Supplementary Materials for

"Estimating the LATE Without the Exclusion Restriction"

Zachary Markovich

A Proofs

A.1 Proof of Prop 1

Note, Assumption 4 implies that $\mathbb{E}\left(\tau_i|\gamma_i \neq 1, X_i = x\right) >= 0$ for all $x \in \mathcal{X}$. Therefore, from the law of total expectation, for all $x \in \mathcal{X}$,

$$\begin{aligned}
\mathbb{E}\left(\tau_i|X_i = x\right) &= P(\gamma_i = 1|X_i = x)\,\mathbb{E}\left(\tau_i|X_i = x, \gamma_i = 1\right) + \left[1 - P(\gamma_i = 1|X_i = x)\right]\mathbb{E}\left(\tau_i|X_i = x, \gamma_i \neq 1\right) \\
&= P(\gamma_i = 1|X_i = x)\,\mathbb{E}\left(\tau_i|X_i = x, \gamma_i = 1\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\int_{\mathcal{X}} f_{X_i|\gamma_i}(x|\gamma_i = 1)\left[\mathbb{E}\left(\tau_i|X_i = x, \gamma_i = 1\right) - \mathbb{E}\left(\tau_i|X_i = x\right)\right] dx \\
&\quad = \int_{\mathcal{X}} f_{X_i|\gamma_i}(x|\gamma_i = 1)\left[1 - P(\gamma_i = 1|X_i = x)\right]\mathbb{E}\left(\tau_i|X_i = x, \gamma_i = 1\right) dx \\
&\quad \geq 0.
\end{aligned}$$

A.2 Proof of Lemma 1

$$\begin{aligned}
\mathrm{LATE} - \mathrm{ITT} &= \mathbb{E}\left(\tau_i|\gamma_i = 1\right) - \mathbb{E}\left(\tau_i\right) \\
&= \int_{\mathcal{X}} f_{X_i|\gamma_i}(x|\gamma_i = 1)\,\mathbb{E}\left(\tau_i|X_i = x, \gamma_i = 1\right) dx - \int_{\mathcal{X}} f_{X_i}(x)\,\mathbb{E}\left(\tau_i|X_i = x\right) dx \\
&= \int_{\mathcal{X}} f_{X_i|\gamma_i}(x|\gamma_i = 1)\,\mathbb{E}\left(\tau_i|X_i = x, \gamma_i = 1\right) dx - \int_{\mathcal{X}} f_{X_i}(x)\,\mathbb{E}\left(\tau_i|X_i = x\right) dx \\
&\quad + \int_{\mathcal{X}} f_{X_i|\gamma_i}(x|\gamma_i = 1)\,\mathbb{E}\left(\tau_i|X_i = x\right) dx - \int_{\mathcal{X}} f_{X_i|\gamma_i}(x|\gamma_i = 1)\,\mathbb{E}\left(\tau_i|X_i = x\right) dx \\
&= \int_{\mathcal{X}} f_{X_i|\gamma_i}(x|\gamma_i = 1)\left[\mathbb{E}\left(\tau_i|X_i = x, \gamma_i = 1\right) - \mathbb{E}\left(\tau_i|X_i = x\right)\right] dx \\
&\quad + \int_{\mathcal{X}} \mathbb{E}\left(\tau_i|X_i = x\right)\left[f_{X_i|\gamma_i}(x|\gamma_i = 1) - f_{X_i}(x)\right] dx
\end{aligned}$$

A.3 Proof of Lemma 2

First, re-write the CPW estimator as the difference between two ratios of sample means:

