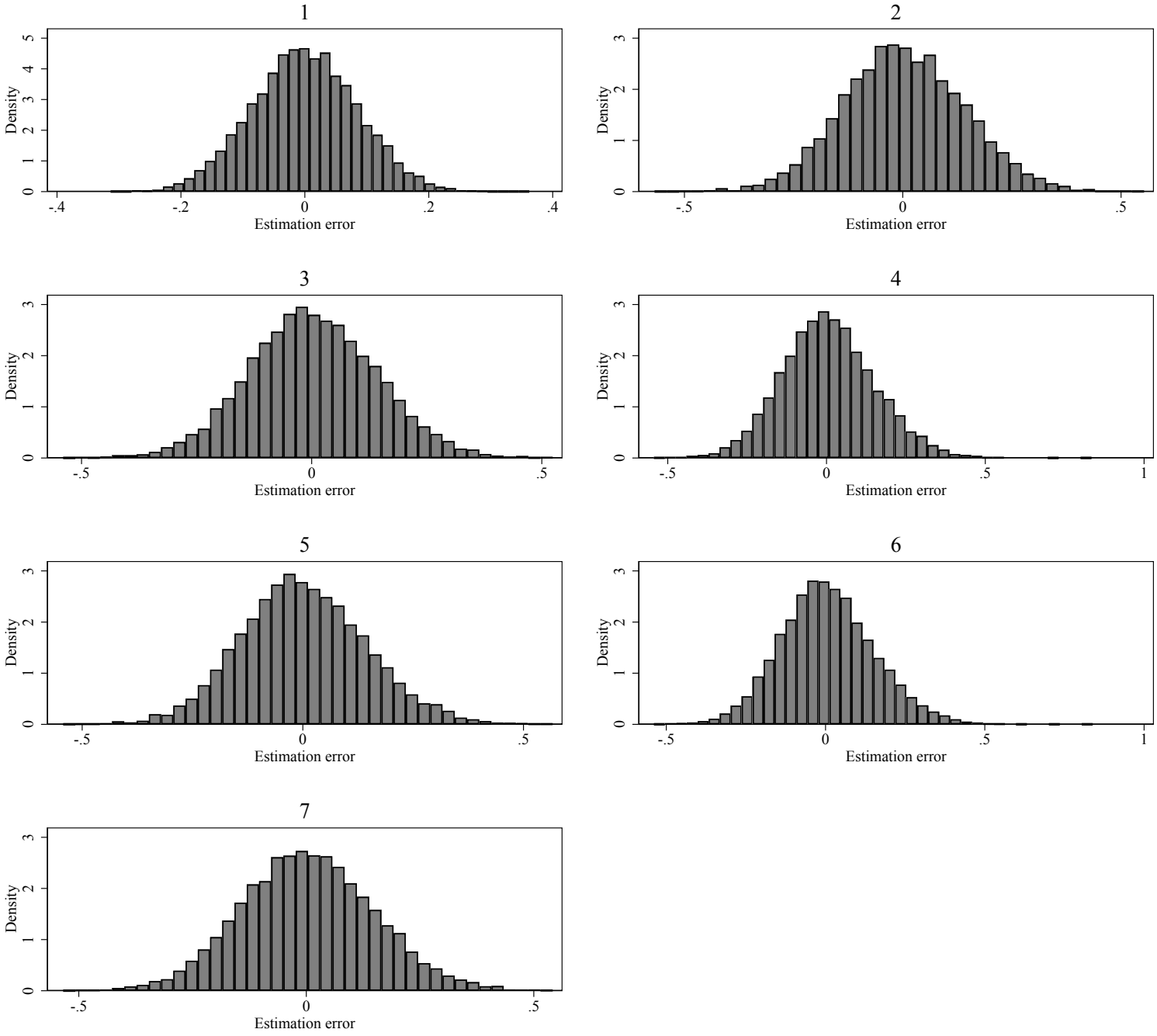
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



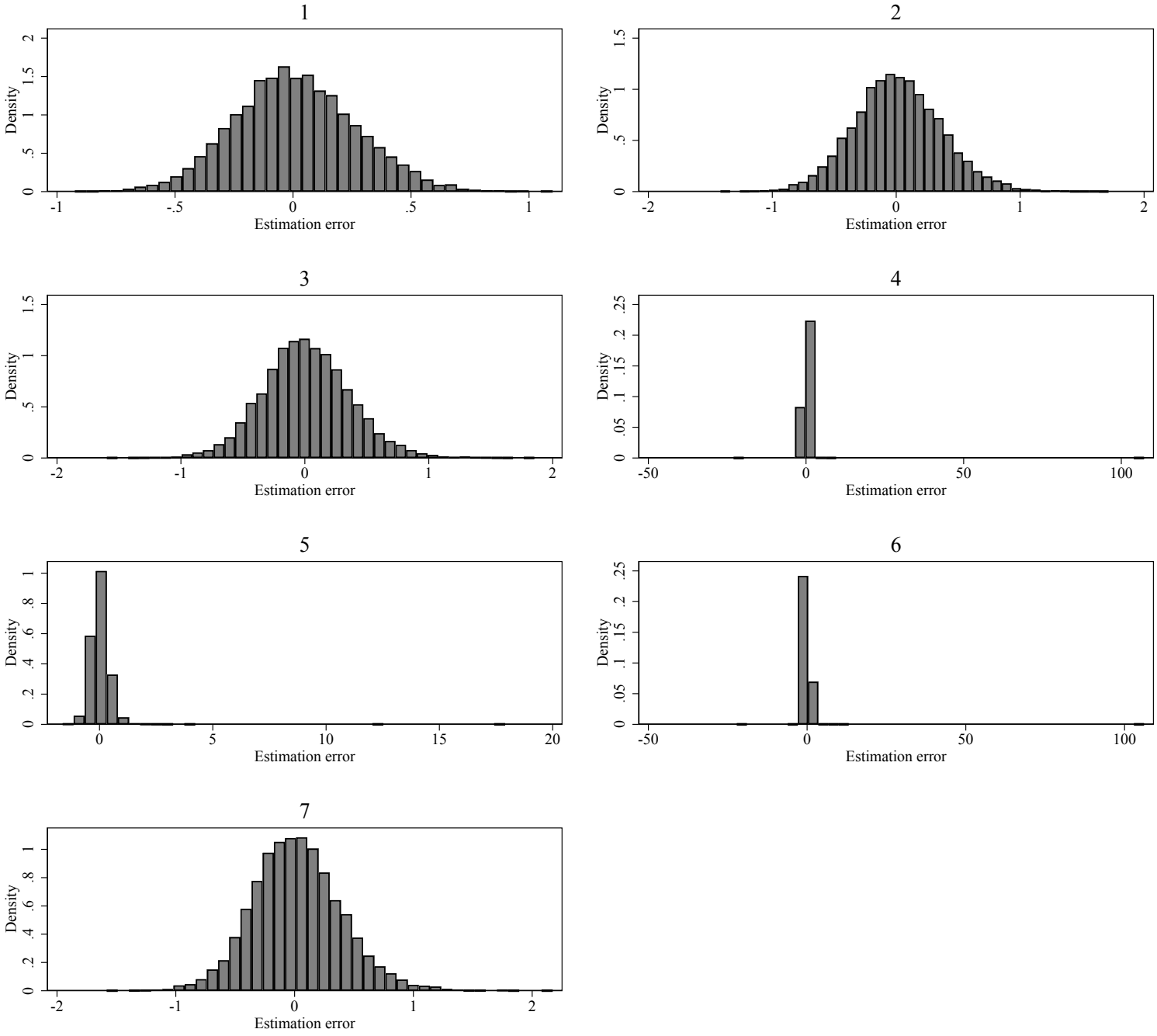
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



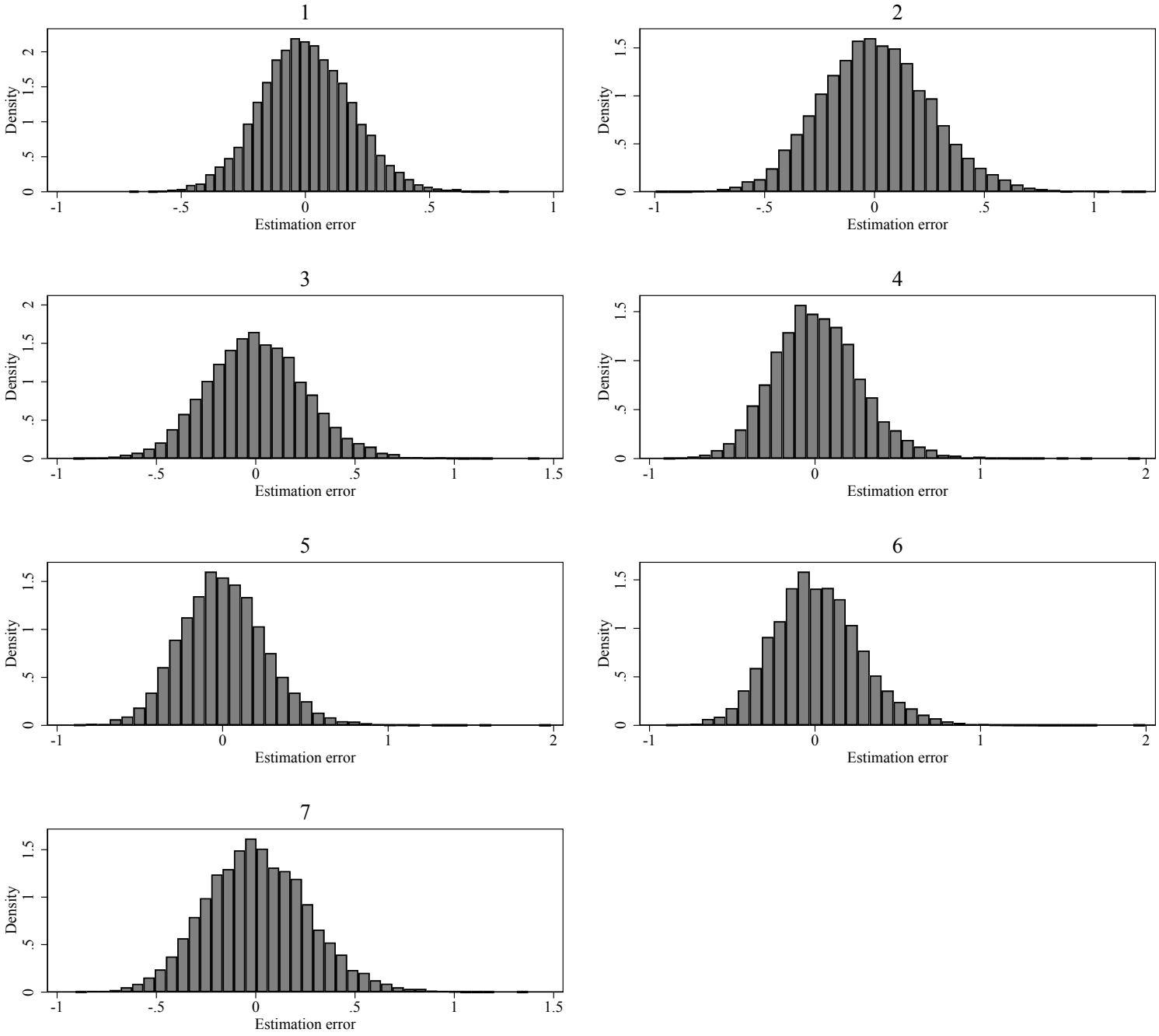
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

Figure B.4: Simulation Results for Design A.1, $\delta = 0.02$, $N = 500$



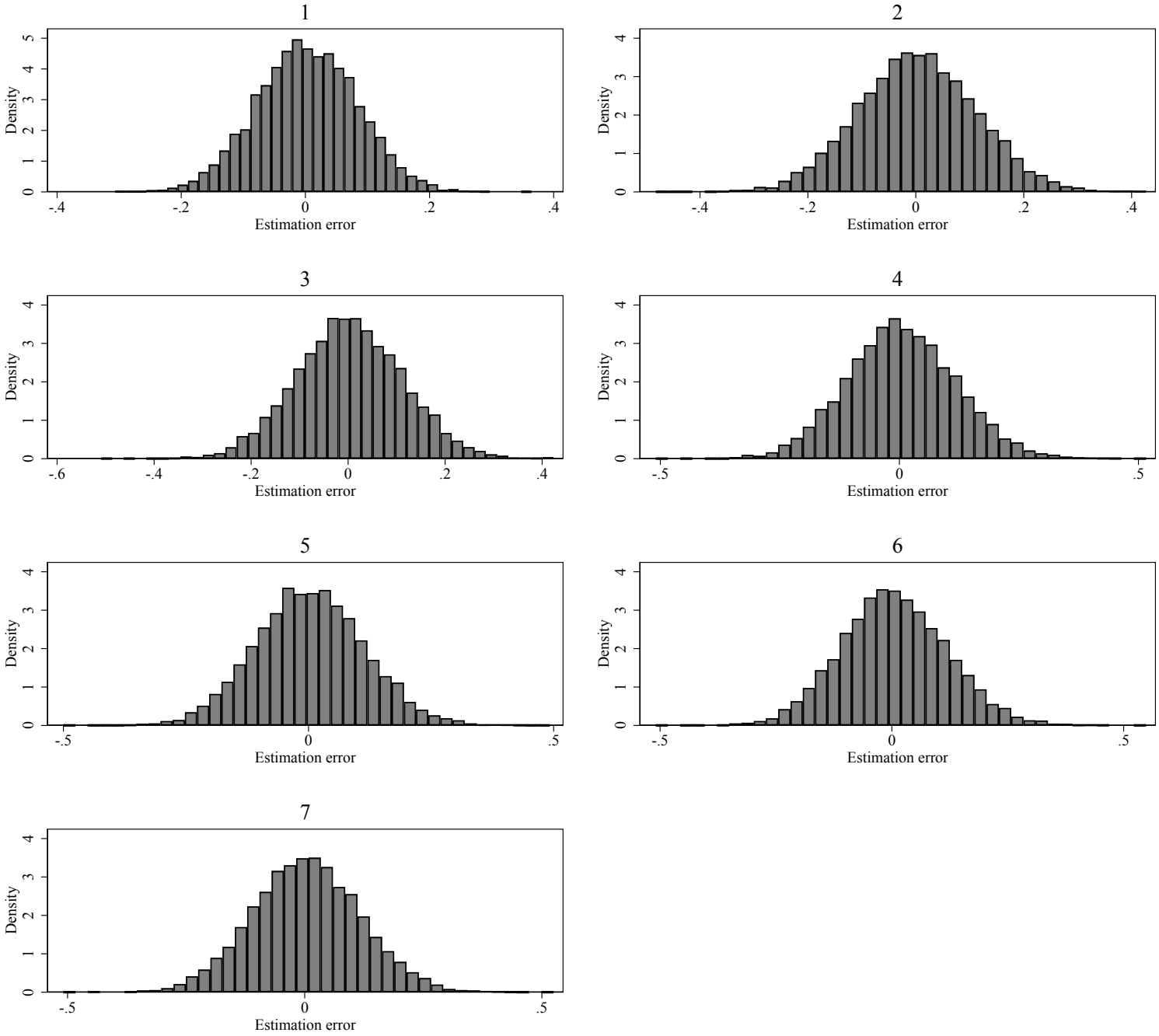
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



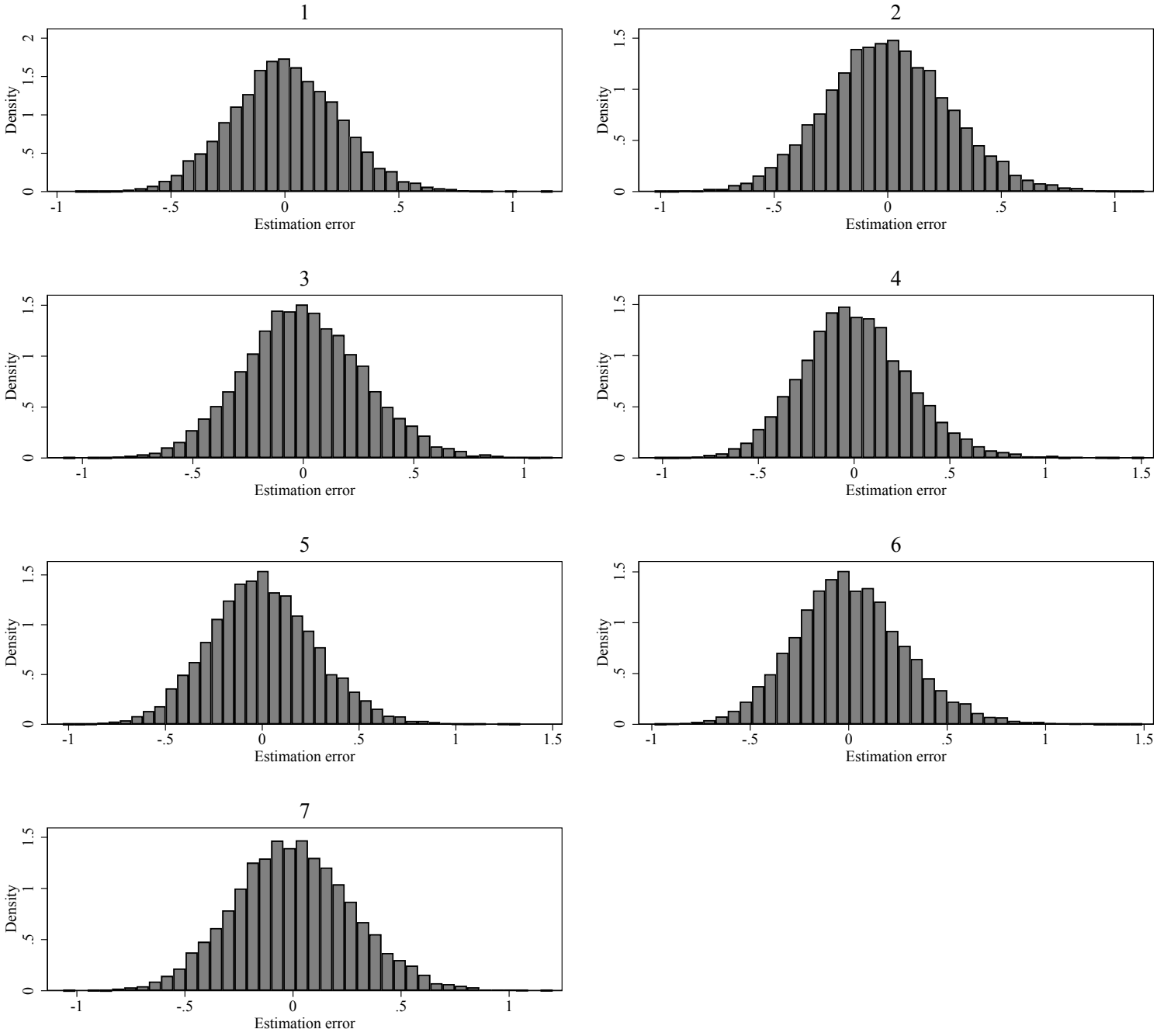
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



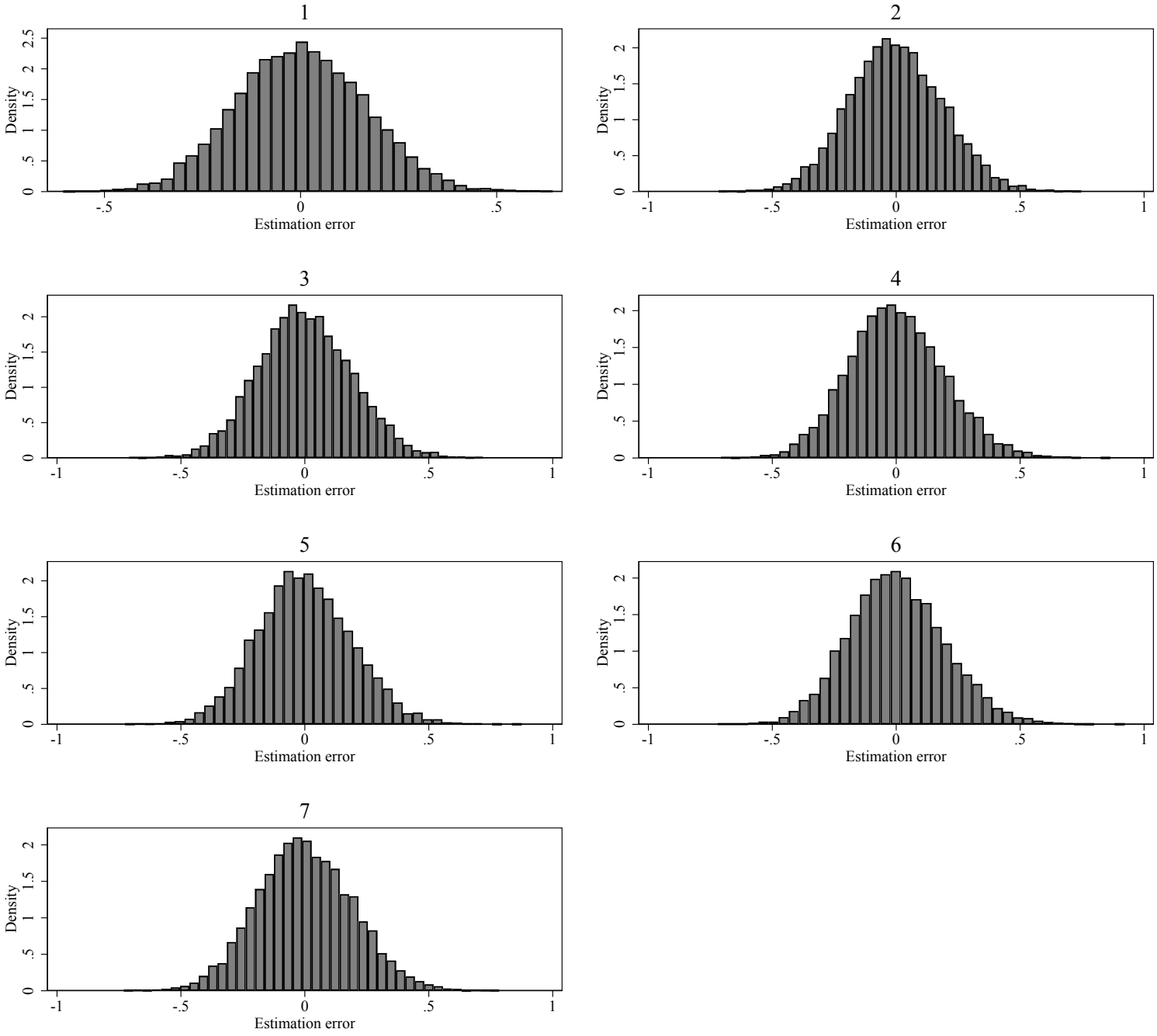
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

Figure B.7: Simulation Results for Design A.1, $\delta = 0.05$, $N = 500$



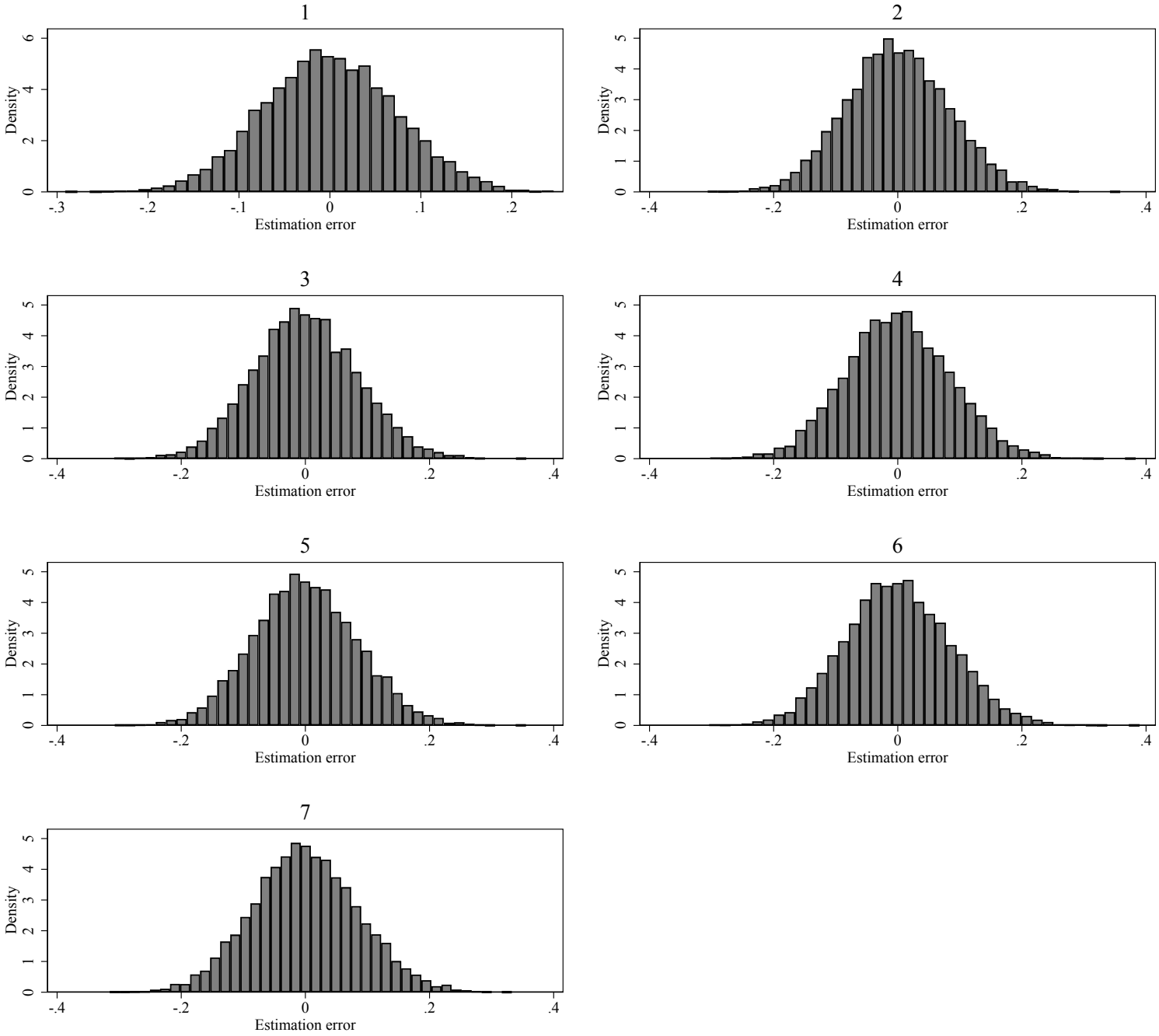
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



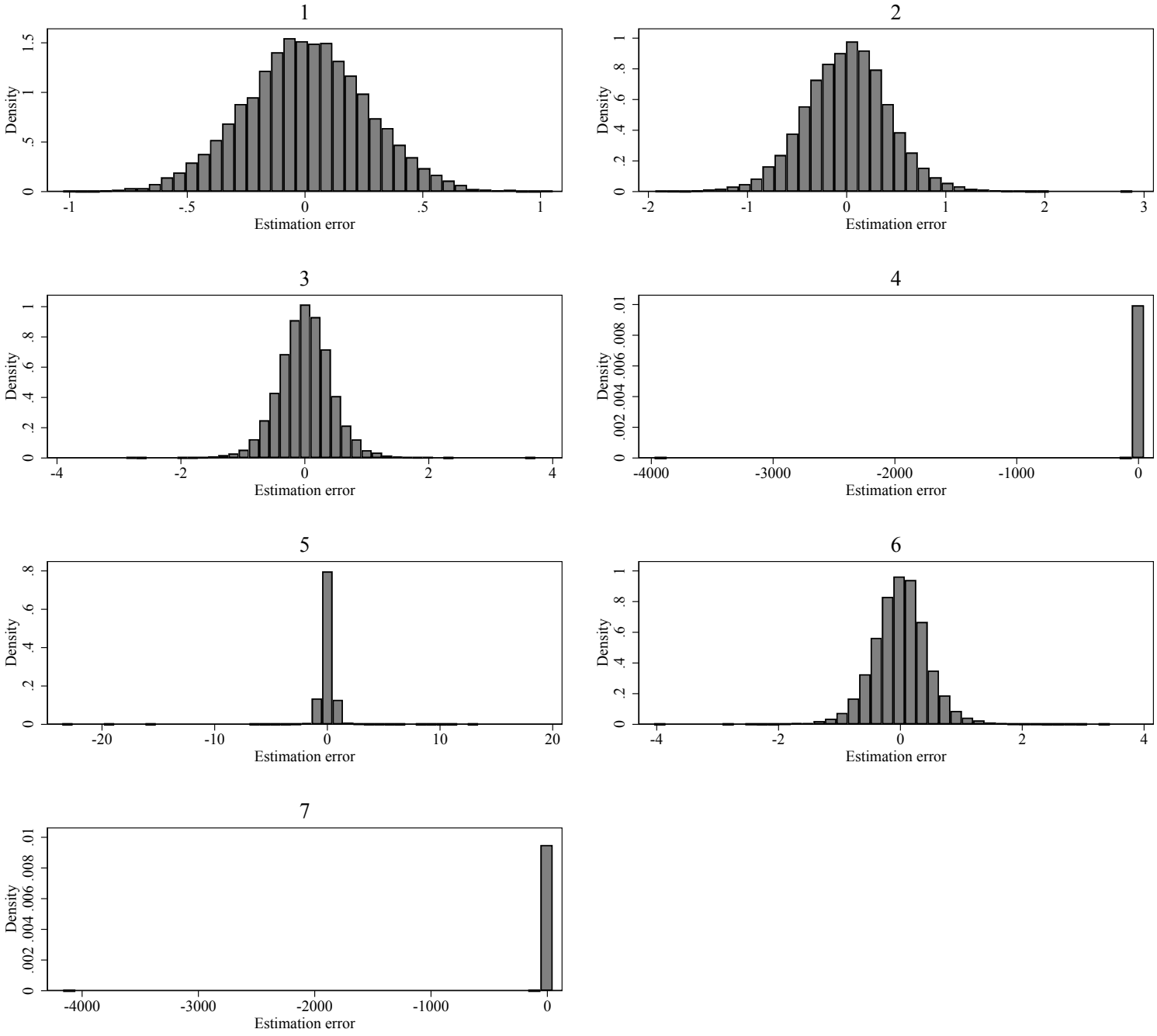
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

Figure B.9: Simulation Results for Design A.1, $\delta = 0.05$, $N = 5{,}000$



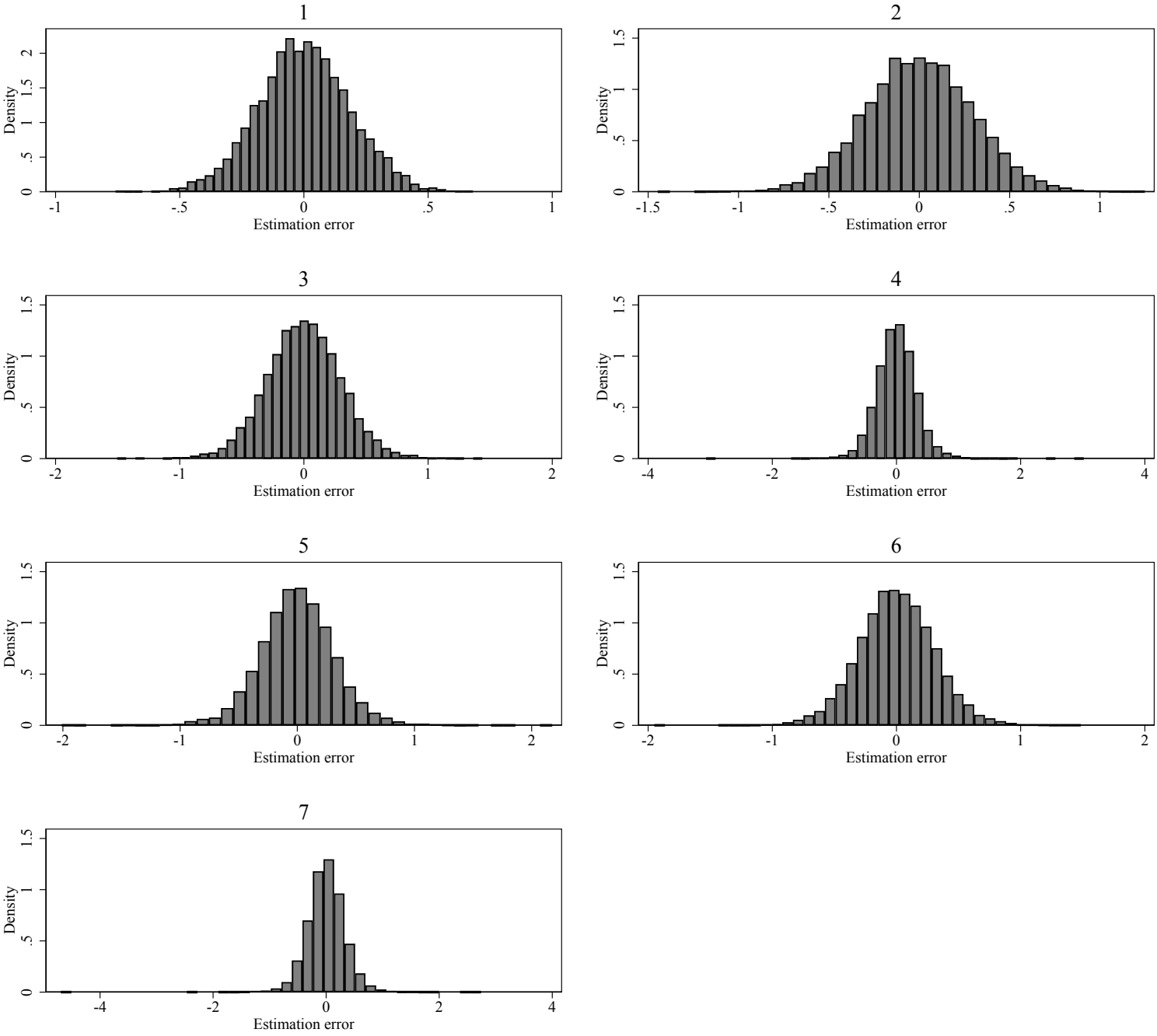
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



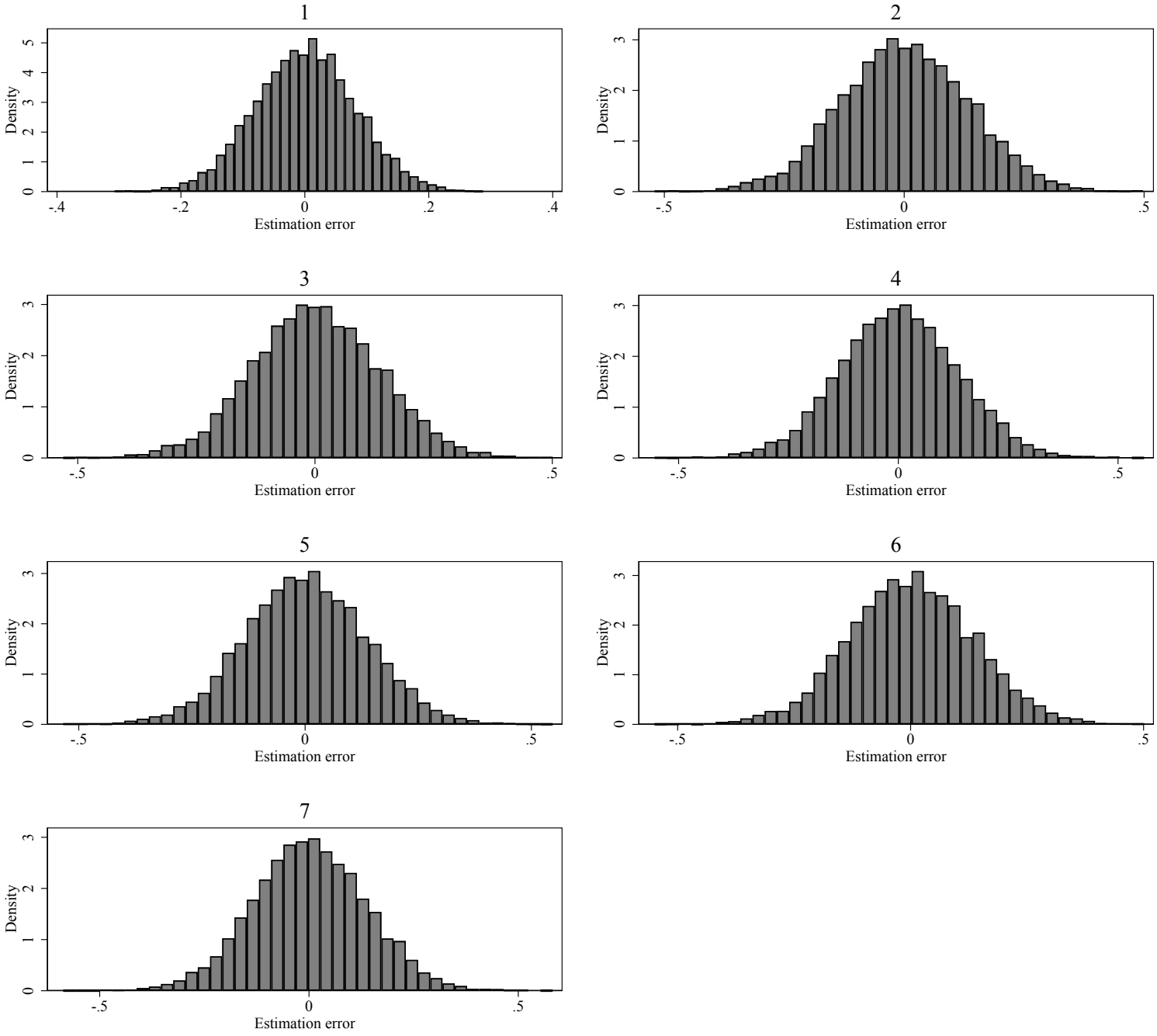
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



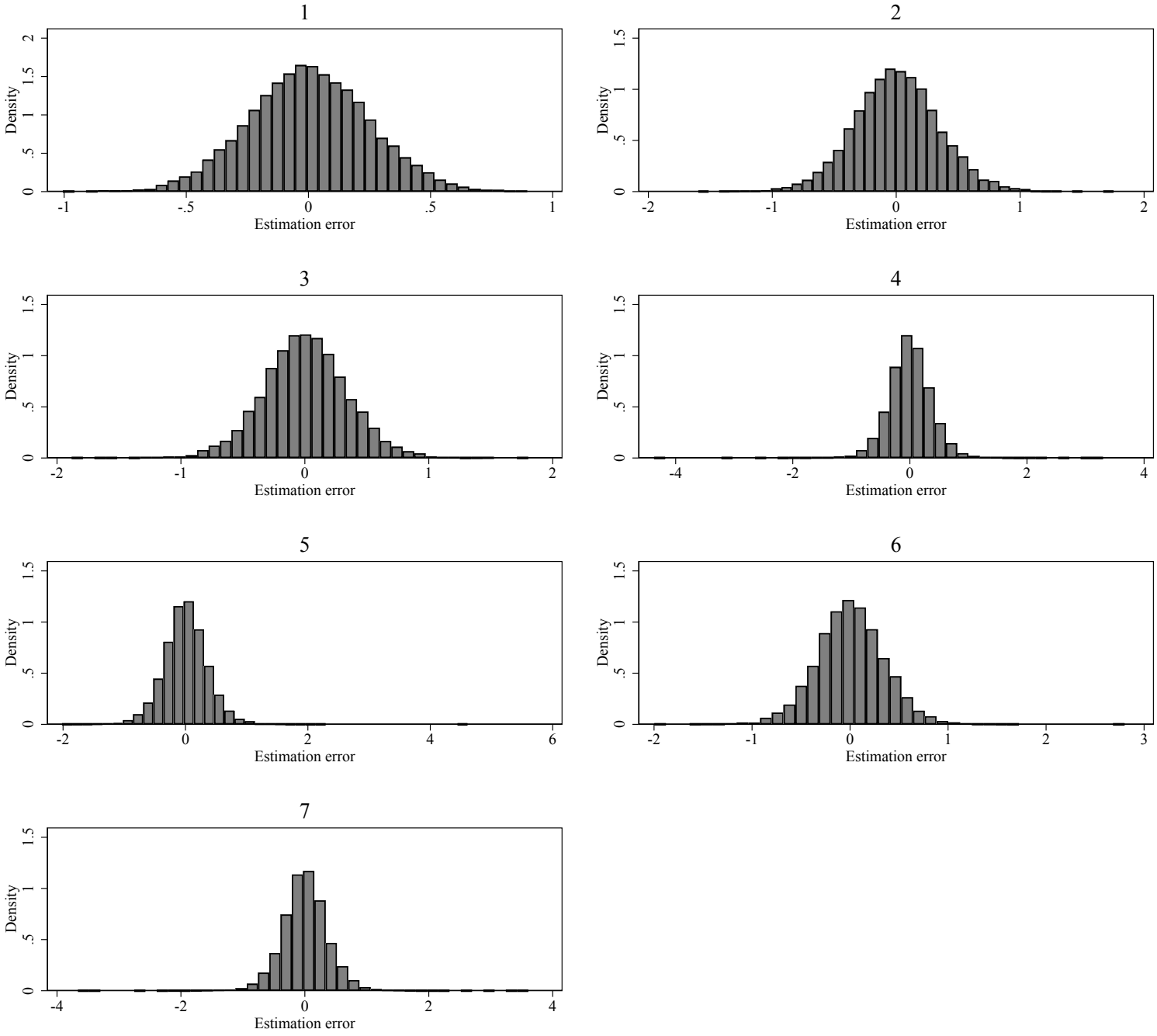
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