$$\begin{aligned}
\widehat{\text{CPW}} &= \sum_{i=1}^{N}(Z_iY_i - (1-Z_i)Y_i)\left[\frac{Z_i\hat{P}_N\left(\gamma_i = 1|X_i\right)}{\sum_{j=1}^{N} Z_j\hat{P}_N\left(\gamma_j = 1|X_j\right)} - \frac{\sum_{i=1}^{N}(1-Z_i)\hat{P}_N\left(\gamma_i = 1|X_i\right)}{\sum_{j=1}^{N}(1-Z_j)\hat{P}_N\left(\gamma_j = 1|X_j\right)}\right] \\
&= \frac{\sum_{i=1}^{N} Y_iZ_i\hat{P}_N\left(\gamma_i = 1|X_i\right)}{\sum_{j=1}^{N} Z_j\hat{P}_N\left(\gamma_j = 1|X_j\right)} - \frac{\sum_{i=1}^{N} Y_i(1-Z_i)\hat{P}_N\left(\gamma_i = 1|X_i\right)}{\sum_{j=1}^{N}(1-Z_j)\hat{P}_N\left(\gamma_j = 1|X_j\right)} \\
&= \frac{(1/\sum_{i=1}^{N} Z_i)\sum_{i=1}^{N} Y_iZ_i\hat{P}_N\left(\gamma_i = 1|X_i\right)}{(1/\sum_{i=1}^{N} Z_i)\sum_{j=1}^{N} Z_j\hat{P}_N\left(\gamma_j = 1|X_j\right)} - \frac{(1/\sum_{i=1}^{N}(1-Z_i))\sum_{i=1}^{N} Y_i(1-Z_i)\hat{P}_N\left(\gamma_i = 1|X_i\right)}{(1/\sum_{i=1}^{N}(1-Z_i))\sum_{j=1}^{N}(1-Z_j)\hat{P}_N\left(\gamma_j = 1|X_j\right)} \\
&= \frac{\hat{\mathbb{E}}_N\left[Y_i\hat{P}_N(\gamma_i = 1|X_i)|Z_i = 1\right]}{\hat{\mathbb{E}}_N\left[\hat{P}_N(\gamma_i = 1|X_i)|Z_i = 1\right]} - \frac{\hat{\mathbb{E}}_N\left[Y_i\hat{P}_N(\gamma_i = 1|X_i)|Z_i = 0\right]}{\hat{\mathbb{E}}_N\left[\hat{P}_N(\gamma_i = 1|X_i)|Z_i = 0\right]}.
\end{aligned}$$

Note, in the above, I have used subscripting with N to denote quantities that depend on the size of the sample. So, $\hat{\mathbb{E}}_N(\cdot)$ represents the sample mean of the given quantity and $\hat{P}_N(\gamma_i = 1|X_i)$ is the estimated probability that $\gamma_i = 1$. Note that $\hat{\mathbb{E}}\left(\right)_N$ is a linear function and is therefore continuous. Since $\hat{P}_N(\gamma_i = 1|X_i)$ is consistent (Assumption 7) and $\hat{\mathbb{E}}\left(\right)_N$ converges to $\mathbb{E}(\cdot)$ by the law of large numbers, by the continuous mapping theorem, their composition will converge to the composition of their limits. Therefore:

$$\begin{aligned}
\hat{\mathbb{E}}_N\left[\hat{P}_N(\gamma_i = 1|X_i)|Z_i = z\right] &\xrightarrow[N\to\infty]{p} \mathbb{E}\left[P(\gamma_i = 1|X_i)|Z_i = z\right] \\
&= \mathbb{E}\left[P(\gamma_i = 1|X_i\right] \\
&= P\left(\gamma_i = 1\right).
\end{aligned}$$

for any $z \in \{0, 1\}$. Similarly,

$$\hat{\mathbb{E}}_N\left[Y_i\hat{P}_N(\gamma_i = 1|X_i)|Z_i = z\right] \xrightarrow[N\to\infty]{p} \mathbb{E}\left[Y_iP(\gamma_i = 1|X_i)|Z_i = z\right].$$

Again, for any $z \in \{0, 1\}$. Therefore, from Slutksy's theorem:

$$
\begin{aligned}
\frac{\hat{\mathbb{E}}_N\left[Y_i \hat{P}_N(\gamma_i = 1|X_i)|Z_i = z\right]}{\hat{\mathbb{E}}_N\left[\hat{P}_N(\gamma_i = 1|X_i)|Z_i = z\right]} &\xrightarrow[N\to\infty]{p} \frac{\mathbb{E}\left[Y_i P(\gamma_i = 1|X_i)|Z_i = z\right]}{P\left(\gamma_i = 1\right)} \\
&= \mathbb{E}\left[Y_i(z)\frac{P(\gamma_i = 1|X_i)}{P\left(\gamma_i = 1\right)}\right] \\
&= \int_{\mathcal{X}} \mathbb{E}\left[Y_i(z)|X_i = x\right] \frac{P(\gamma_i = 1|X_i = x)}{P\left(\gamma_i = 1\right)} f_{X_i}(X_i = x)dx \\
&= \int_{\mathcal{X}} \mathbb{E}\left[Y_i(z)|X_i = x\right] f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1)dx.
\end{aligned}
$$

Where the first equality follows from Assumption 1, the second from the law of iterated expectation, and the third from Bayes theorem.

Now substituting this identity into the formula for the CPW estimator:

$$
\begin{aligned}
\widehat{\text{CPW}} &\xrightarrow[N\to\infty]{p} \int_{\mathcal{X}} \mathbb{E}\left[Y_i(1)|X_i = x\right] f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1)dx - \int_{\mathcal{X}} \mathbb{E}\left[Y_i(0)|X_i = x\right] f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1)dx \\
&= \int_{\mathcal{X}} \mathbb{E}\left[Y_i(1) - Y_i(0)|X_i = x\right] f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1)dx \\
&= \int_{\mathcal{X}} \mathbb{E}\left[\tau_i|X_i = x\right] f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1)dx.
\end{aligned}
$$

Which completes the proof of the first identity in the lemma. For the second, note that, from the previous result,

$$
\begin{aligned}
\text{LATE} - \widehat{\text{CPW}} &\xrightarrow[N\to\infty]{p} \mathbb{E}\left(\tau_i|\gamma_i = 1\right) - \int_{\mathcal{X}} f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1)\,\mathbb{E}\left(\tau_i|X_i = x\right)dx \\
&= \int_{\mathcal{X}} f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1)\,\mathbb{E}\left(\tau_i|X_i = x, \gamma_i = 1\right)dx - \int_{\mathcal{X}} f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1)\,\mathbb{E}\left(\tau_i|X_i = x\right)dx \\
&= \int_{\mathcal{X}} f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1)\left[\mathbb{E}\left(\tau_i|X_i = x, \gamma_i = 1\right) - \mathbb{E}\left(\tau_i|X_i = x\right)\right]dx.
\end{aligned}
$$

A.4 Proof of Prop 3

First consider the case when $\forall x \in \mathcal{X}$, $\mathbb{E}(\tau_i|X_i = x, \gamma_i = 1) \geq \mathbb{E}\left(\tau_i|X_i = x\right) \geq 0$ (condition a. from the Assumption 1). Note that this implies that $\mathbb{E}\left(\tau_i|\gamma_i\right) \geq 0$, $\int_{\mathcal{X}} f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1)\,\mathbb{E}\left(\tau_i|X_i = x\right)dx \geq 0$, and

$\int_{\mathcal{X}} f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1) \left[\mathbb{E}\left(\tau_i|X_i = x, \gamma_i = 1\right) - \mathbb{E}\left(\tau_i|X_i = x\right)\right] dx \geq 0$. This establishes that case (a) implies Assumption 8, so the conclusion of Proposition 2: $\lim_{N\to\infty} P(|\text{LATE}| \geq |\widehat{\text{CPW}}|) = 1$, will hold.

Now consider the case when, $\forall x \in \mathcal{X}$, $\mathbb{E}(\tau_i|X_i = x, \gamma_i = 1) \leq \mathbb{E}\left(\tau_i|X_i = x\right) \leq 0$ (condition b from Assumption 1). Note that this implies that $\mathbb{E}\left(\tau_i|\gamma_i\right) \leq 0$, $\int_{\mathcal{X}} f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1)\, \mathbb{E}\left(\tau_i|X_i = x\right) dx \leq 0$, and $\int_{\mathcal{X}} f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1) \left[\mathbb{E}\left(\tau_i|X_i = x, \gamma_i = 1\right) - \mathbb{E}\left(\tau_i|X_i = x\right)\right] dx \leq 0$. So

$$
\begin{aligned}
|\text{LATE}| - |\widehat{\text{CPW}}| \xrightarrow[N\to\infty]{p} & \left|\, \mathbb{E}\left(\tau_i|\gamma_i = 1\right) \right| - \left| \int_{\mathcal{X}} f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1)\, \mathbb{E}\left(\tau_i|X_i = x\right) dx \right| \\
&= -\, \mathbb{E}\left(\tau_i|\gamma_i = 1\right) + \int_{\mathcal{X}} f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1)\, \mathbb{E}\left(\tau_i|X_i = x\right) dx \\
&= -\left[\mathbb{E}\left(\tau_i|\gamma_i = 1\right) - \int_{\mathcal{X}} f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1)\, \mathbb{E}\left(\tau_i|X_i = x\right) dx\right] \\
&= -\left(\int_{\mathcal{X}} f_{X_i|\gamma_i}(X_i = x|\gamma_i = 1) \left[\mathbb{E}\left(\tau_i|X_i = x, \gamma_i = 1\right) - \mathbb{E}\left(\tau_i|X_i = x\right)\right] dx\right) \\
&\geq 0.
\end{aligned}
$$