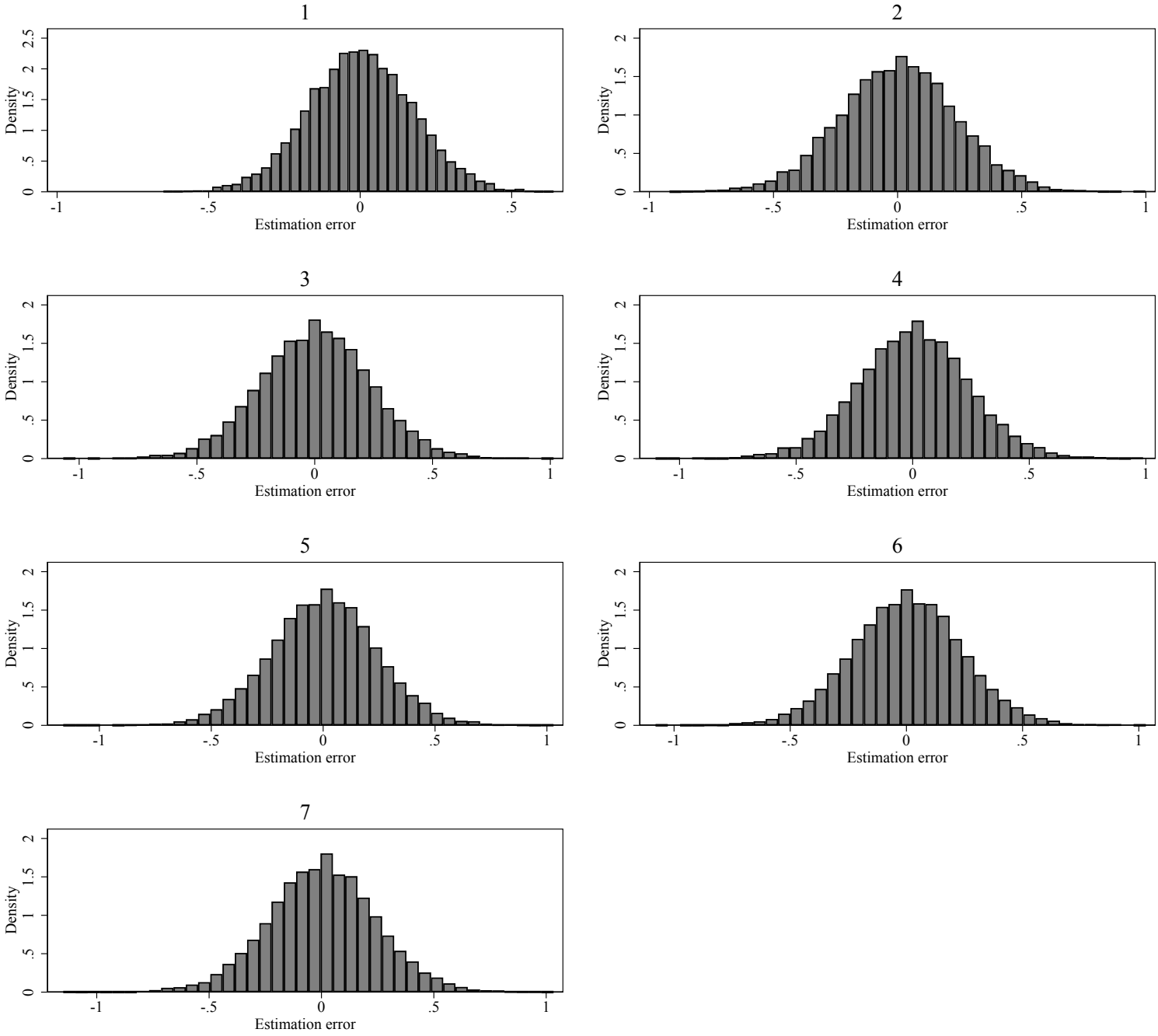
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



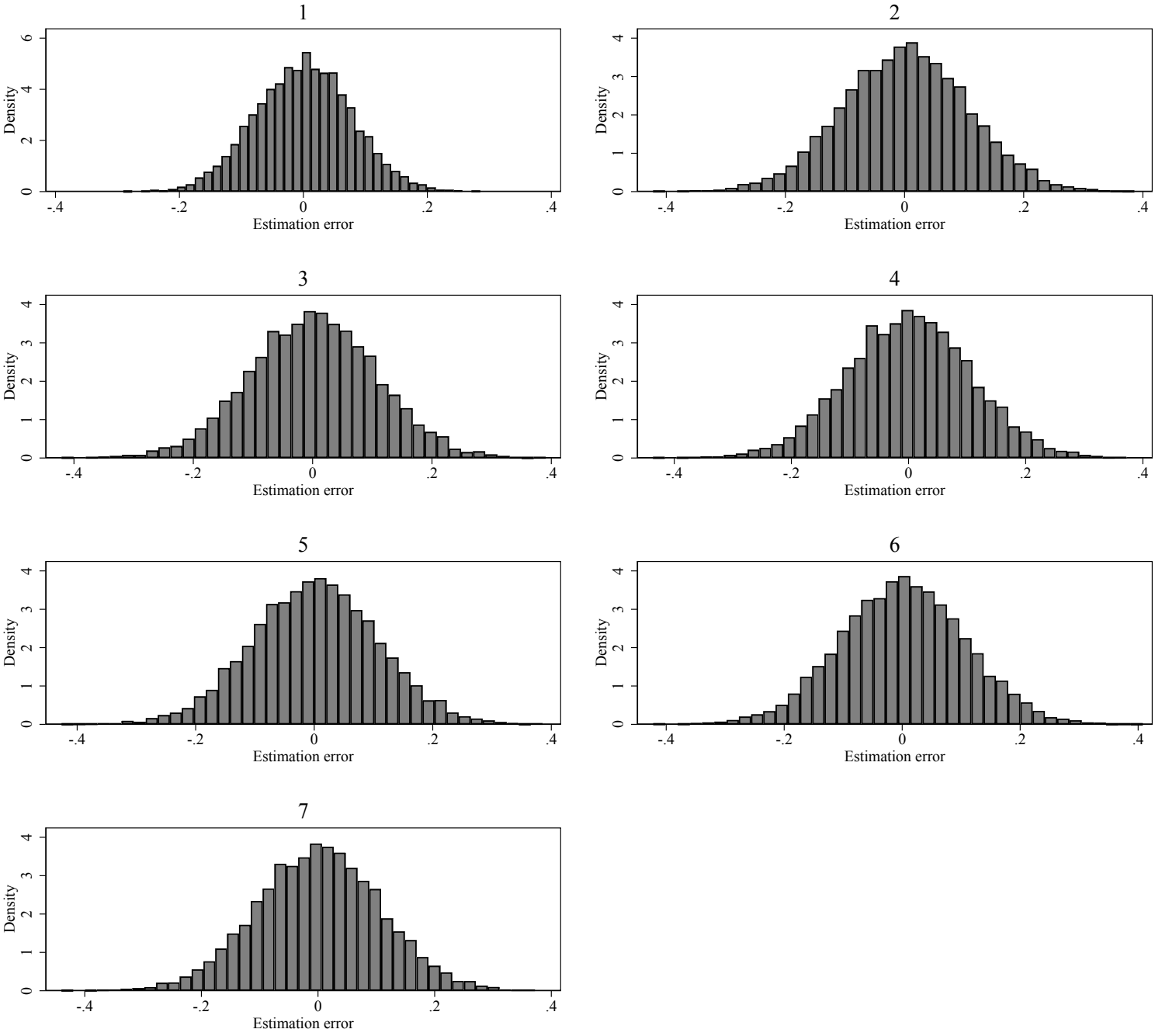
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



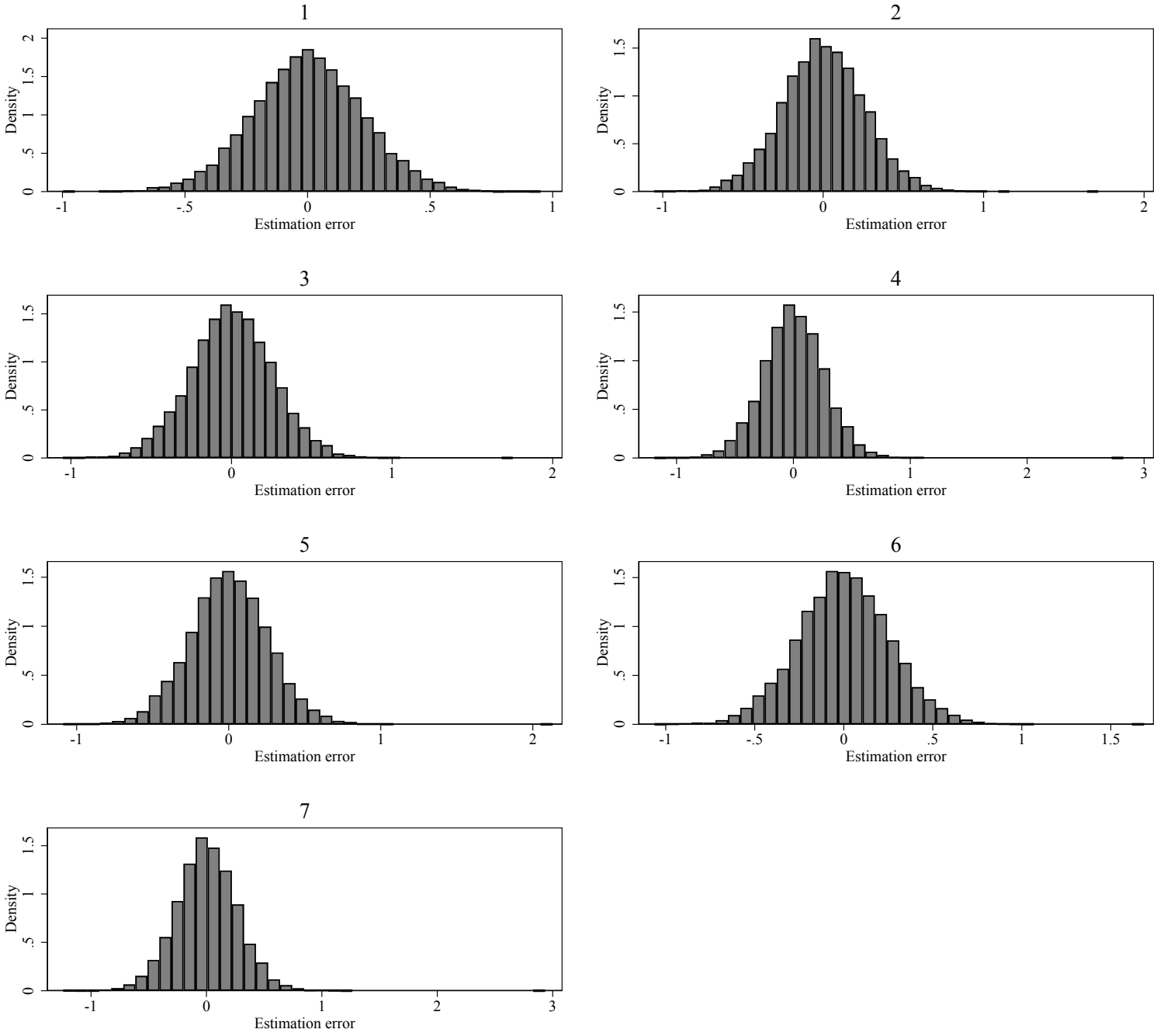
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



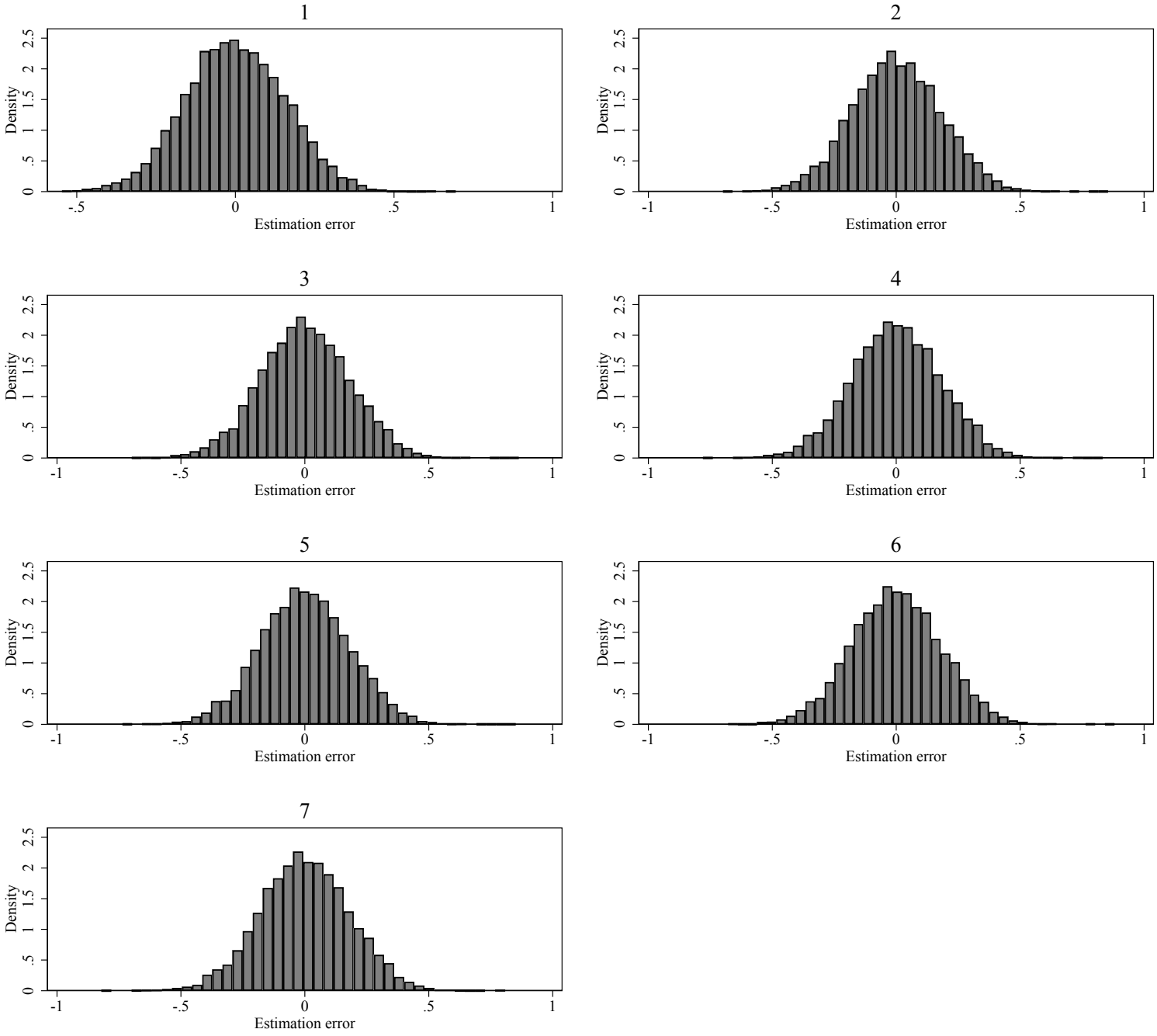
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



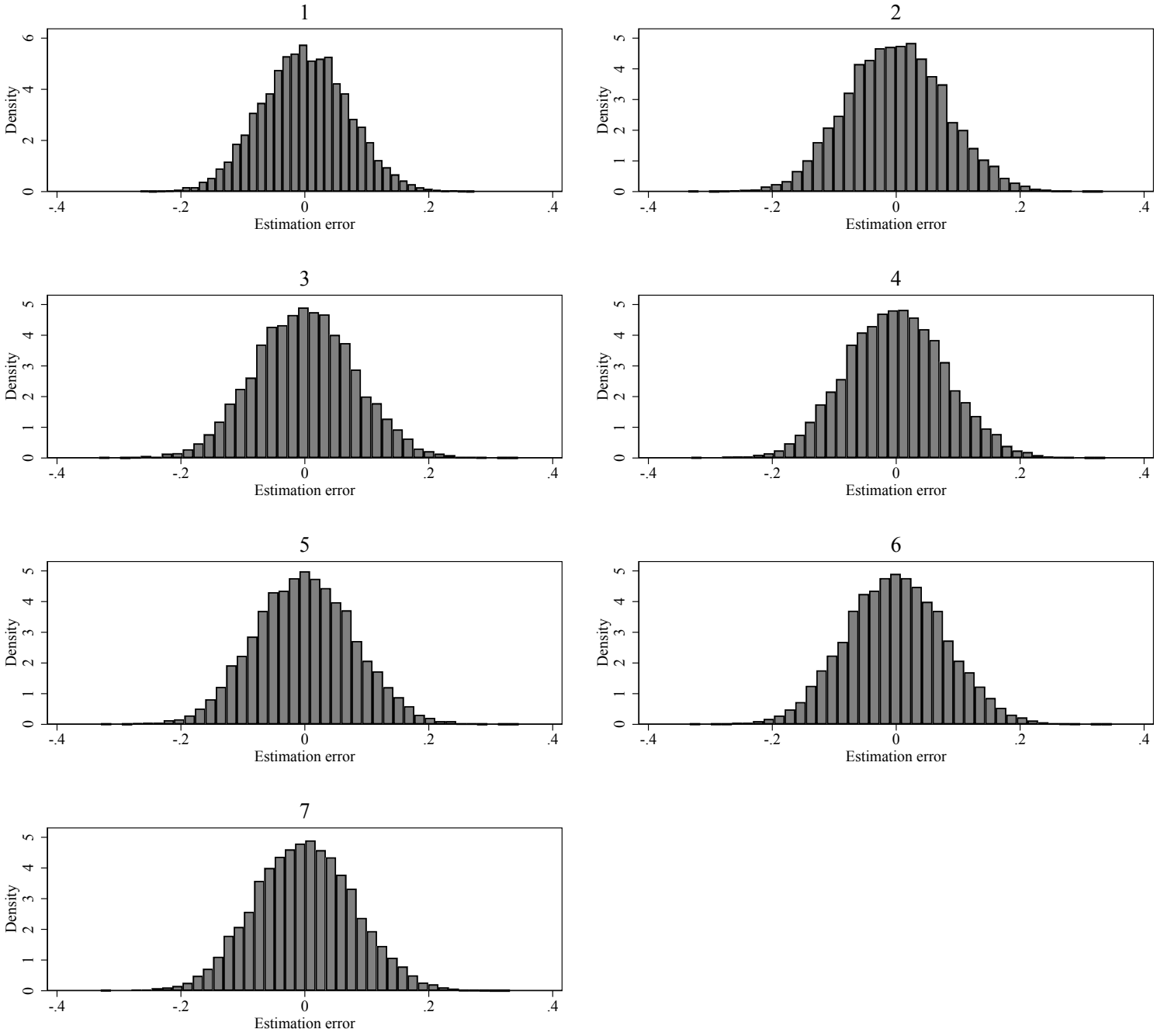
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