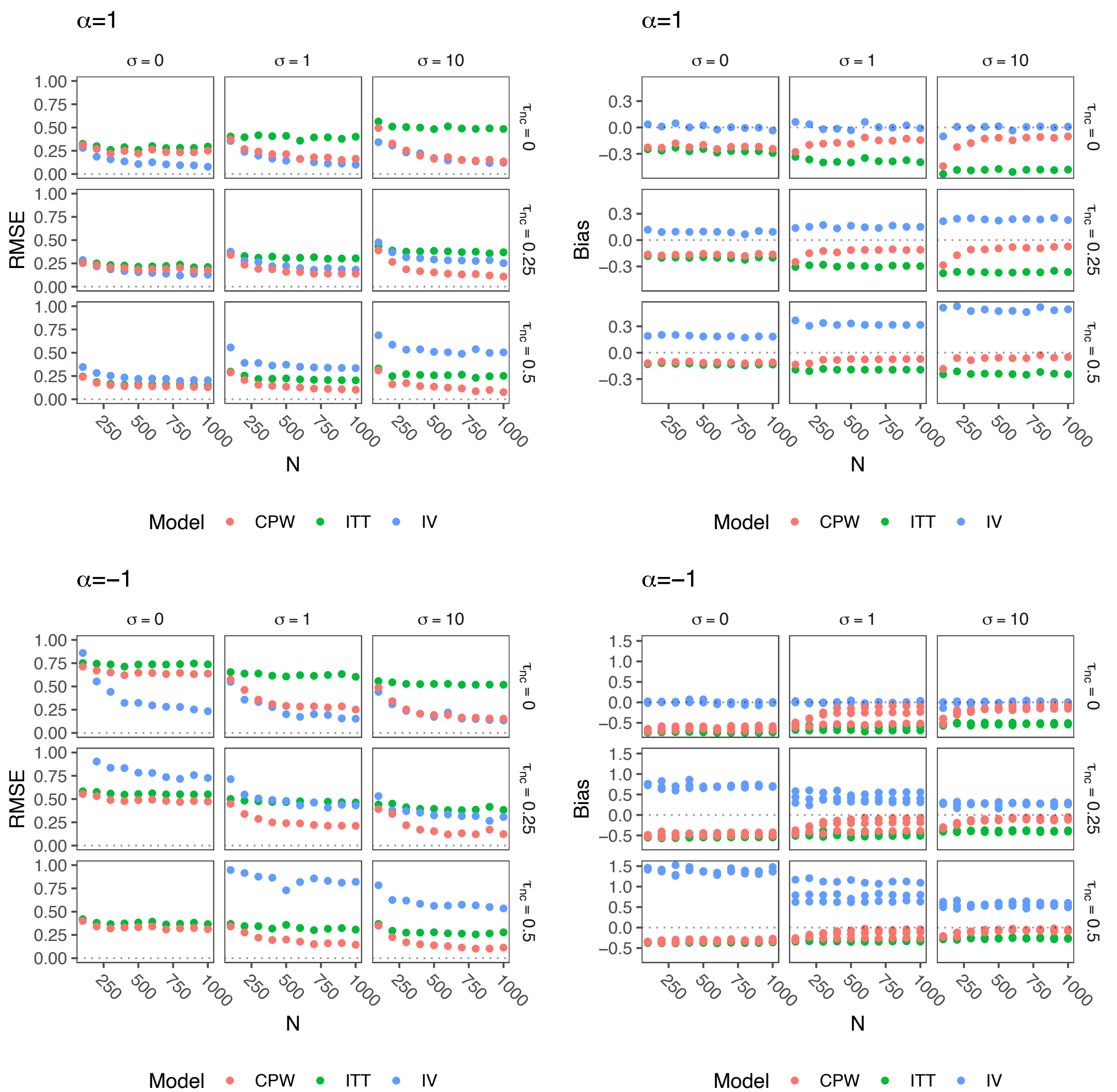
B Additional Simulation Results

B.1 Varying Compliance Rate

One limitation of the simulations in the main paper is that they do not allow the percent of the sample that are compliers to vary. To address this shortcoming, the simulation results presented in this subsection add an offset term to the equation that defines compliance probability so that, $\mu_i = \text{Logit}^{-1}(X_i'\beta + \alpha)$. Recall that μ_i is the probability that unit i is a complier, so positive values of α increase the fraction of the sample that is a complier, while negative values decrease it. The plots below provide simulation results when α is set equal to -1 and 1. Note that the effect of alpha depends on σ. When σ is large, X_i is a strong predictor of γ_i so changing alpha will not impact the ultimate distribution of γ_i as much. When $\alpha = 1$, the fraction of compliers is .74, .63, and .498 for σ set to 0, 1, and 10, respectively. The equivalent values for $\alpha = -1$ are .28, .361, and .478 instead. Although other details of the simulations are the same as those presented in the main text.

Figure B.2 presents the results of this analysis. The patterns generally match those seen in the main text. The CPW estimator does best when X is strongly predictive of γ_i while the performance of the IV estimator remains dependent on the magnitude of the exclusion restriction violation.

Figure B.1: Simulation Results With Varying Compliance Rates

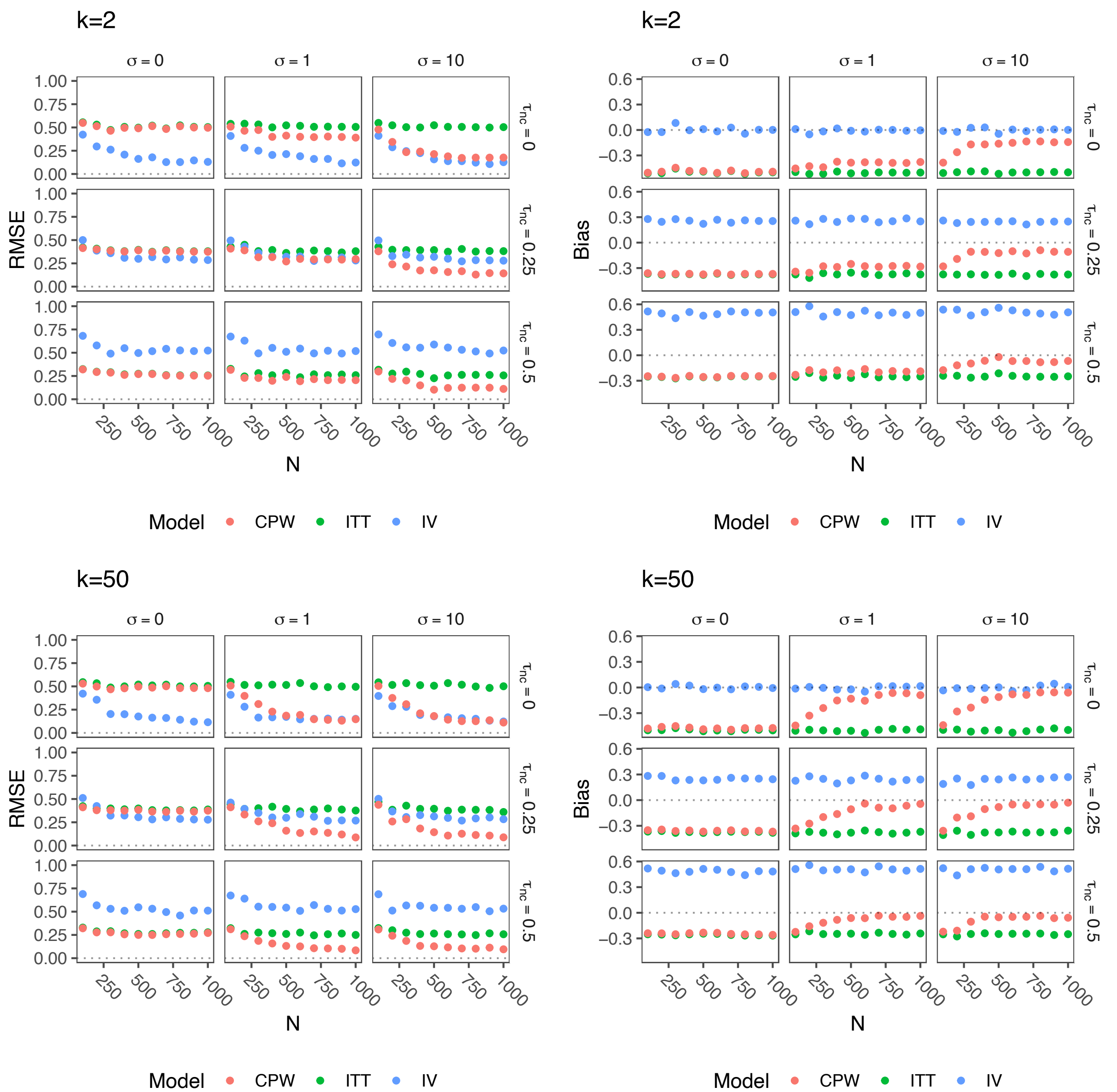


Note: Figure presents results from simulations comparing efficacy of the ITT, IV, and CPW estimators. The vertical facets identify the value of σ while the horizontal facets identify the value of τ_{NC}. Each point represents the results from 100 simulations with those parameters. Top panels show the results when $\alpha = 1$, while bottom panels present the case when $\alpha = -1$

B.2 Varying K

This subsection presents additional simulation results where the dimensionality of the X_i, K, is allowed to vary from 10. In particular, I consider the case when $K = 2$ and $K = 50$. The results seen here suggest that the rate of convergence of the estimator decreases as K increases, but the main results about the asymptotic conservatism of the CPW estimator still hold.