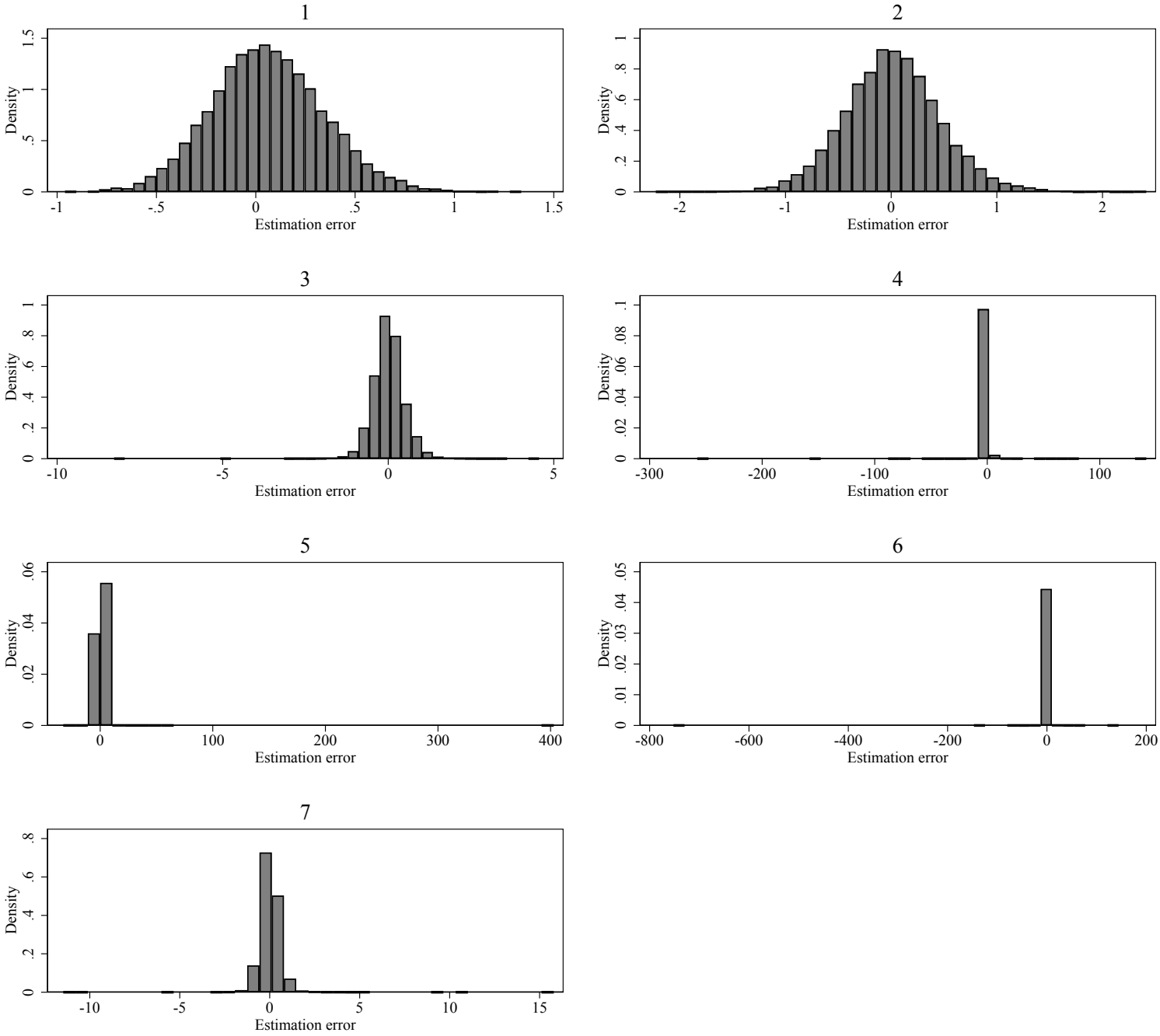
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ $(= \hat{\tau}_{a,1})$. "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

Figure B.18: Simulation Results for Design A.2, $\delta = 0.05$, $N = 5{,}000$



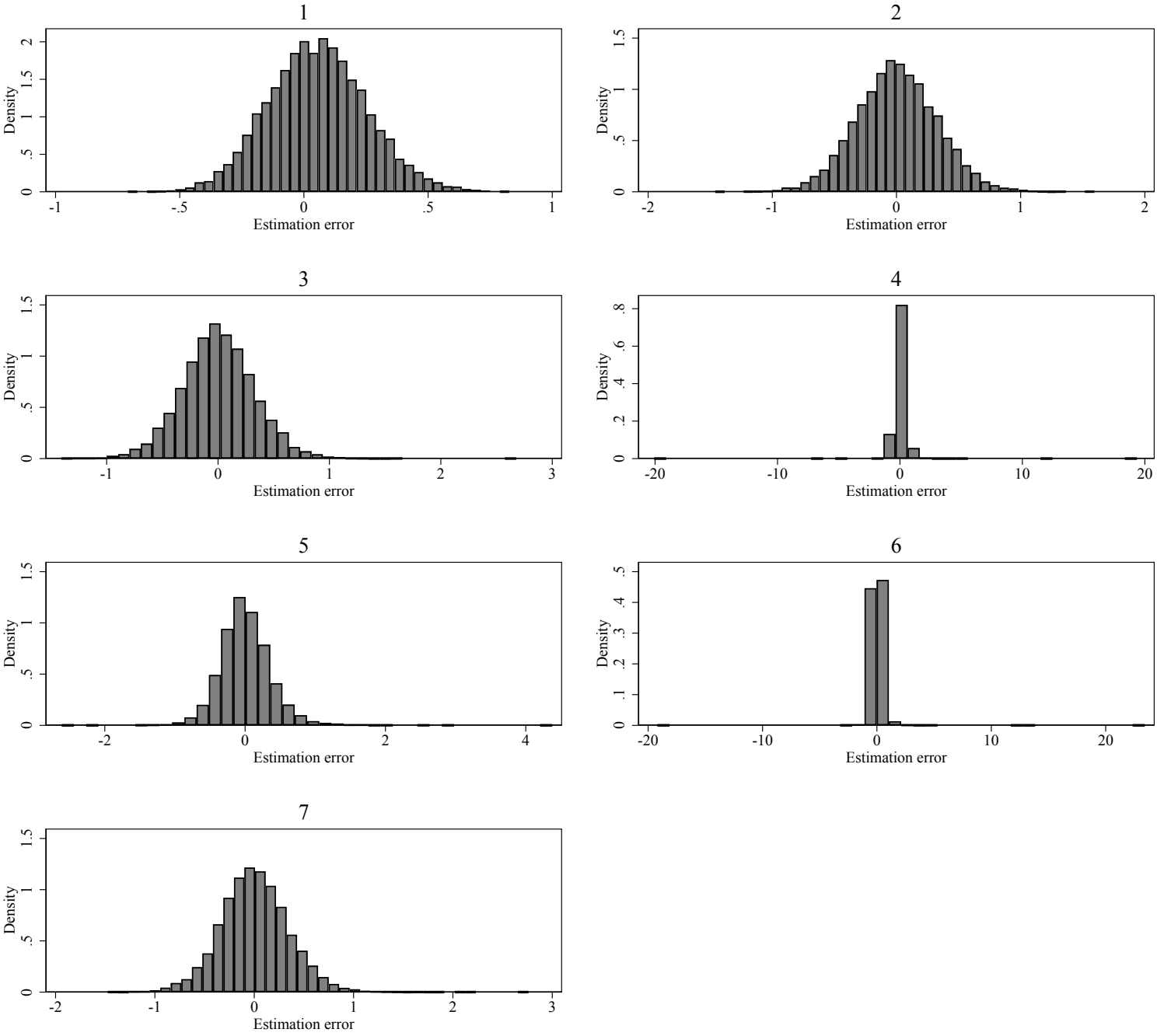
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



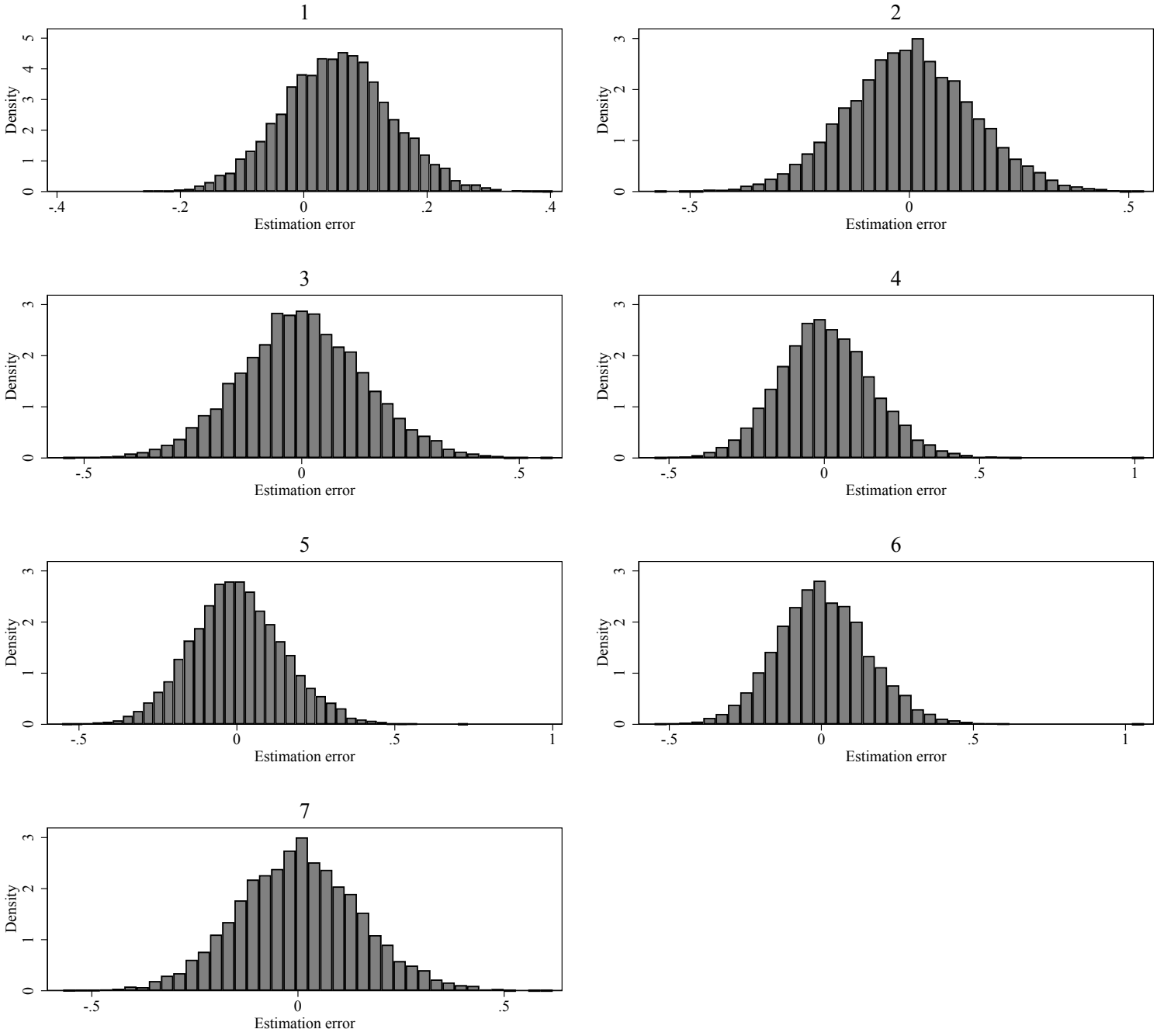
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



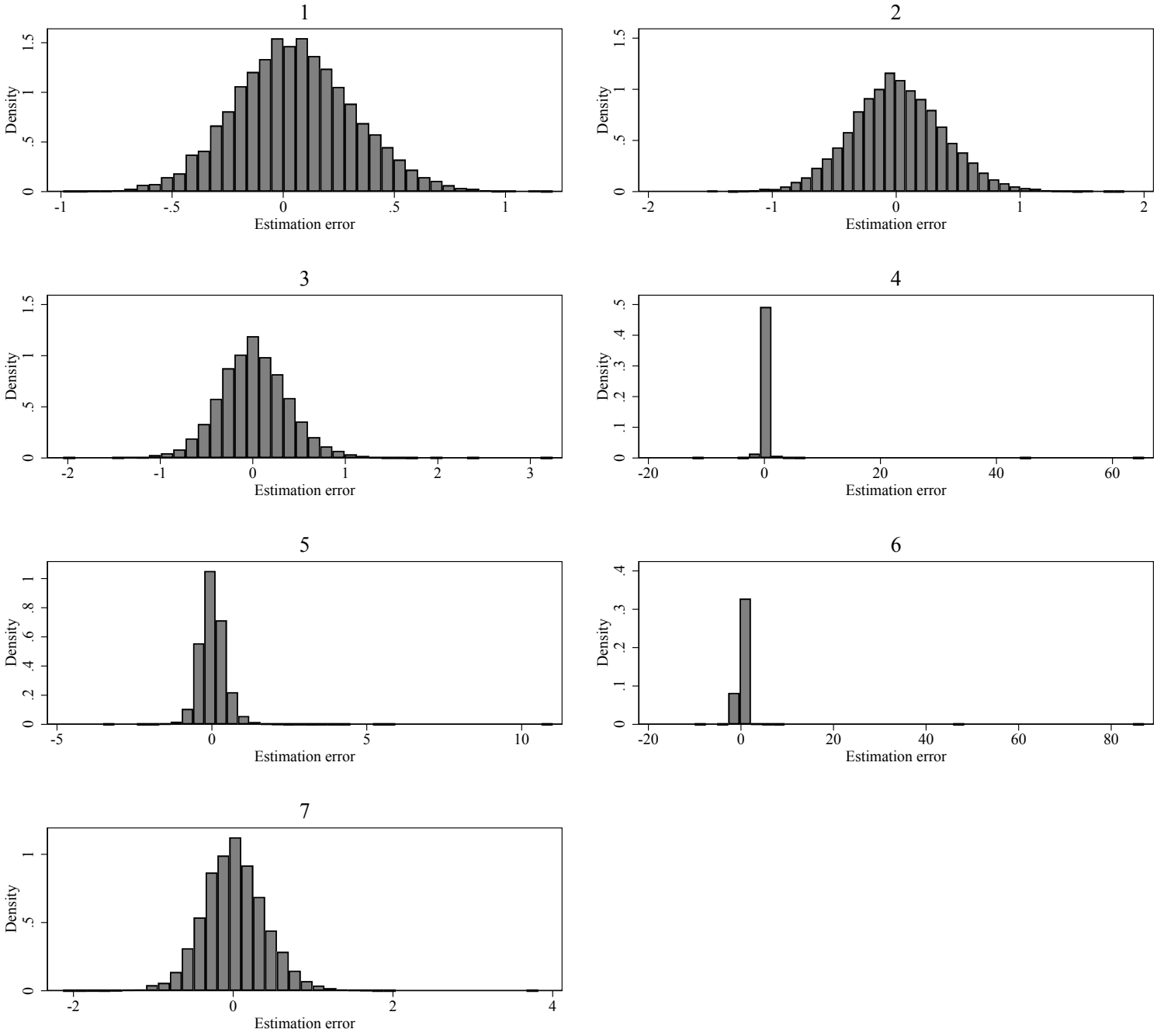
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



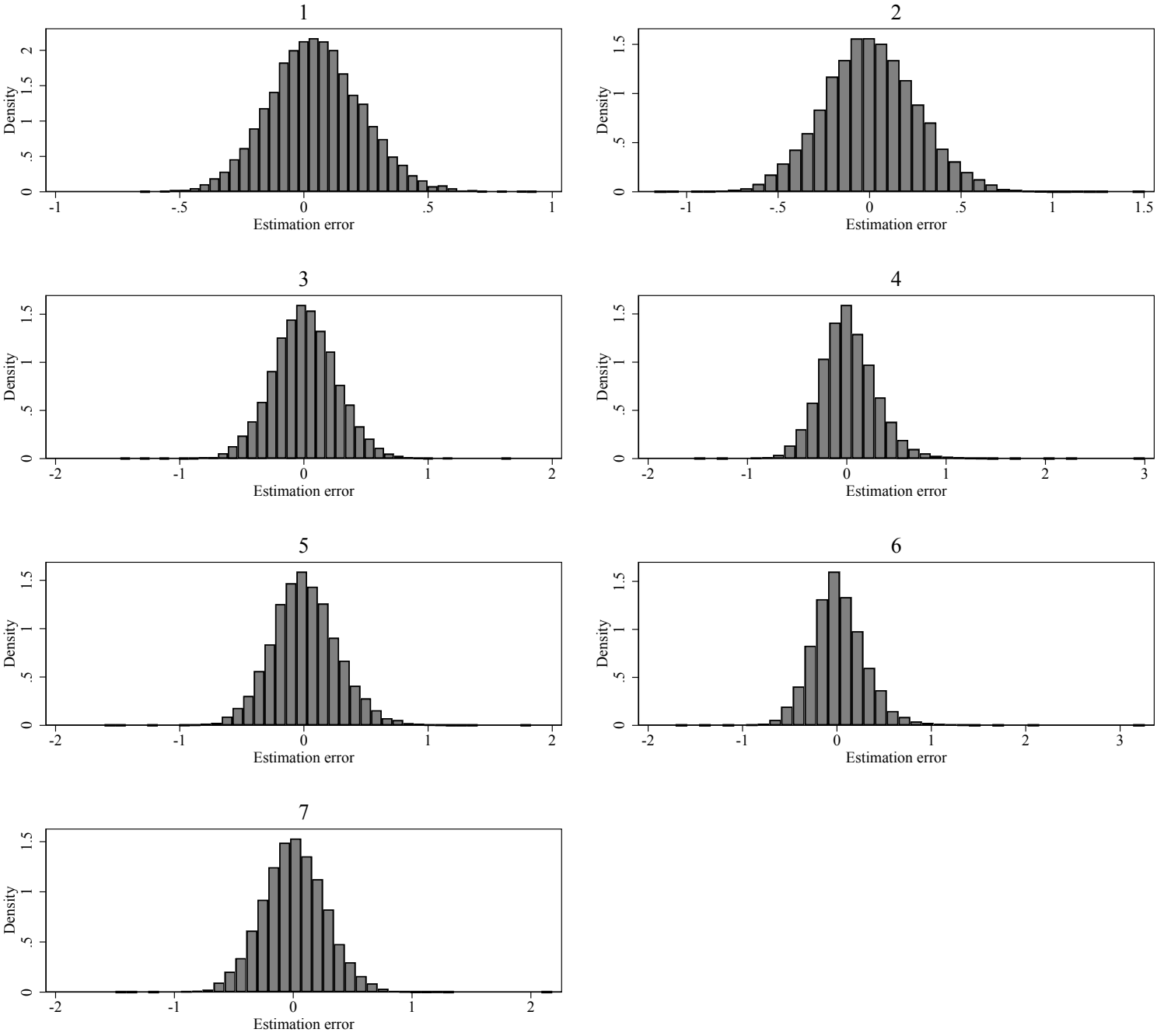
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



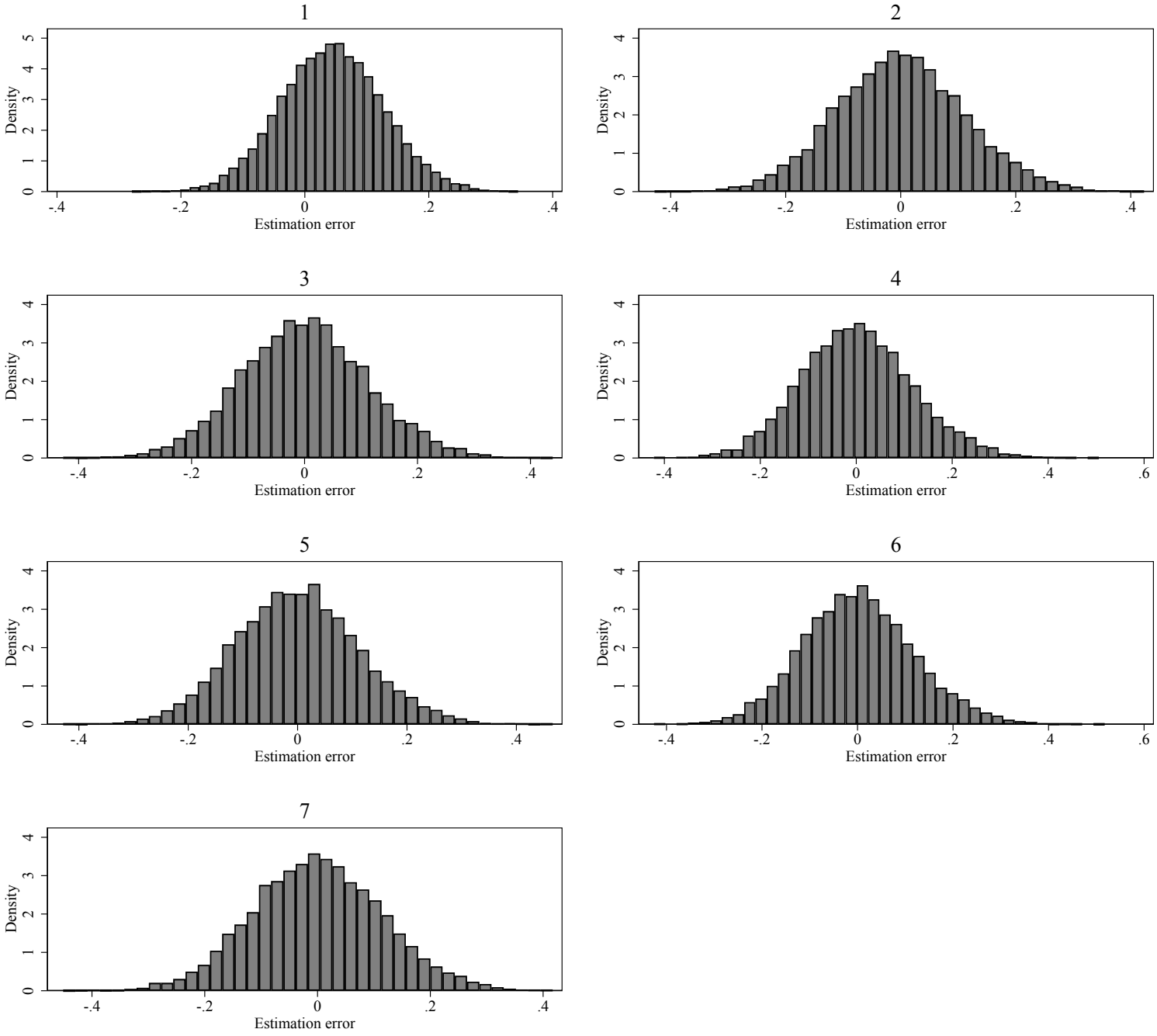
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