Figure B.2: Simulation Results With Varying K

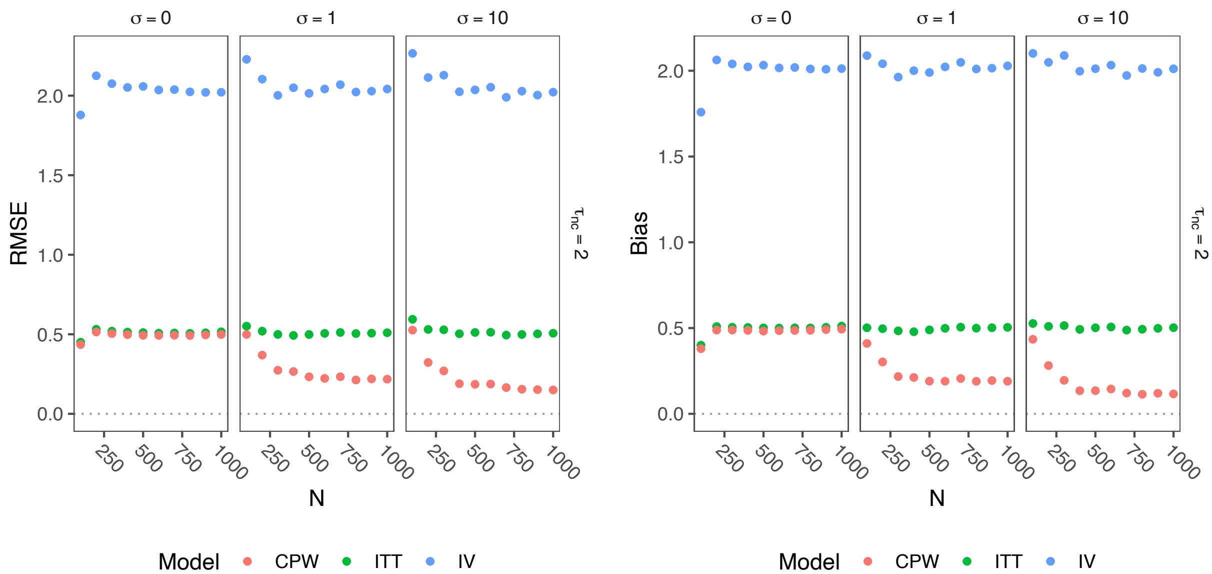


Note: Figure presents results from simulations comparing efficacy of the ITT, IV, and CPW estimators. The vertical facets identify the value of σ while the horizontal facets identify the value of τ_{NC}. Each point represents the results from 100 simulations with those parameters. Top panels show the results when $K = 2$. Bottom panel shows reuslts when $K = 50$

B.3 Assumption 8 Fails

Simulations can also be used to consider the consequences of the failure of Assumption 8. In particular, I accomplish this by setting τ_{NC} equal to two in the simulation set up from the main text, which implies that the effect of assignment to treatment on the outcome is larger for non-compliers than it is for compliers. Figure B.3 presents the results of this analysis, which show that while the bias of the CPW estimator is now positive, it actually still delivers a lower RMSE than the other estimators considered

Figure B.3: Simulation Results When the Effect of Assignment to Treatment Is Larger For Non-Compliers