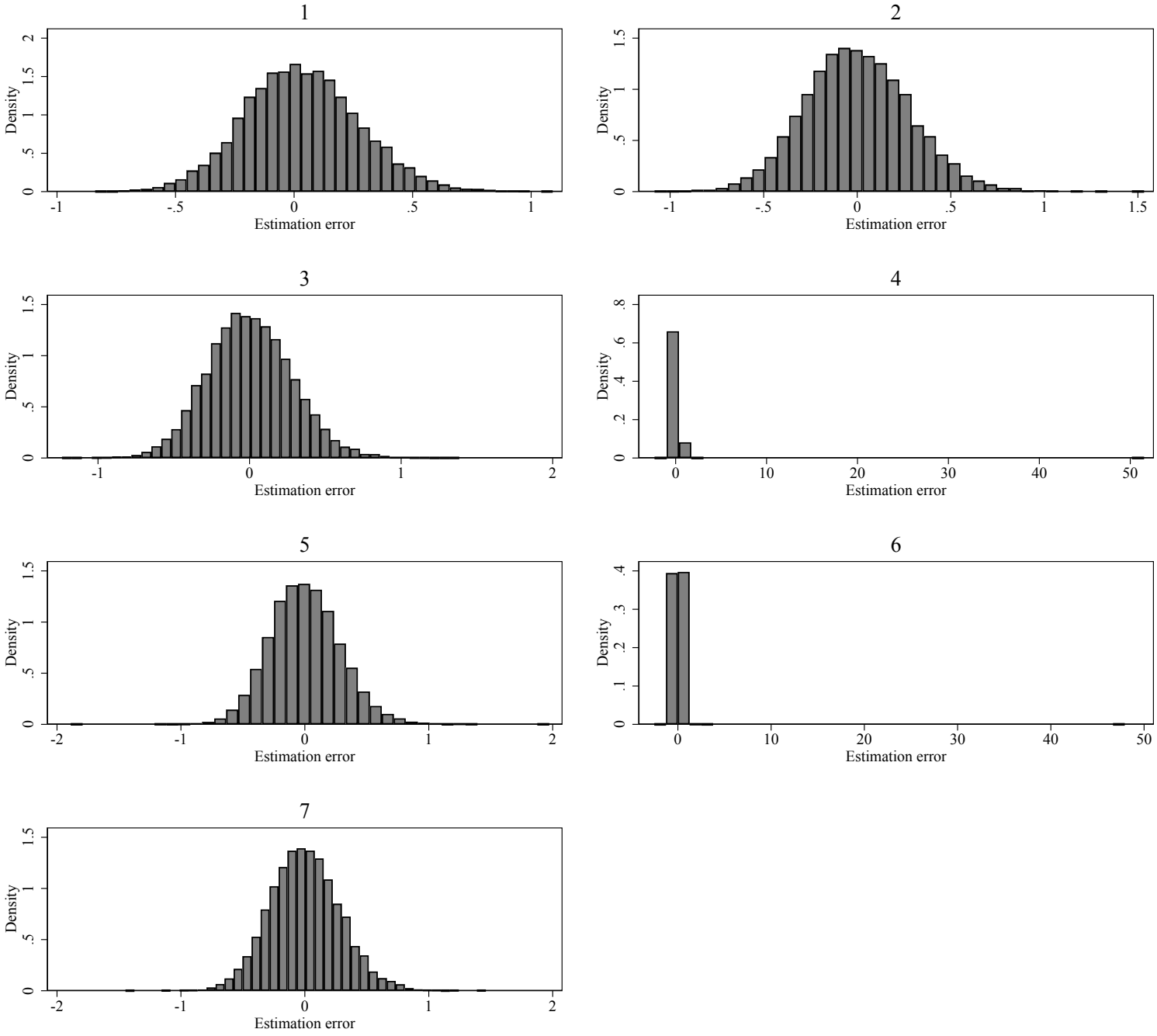
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



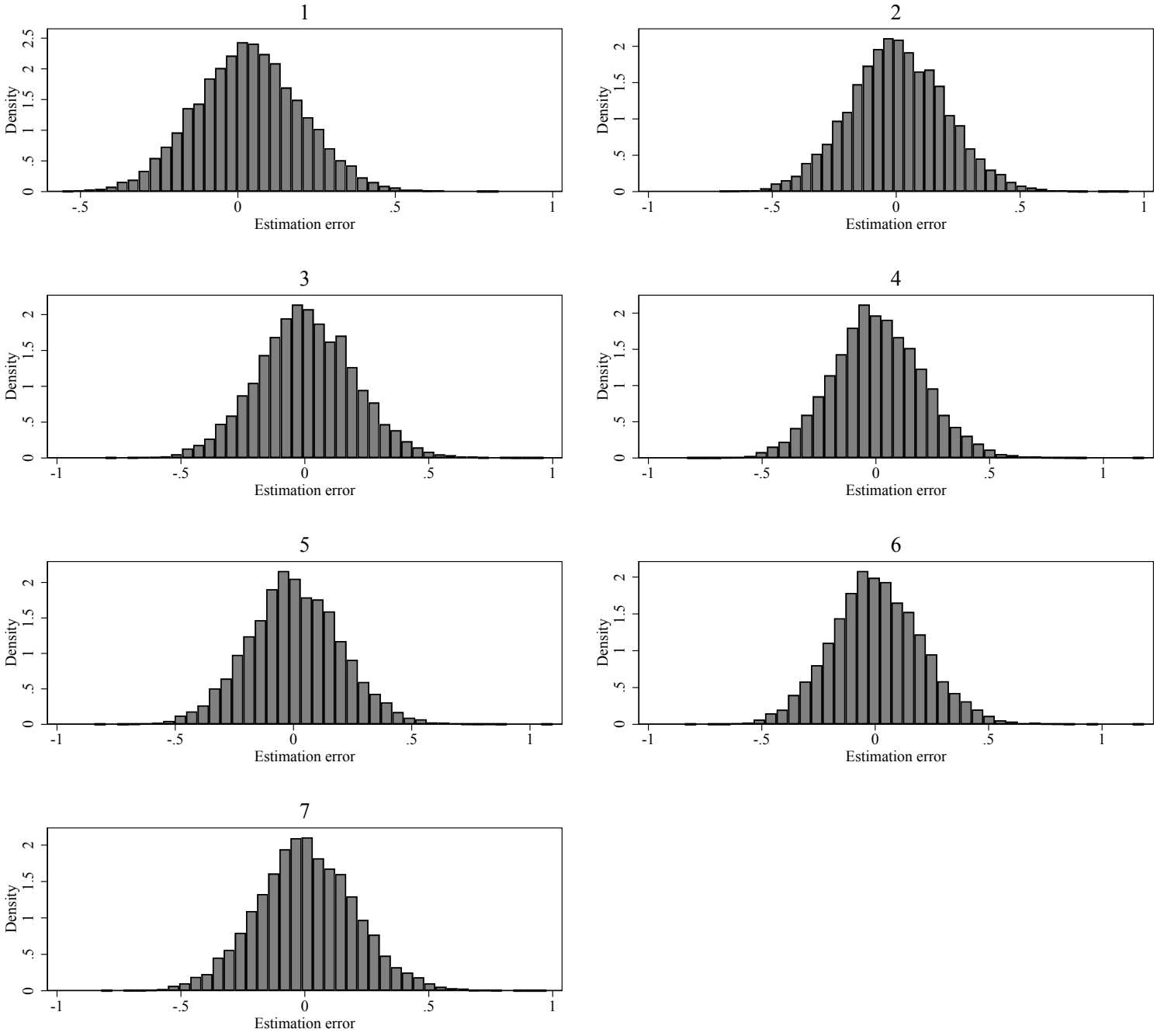
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

Figure B.25: Simulation Results for Design B, $\delta = 0.05$, $N = 500$



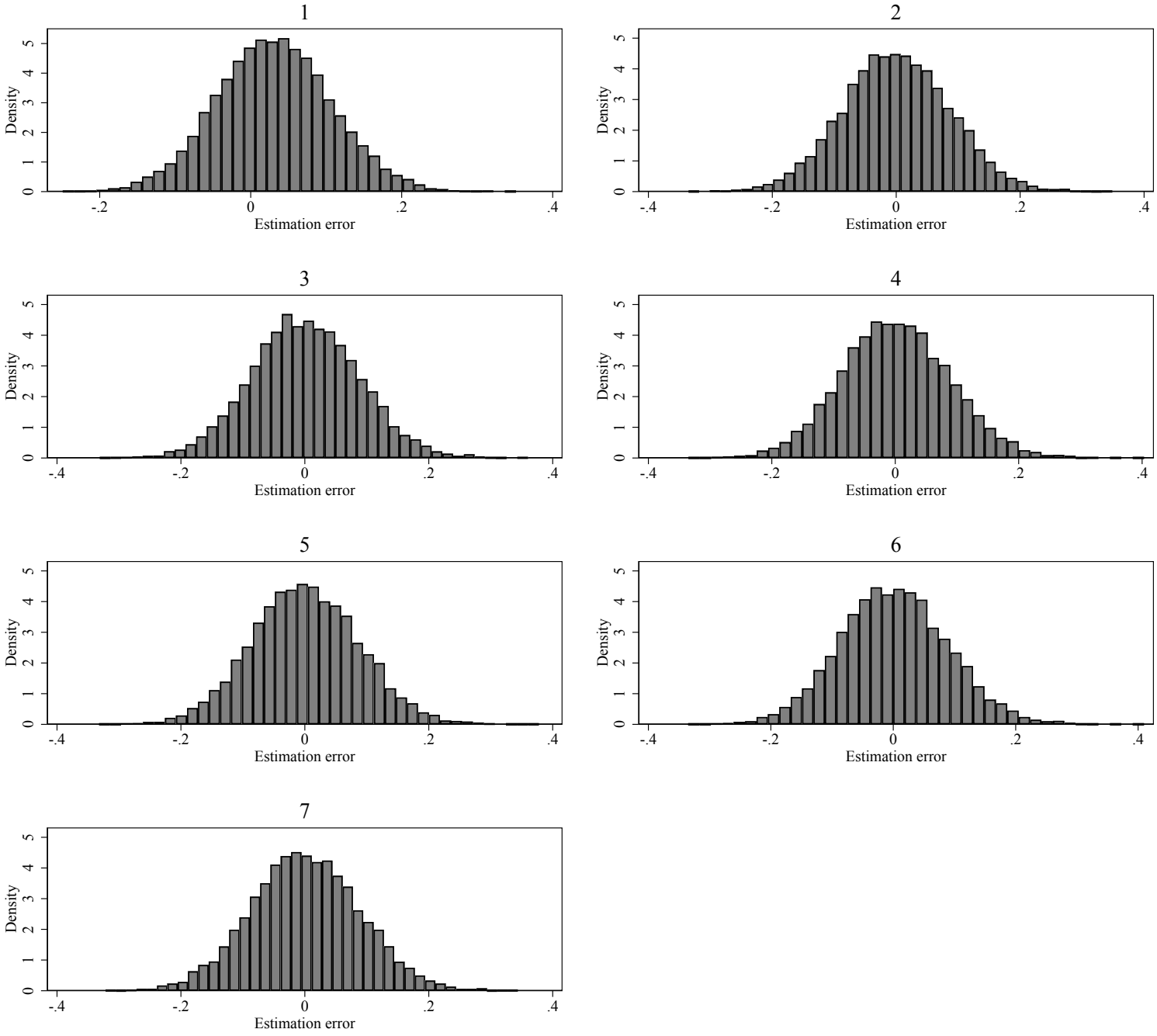
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



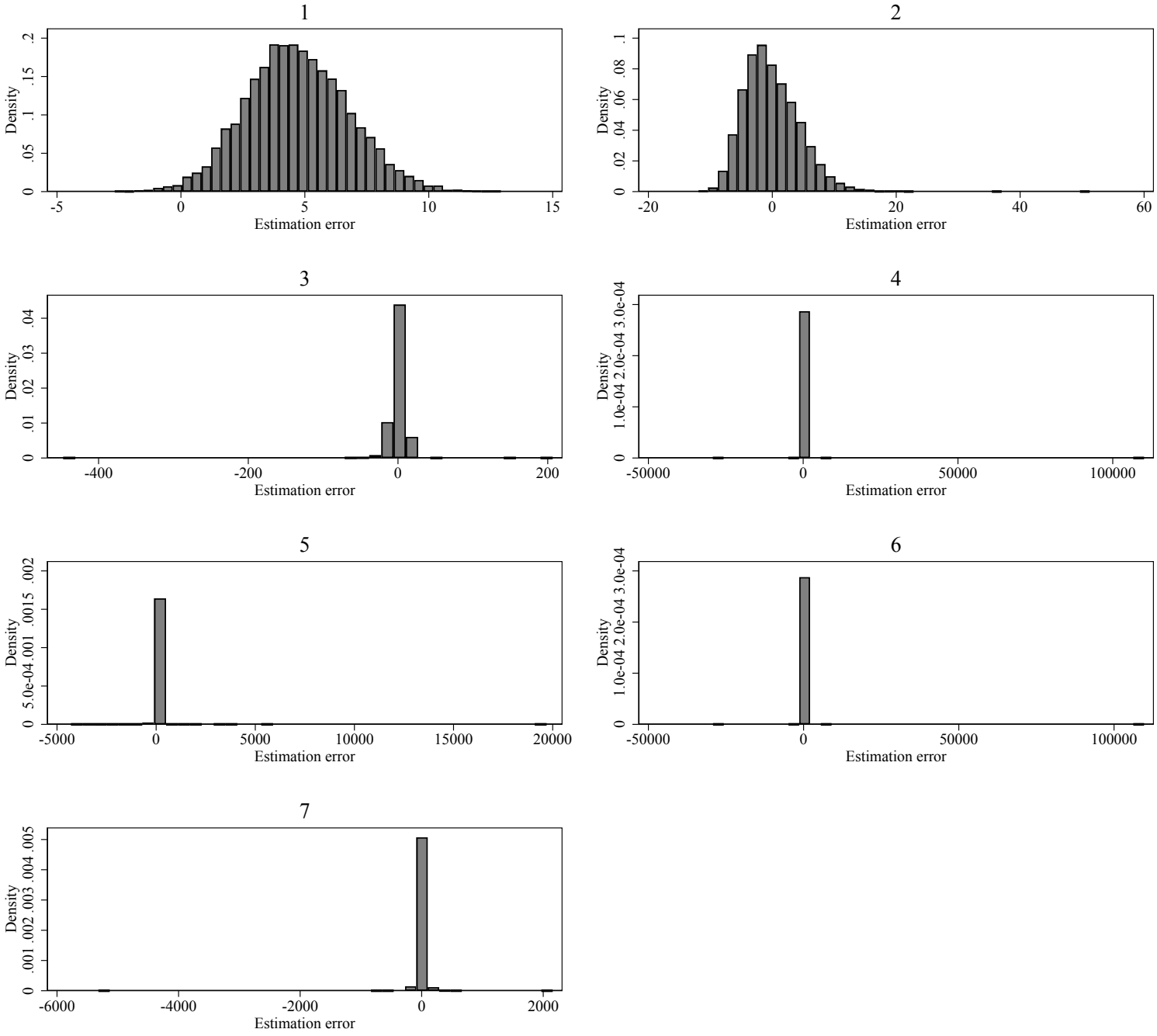
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



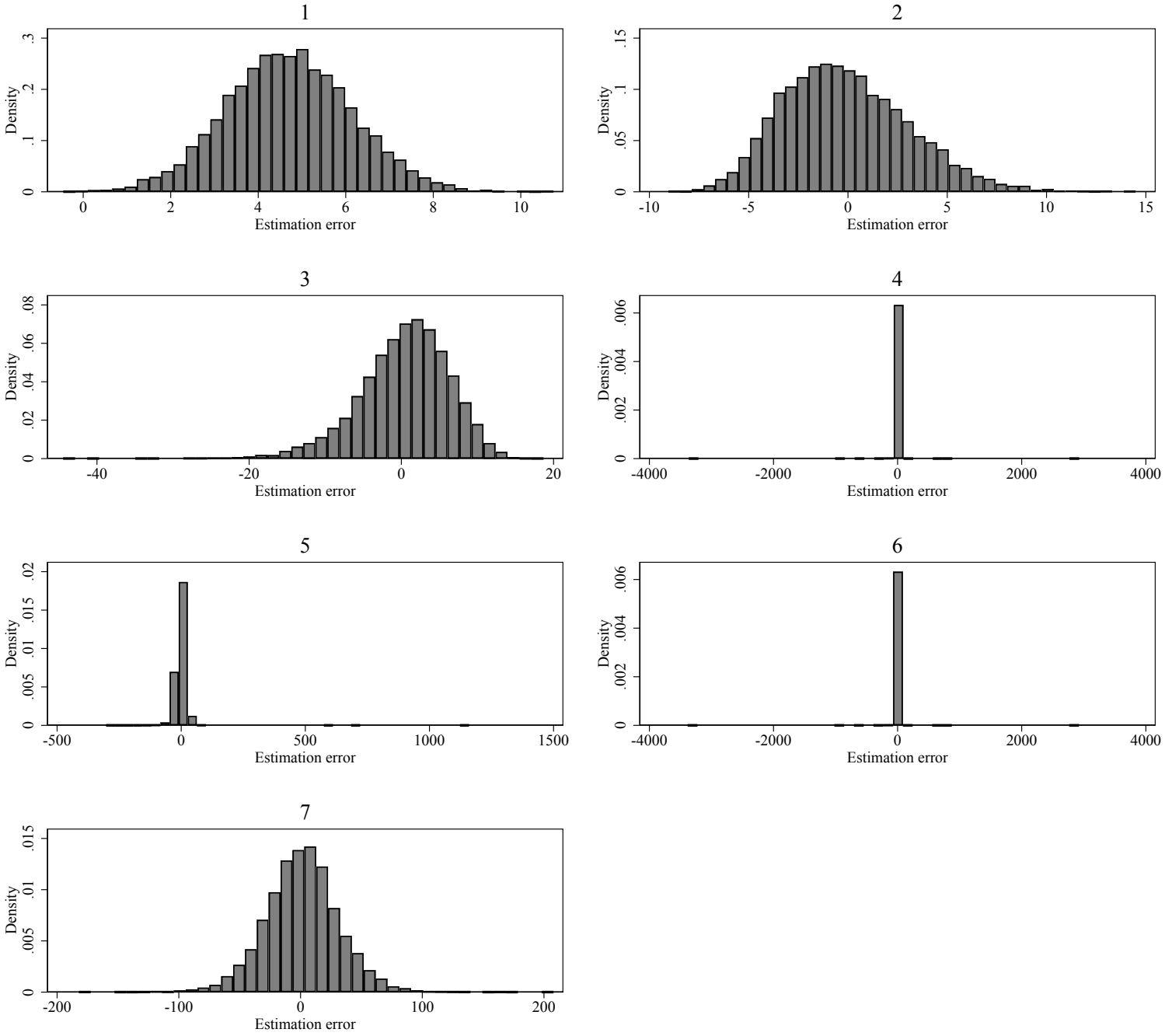
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



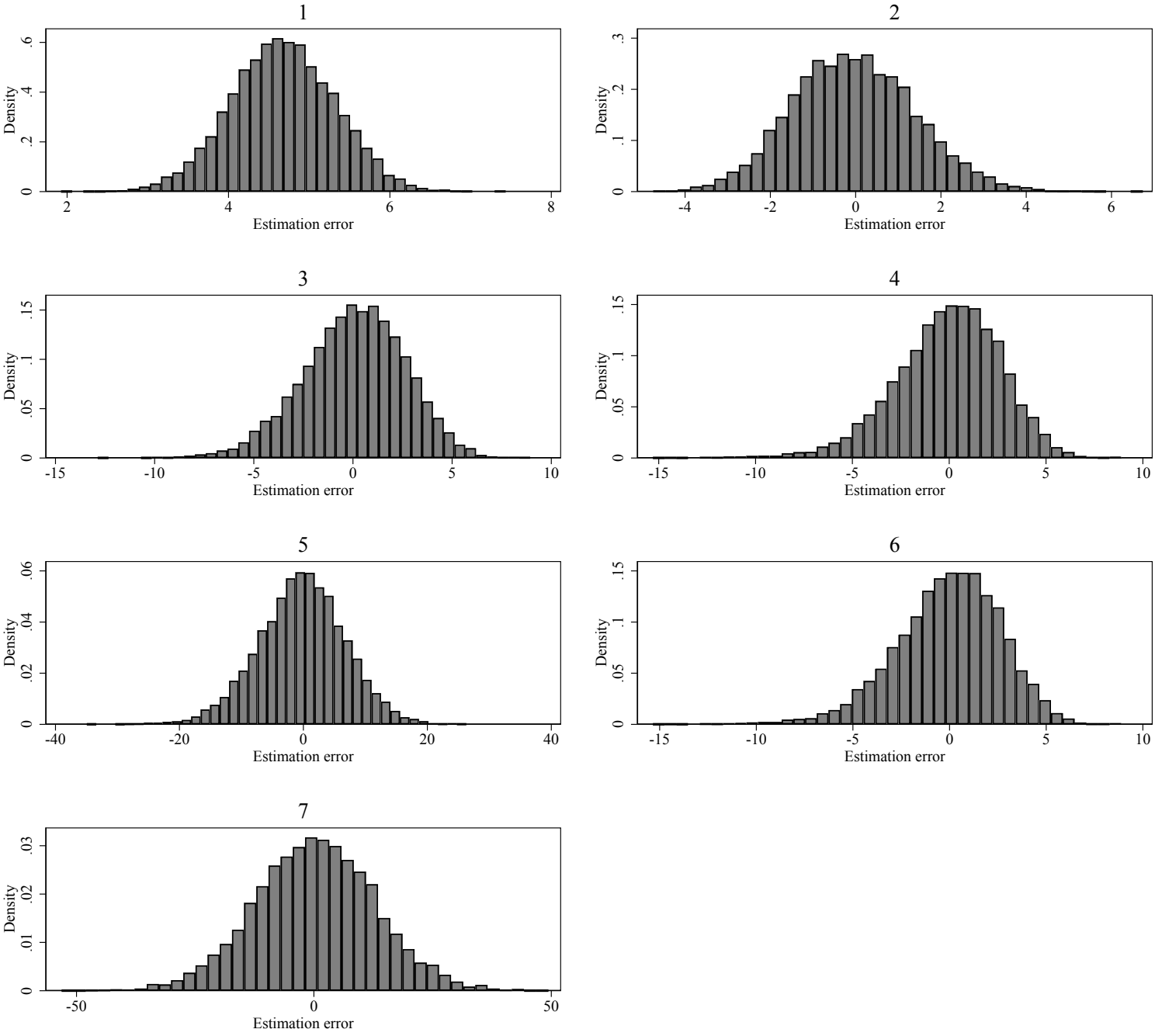
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



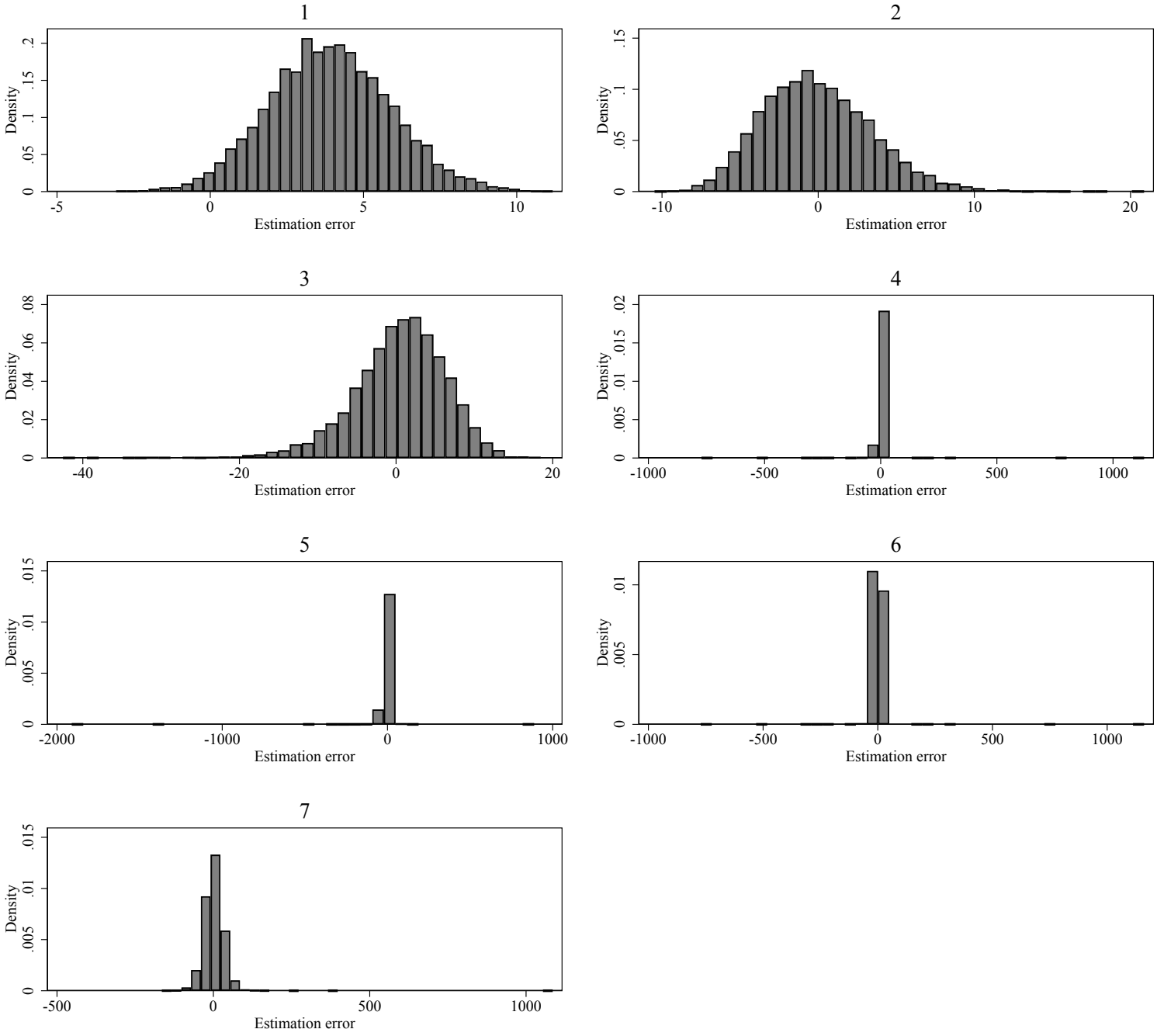
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



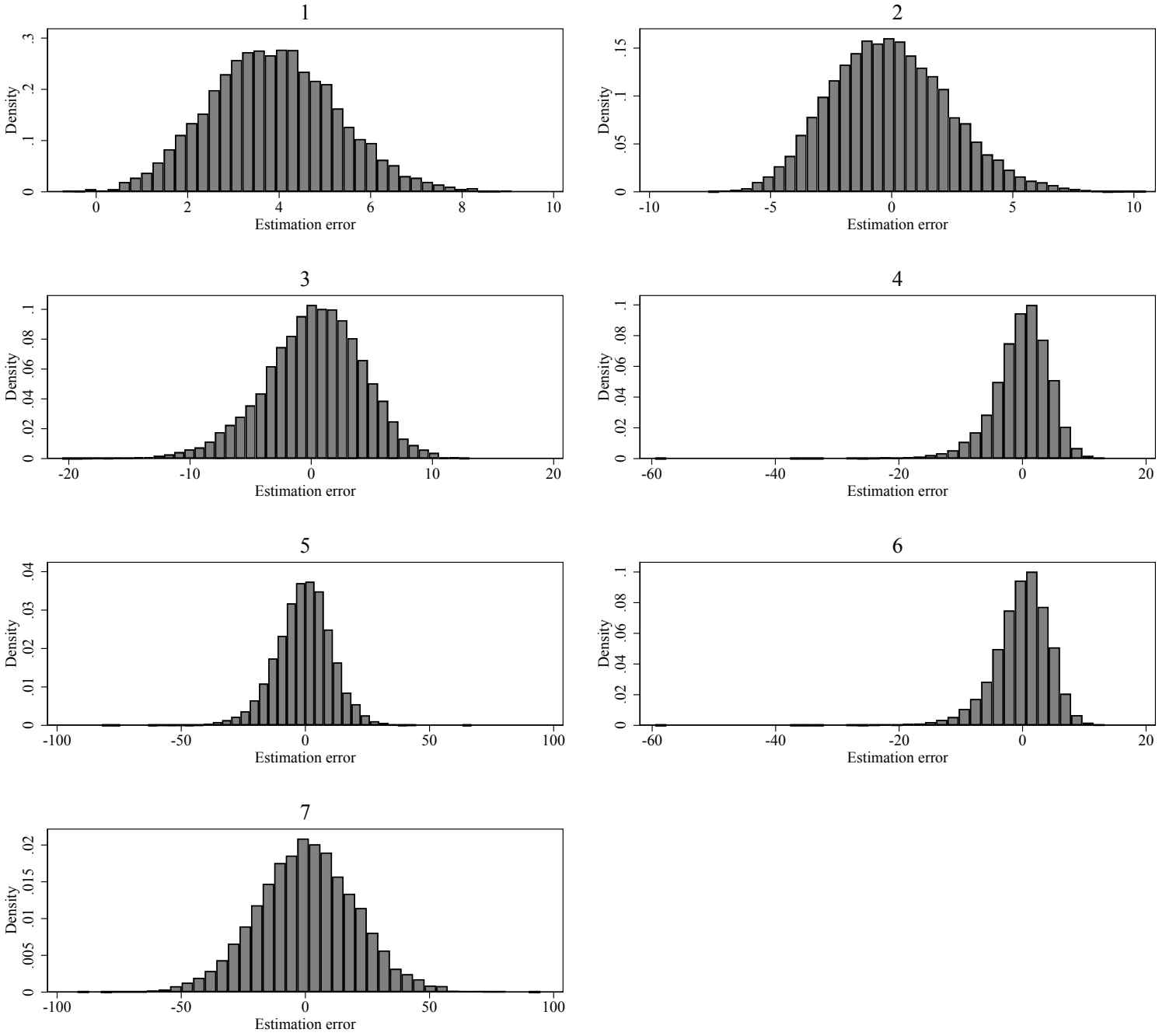
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ $(= \hat{\tau}_{a,1})$. "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



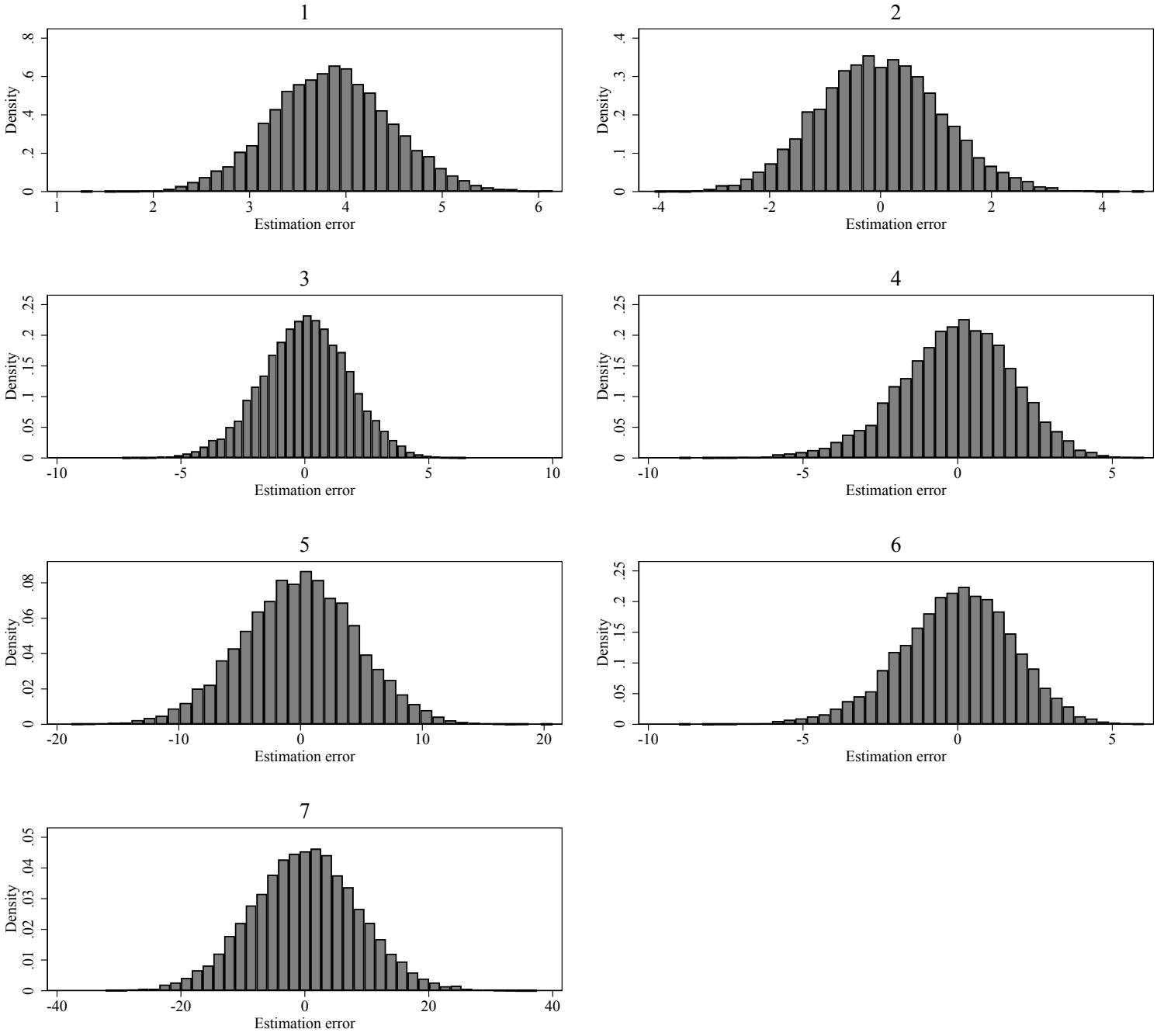
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



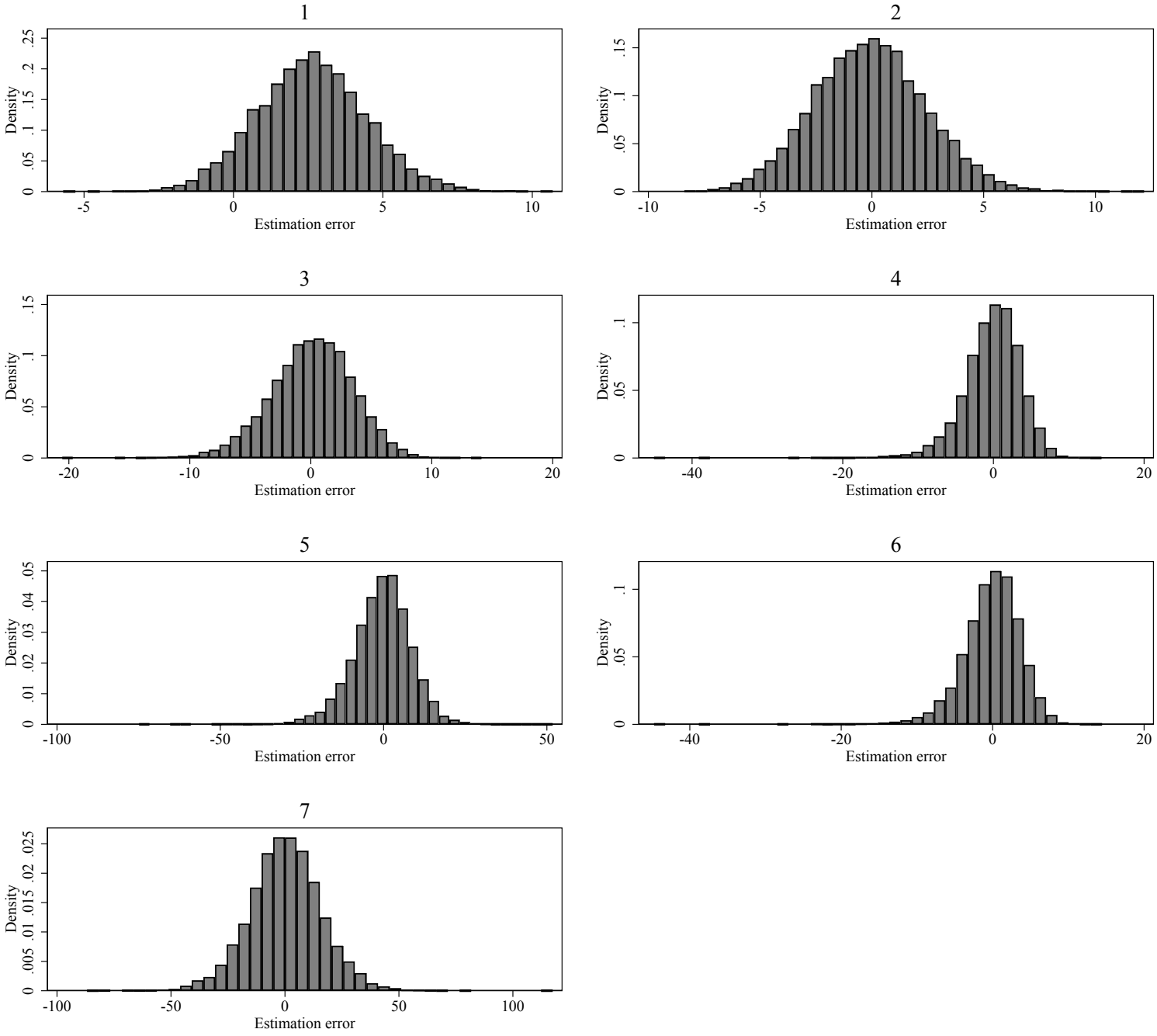
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ $(= \hat{\tau}_{a,1})$. "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