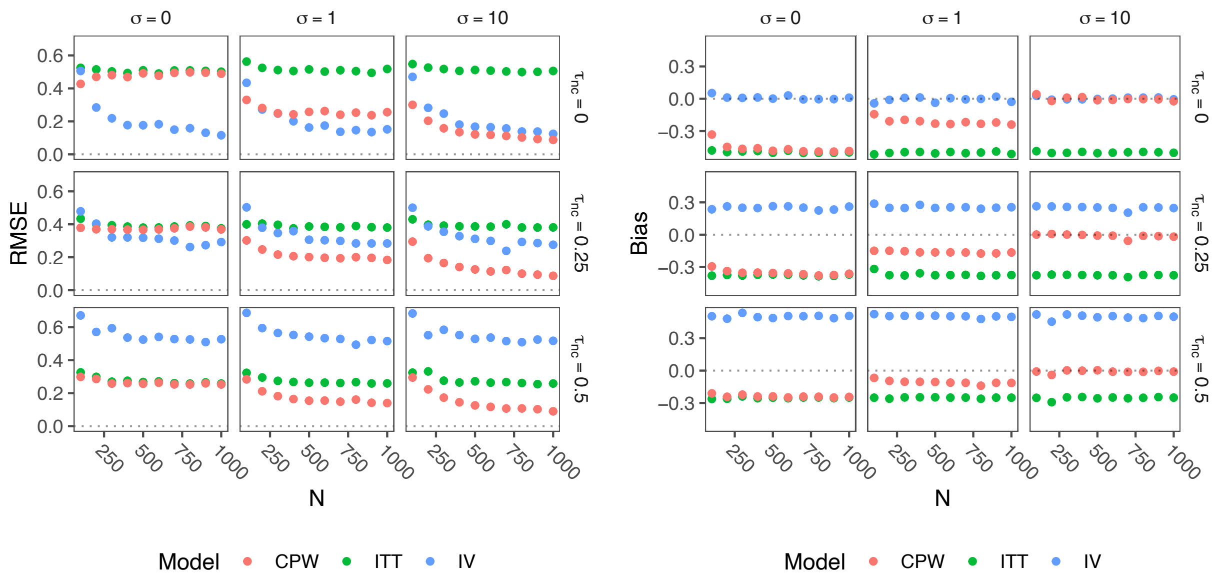
Note: Figure presents results from simulations comparing efficacy of the ITT, IV, and CPW estimators. The vertical facets identify the value of σ while the horizontal facets identify the value of τ_{NC}. Each point represents the results from 100 simulations with those parameters.

B.4 Parametric Approaches to Estimating the Probability of Compliance

The simulation results presented in the main paper only consider one model for estimating the probability of compliance. In this section, I present results which instead use the estimator proposed by Aronow and Carnegie (2013). This estimator differs from that used in the main paper principally in that it is based on a parametric linear model of compliance probability, while the estimator used in the main paper is non-parametric. This estimator can also easily estimate the probability of compliance even in the face of two-sided non-compliance, while the approach used in the main paper assumes one-sided non-compliance. Figure B.4 presents the results of this analysis under the same simulation set up used in the main paper when the model used to estimate the probability that each unit is a complier is correctly specified. Overall, it shows that the results generally match those seen in the same paper, although the estimator is somewhat more efficient, likely due to its ability to capitalize on the correctly specified linear model.

The results presented in Figure B.5 consider the performance of this parametric estimator when the model is incorrectly specified. Specifically, I tweak the data generating process so that the probability that each unit is a complier is generated as a linear function of all pairwise interactions between each element of X_i (i.e. $\mu_i = \text{Logit}^{-1}(\sum_{j=1}^{K}\sum_{l=1}^{K} X_{ij}X_{il}\beta_{il})$ where β is a matrix of independent standard normal draws). Unsuprisingly, the CPW estimator's performance is worse under these circumstances; however, it

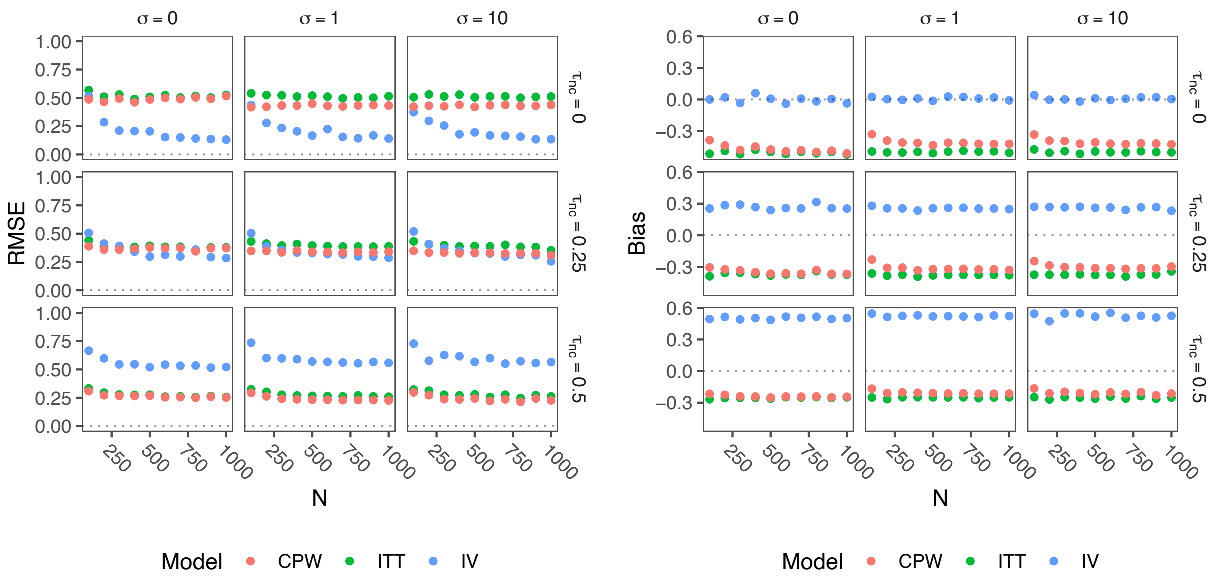
Figure B.4: Simulation Results Using Correctly Specified Linear Model For Compliance Probability



Note: Figure presents results from simulations comparing efficacy of the ITT, IV, and CPW estimators. The vertical facets identify the value of σ while the horizontal facets identify the value of τ_{NC}. Each point represents the results from 100 simulations with those parameters.

still performs well and the bias is still towards 0.

Figure B.5: Simulation Results Using Incorrectly Misspecified Linear Model For Compliance Probability



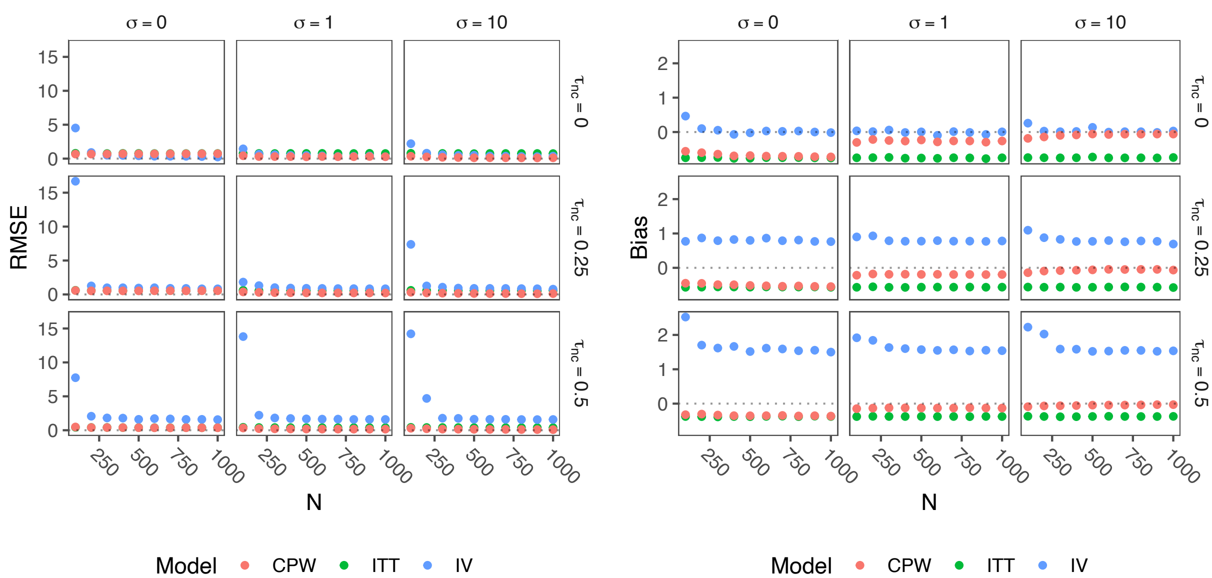
Note: Figure presents results from simulations comparing efficacy of the ITT, IV, and CPW estimators. The vertical facets identify the value of σ while the horizontal facets identify the value of τ_{NC}. Each point represents the results from 100 simulations with those parameters.

B.5 Two Sided Non-Compliance

One limitation of the non-parametric approach to estimating the probability that each unit is a complier used in the main paper is that it assumes non-compliance is one sided. This Assumption can be weakened by instead using the parametric estimator proposed by Aronow and Carnegie (2013). To explore this possibility, I tweaked the simulation set up so that the probability that a unit was a complier was generated in the same way as before ($\mu_i = \text{Logit}^{-1}(X_i'\beta)$); however, I assigned units to be defiers with probability $(1 - \mu_i)/2$, and always and never takers each with probability $(1 - \mu_i)/4$. I assumed that the effect of assignment to treatment for always and never takers was τ_{NC}, which I allowed to vary, but was always -1 (ie, the negative LATE) for defiers. Note that this setup still satisfies all of the Assumptions required for the conservatism of the CPW estimator.

Figure B.6 presents the results of this analysis and shows the CPW estimator still performs as expected with the general pattern of results matching that seen in the main paper.

Figure B.6: Simulation Results Under Two Sided Non-Compliance



Note: Figure presents results from simulations comparing efficacy of the ITT, IV, and CPW estimators. The vertical facets identify the value of σ while the horizontal facets identify the value of τ_{NC}. Each point represents the results from 100 simulations with those parameters.