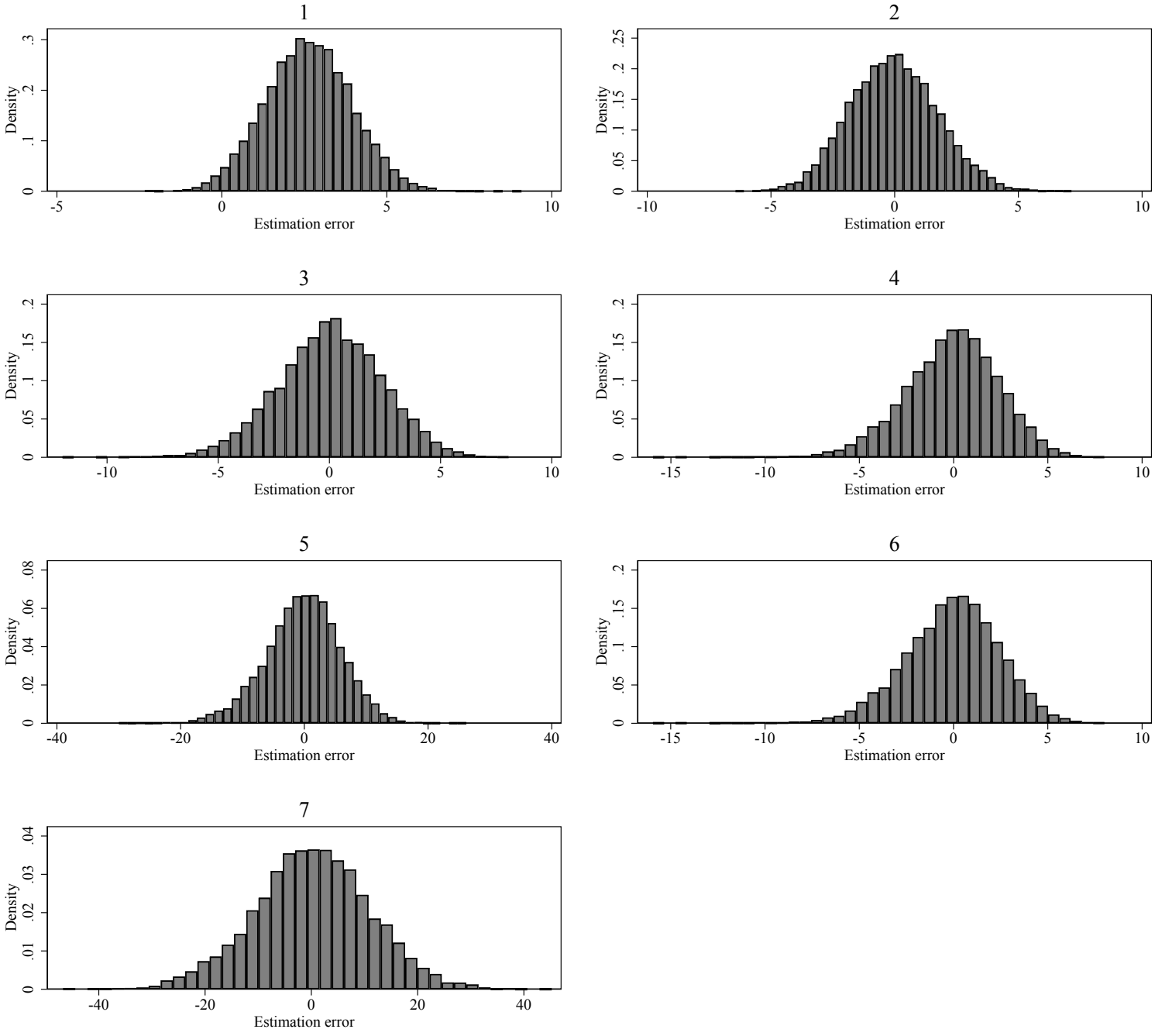
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



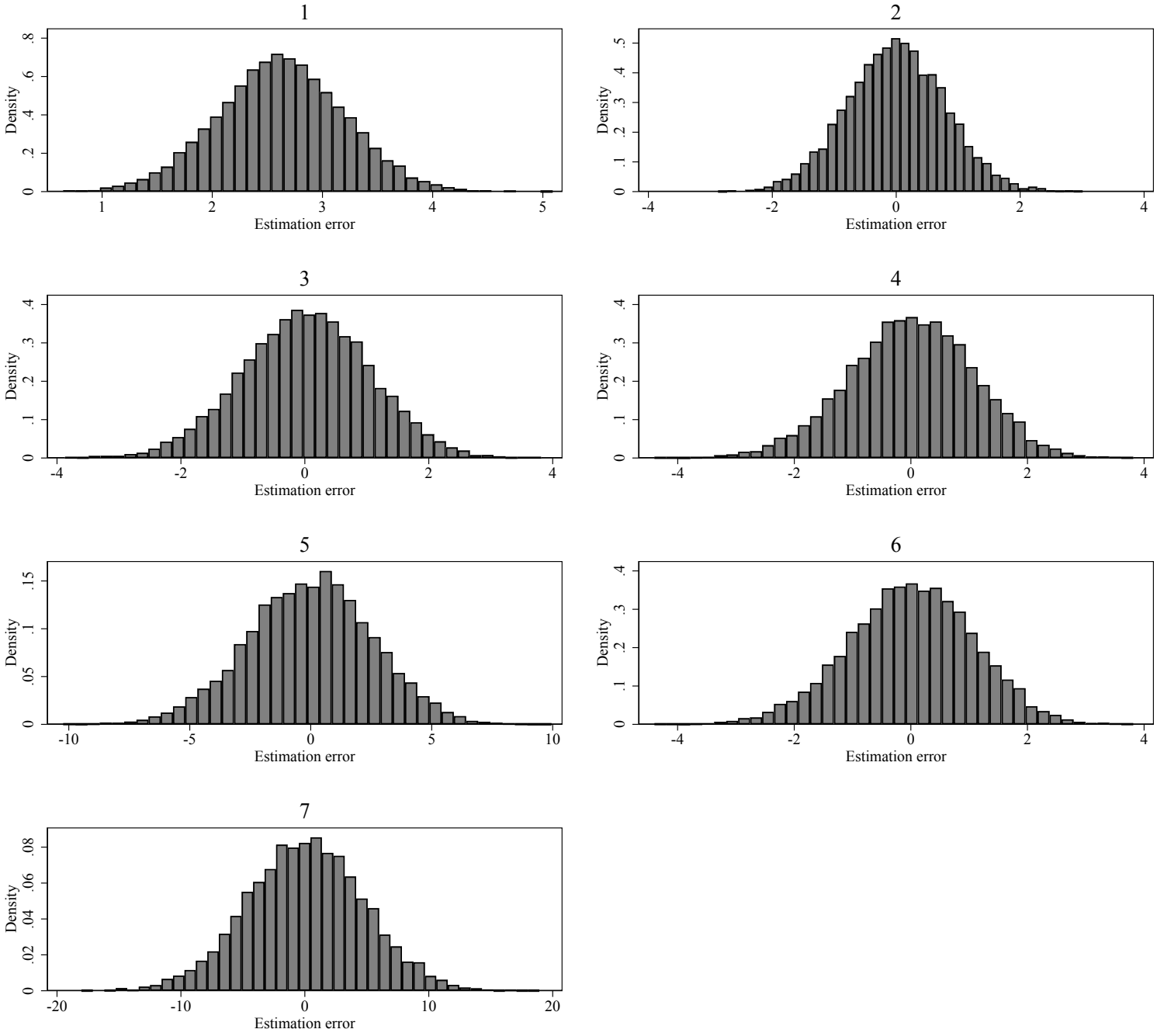
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

Figure B.35: Simulation Results for Design C, $\delta = 0.05$, $N = 1{,}000$



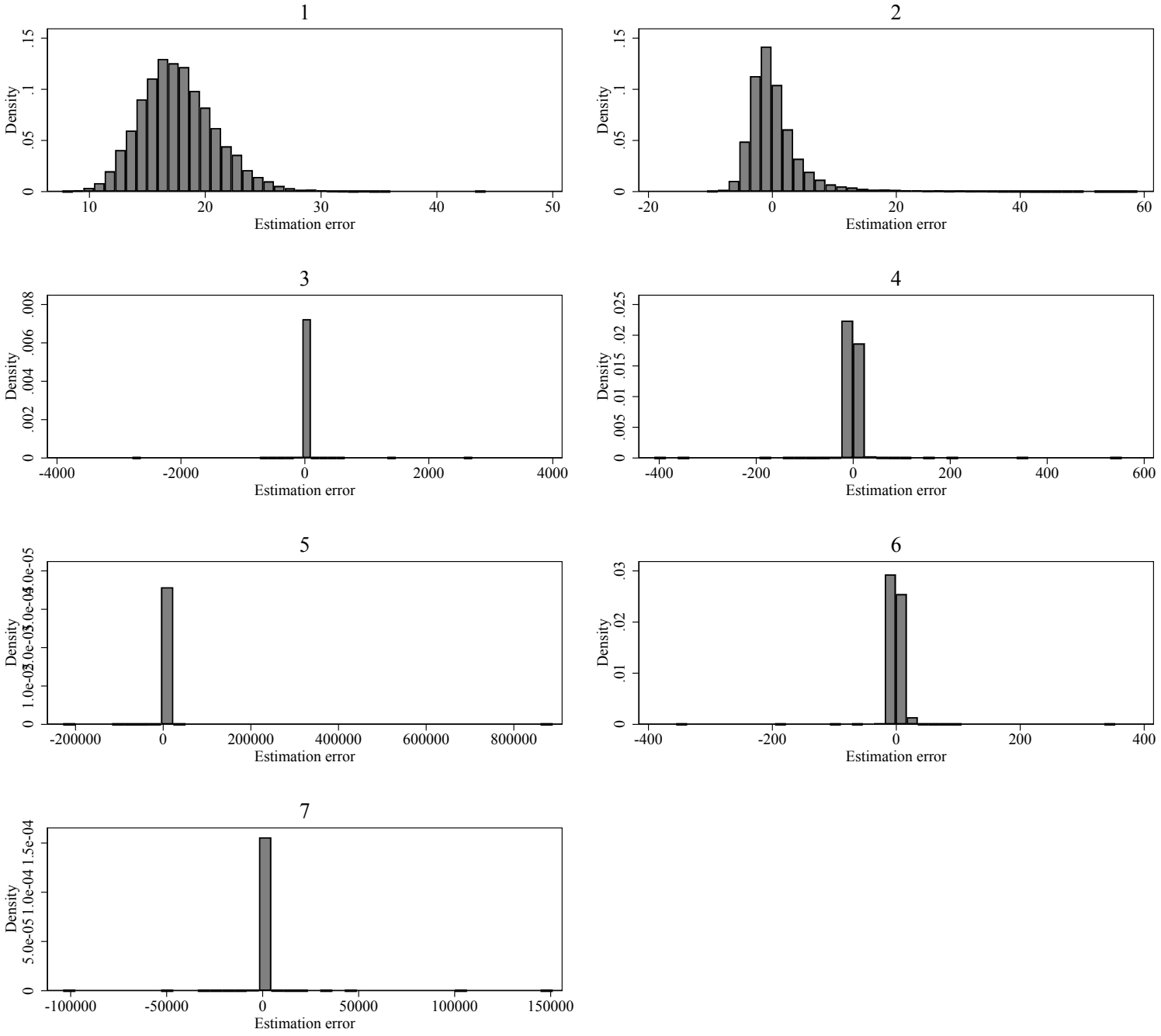
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



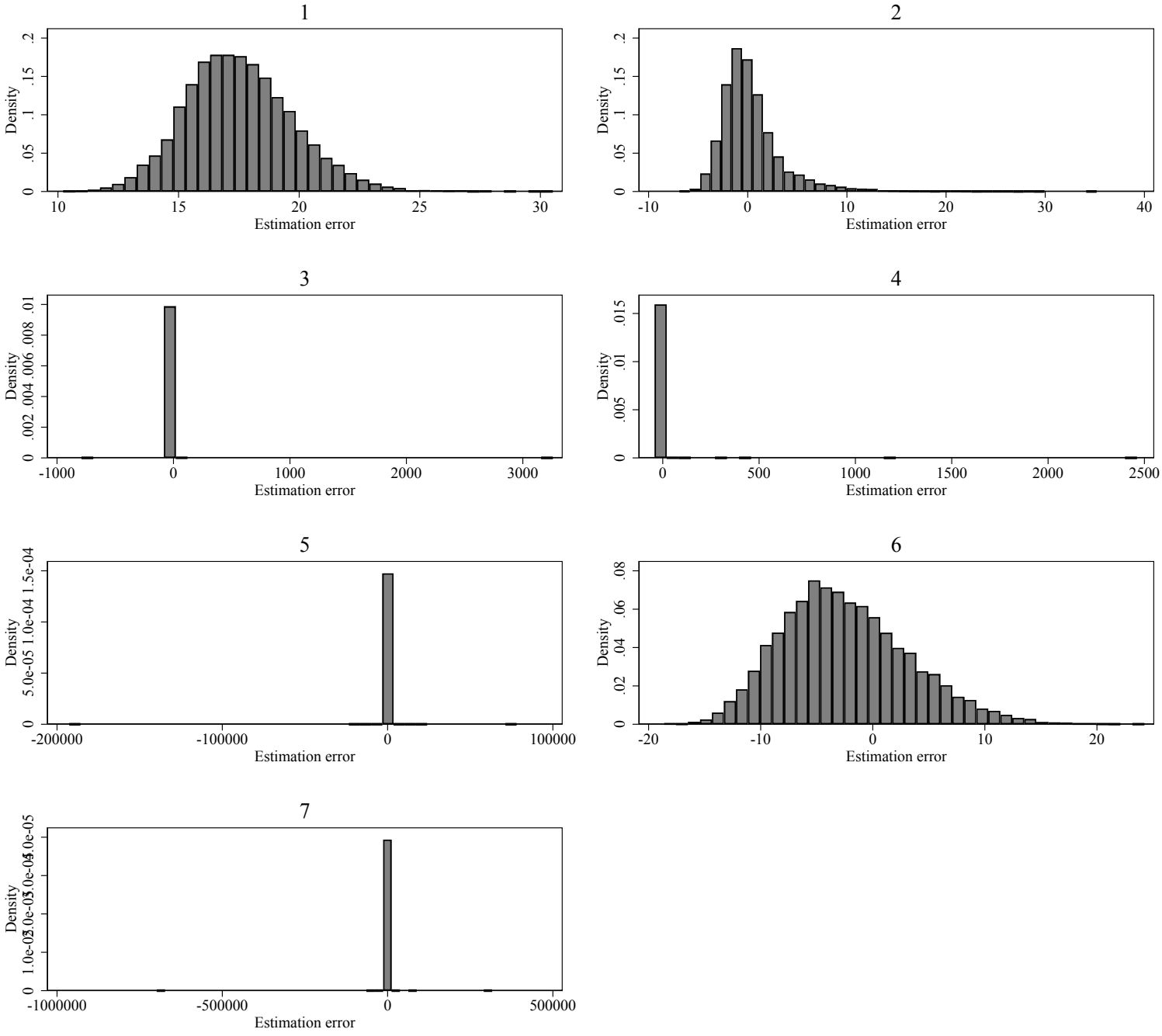
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



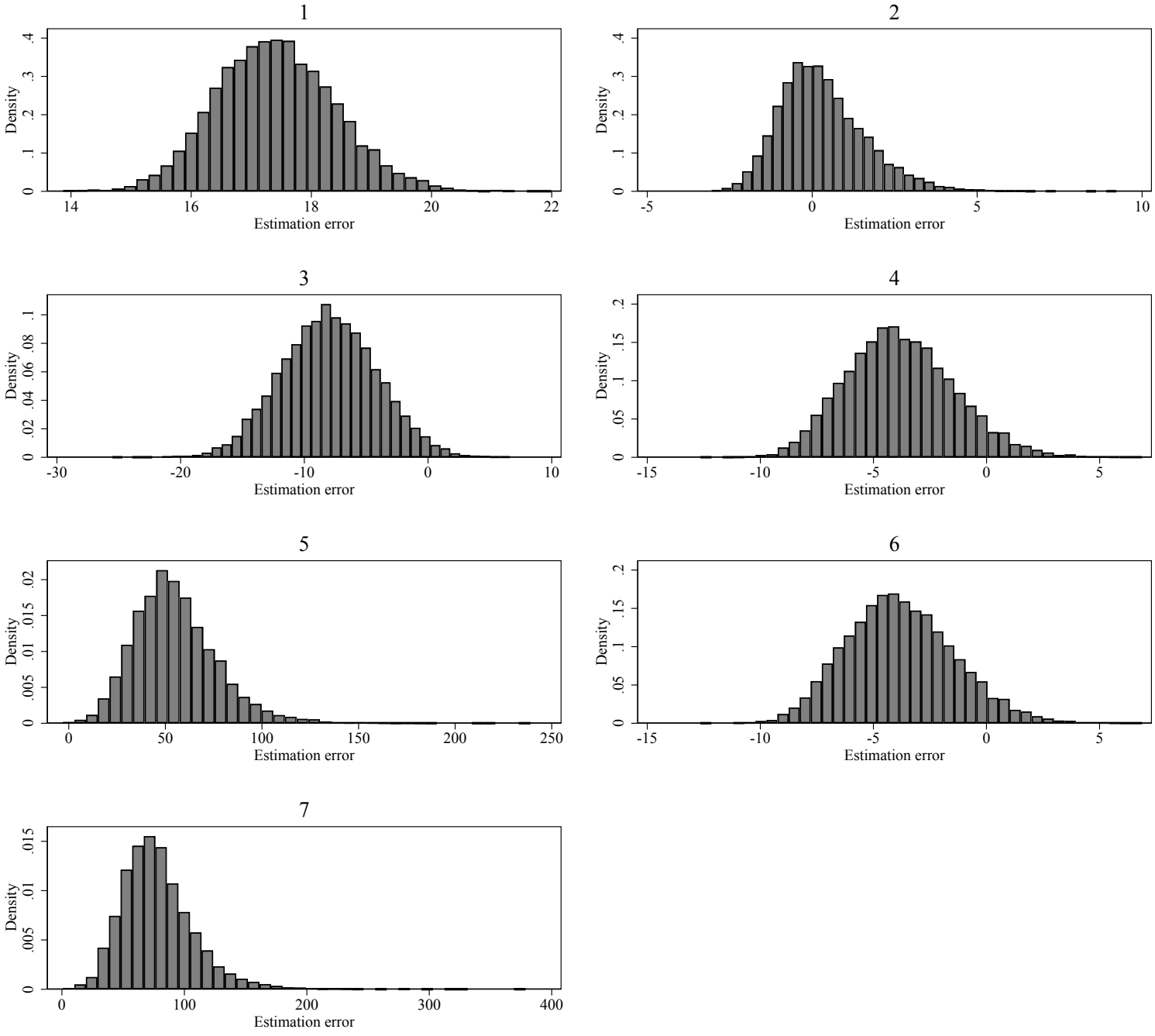
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



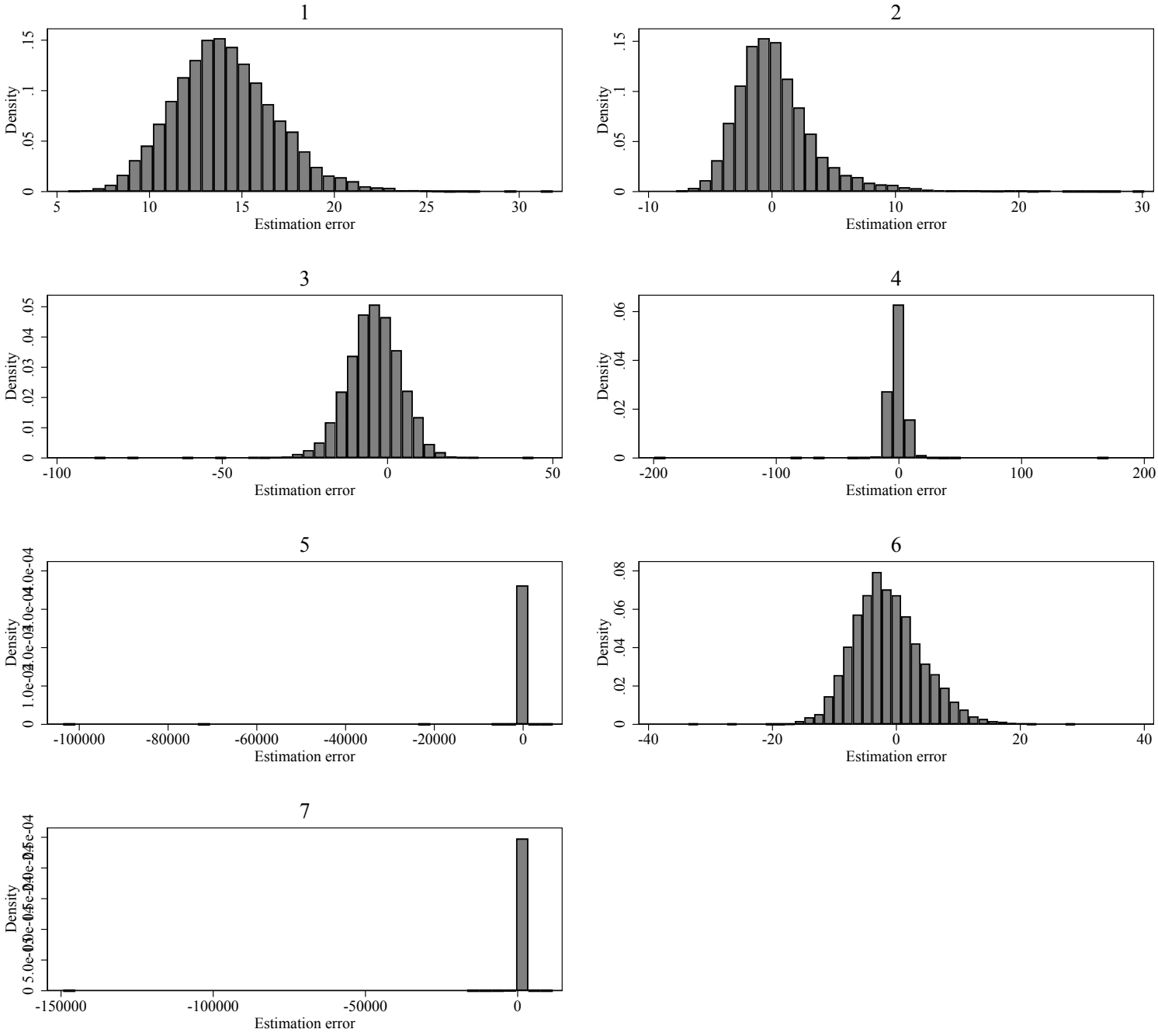
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ $(= \hat{\tau}_{a,1})$. "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

Figure B.39: Simulation Results for Design D, $\delta = 0.01$, $N = 5{,}000$



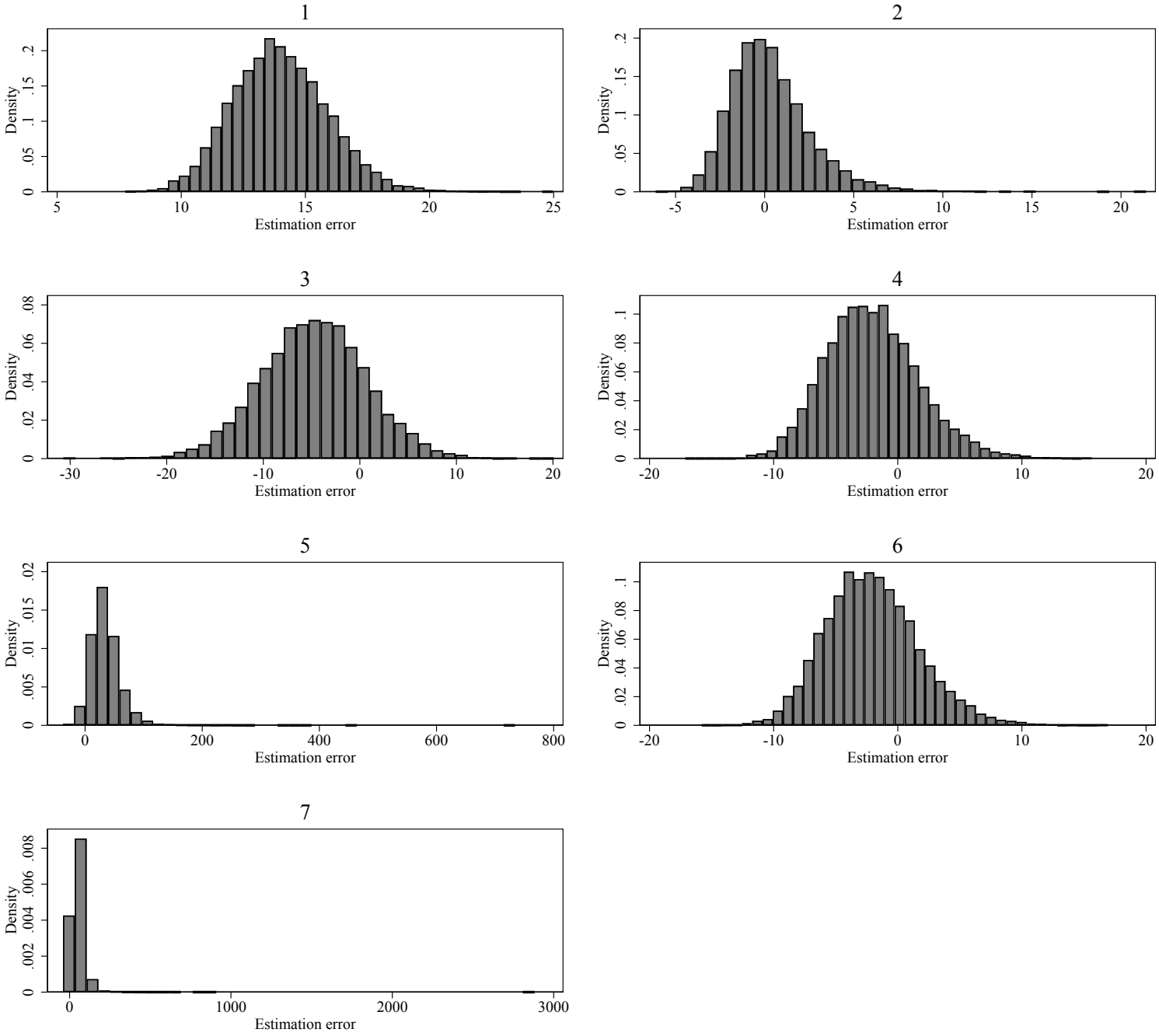
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

Figure B.40: Simulation Results for Design D, $\delta = 0.02$, $N = 500$



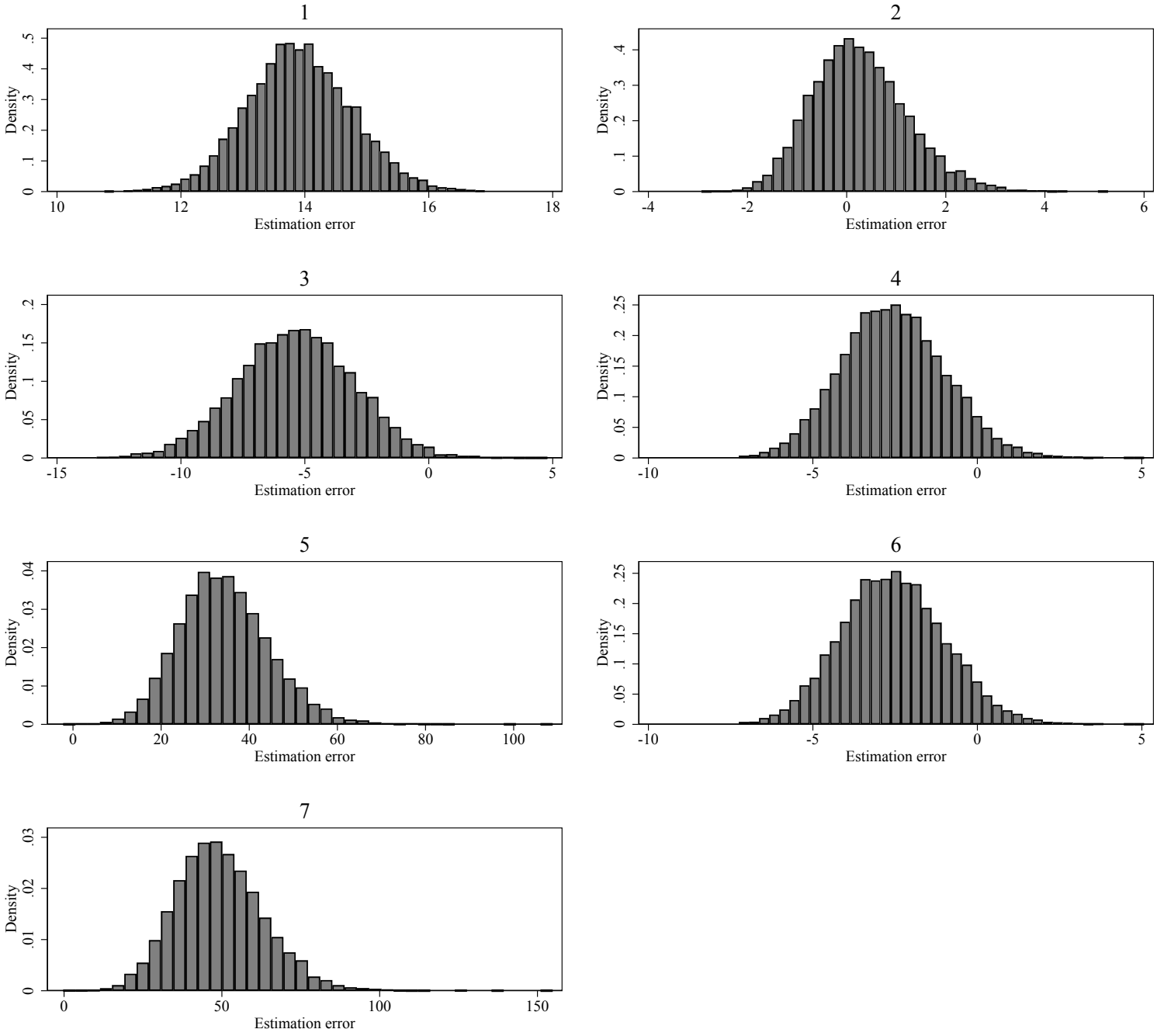
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



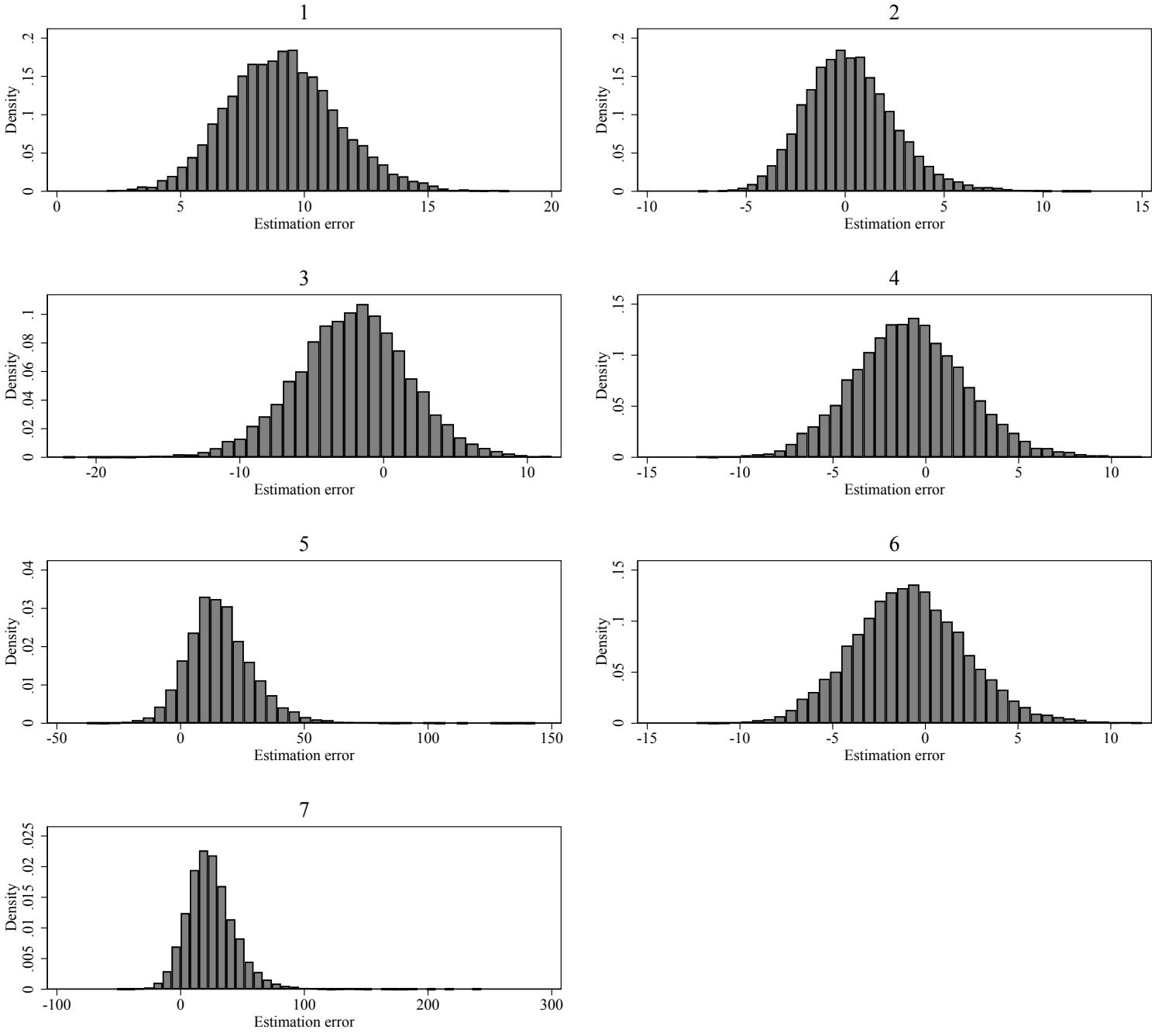
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



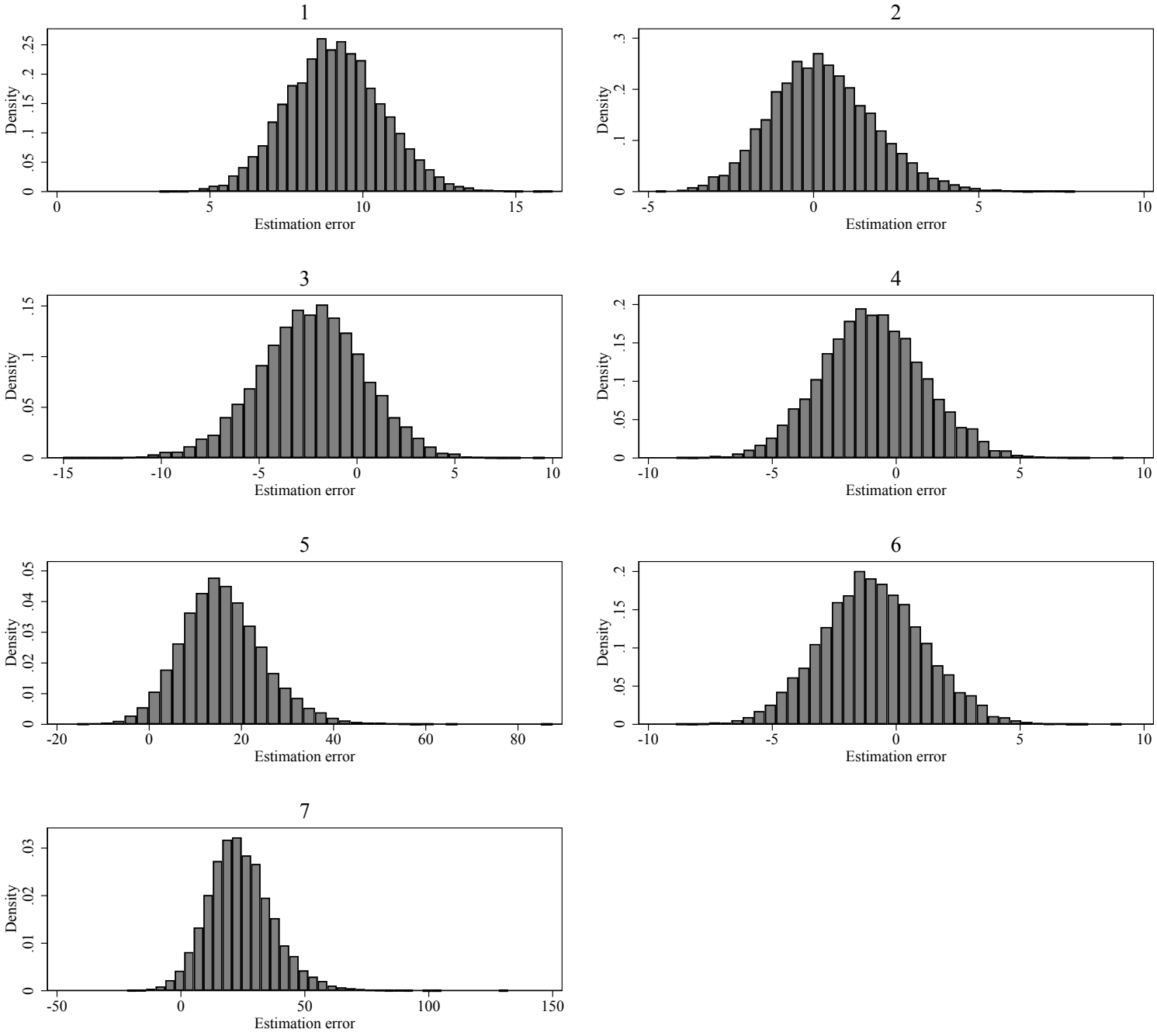
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



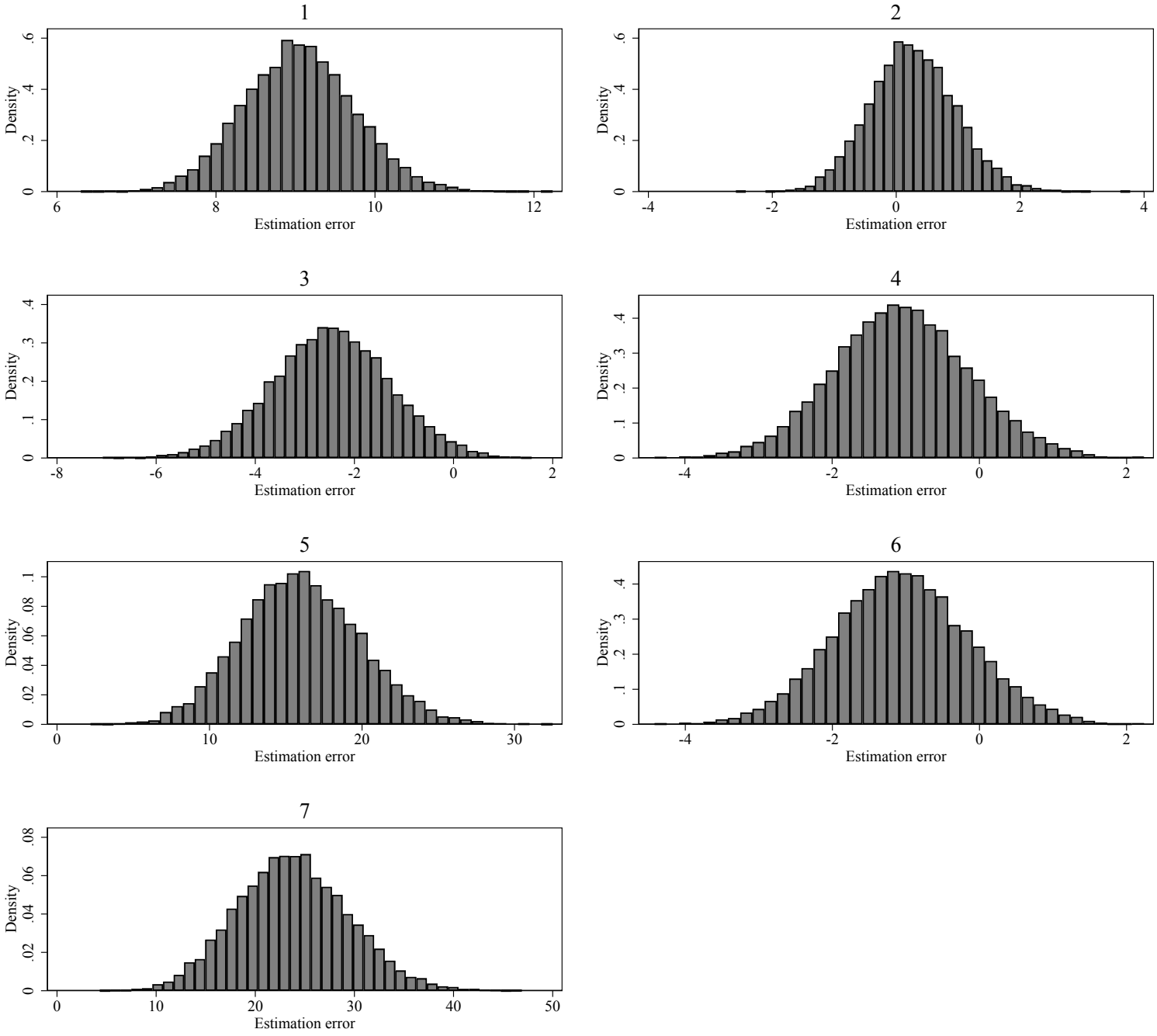
Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

Figure B.44: Simulation Results for Design D, $\delta = 0.05$, $N = 1{,}000$



Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.

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



Notes: The details of this simulation design are provided in Section 4. "1" corresponds to the linear IV estimator that controls for X. "2" corresponds to $\hat{\tau}_{cb}$, where the instrument propensity score is estimated using the approach of Imai and Ratkovic (2014), controlling for X. "3" corresponds to $\hat{\tau}_{t,norm}$. "4" corresponds to $\hat{\tau}_{a,10}$. "5" corresponds to $\hat{\tau}_a$. "6" corresponds to $\hat{\tau}_t$ (= $\hat{\tau}_{a,1}$). "7" corresponds to $\hat{\tau}_{a,0}$. The weighting estimators, other than $\hat{\tau}_{cb}$, use a logit to estimate the instrument propensity score, also controlling for X. Results are based on 10,000 replications.