Supplemental Information: Matching Methods for Causal Inference with Time-Series Cross-Section Data

Intended for online publication only

Kosuke Imai In Song Kim Erik Wang

Supplemental Information: Table of Contents

Appendix A Specifying the Future Treatment Sequence

Formally, suppose that after a policy change, for some observations, the treatment will be in place at least for F time periods. We may be interested in estimating the ATT of stable policy change relative to no policy change among these treated observations. In this case, the ATT can be defined as,

$$\begin{aligned}&\mathbb{E}\left[Y_{i,t+F}\left(\{X_{i,t+\ell}\}_{\ell=1}^{F}=\mathbf{1}_F, X_{it}=1, X_{i,t-1}=0, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) - \right.\\ &\qquad \left. Y_{i,t+F}\left(\{X_{i,t+\ell}\}_{\ell=1}^{F}=\mathbf{0}_F, X_{it}=0, X_{i,t-1}=0, \{X_{i,t-\ell}\}_{\ell=2}^{L}\right) \mid \{X_{i,t+\ell}\}_{\ell=1}^{F}=\mathbf{1}_F, X_{it}=1, X_{i,t-1}=0\right]\end{aligned} \tag{1}$$

where $\mathbf{1}_F$ and $\mathbf{0}_F$ are F dimensional vectors of ones and zeros, respectively. In our two applications, this alternative quantity represents the causal effect of democratization on economic growth (without reverting to an authoritarian regime) and that of continuing war on inheritance taxation (without ending it), respectively.

The difference between equations (8) and (1) is that the latter specifies the future treatment sequence. The treated (matched control) observations are those who remain under the treatment (control) condition throughout F time periods after the administration of the treatment whereas the matched control units receive no treatment at least for F time periods after the treatment is given. The matched set changes to,

$$\mathcal{M}_{it} \;=\; \{i' : i' \neq i, X_{i't}=X_{i't+1}=\ldots=X_{i't+F}=0, X_{i't'}=X_{it'} \text{ for all } t'=t-1,\ldots,t-L\} \tag{2}$$

To estimate this ATT, we apply the idea of marginal structural models (MSMs) in order to make covariate adjustments while avoiding post-treatment bias (Robins *et al.*, 2000). Note that the identification assumption is unchanged. We first constrain the matched set for each treated observation (i,t) such that the matched control units do not receive the treatment at least after time $t+F$. We then estimate the propensity score by modeling the treatment assignment, for example, using the logistic regression,

$$e_{it}(\{\mathbf{U}_{i,t-\ell}\}_{\ell=1}^{L}) \;=\; \Pr(X_{it}=1 \mid \mathbf{U}_{i,t-1},\ldots,\mathbf{U}_{i,t-L}) \;=\; \frac{1}{1+\exp(-\sum_{\ell=1}^{L}\boldsymbol{\beta}_\ell^\top \mathbf{U}_{i,t-\ell})}. \tag{3}$$

Unlike the above setting, the model must be fit to all observations including those who are not in the matched sets in order to model the entire treatment sequence. Using the result from MSMs, the weights are then computed as,

$$w_{it}^{i'} \;=\; \prod_{f=0}^{F} \frac{e_{i,t+f}(\{\mathbf{U}_{i,t+f-\ell}\}_{\ell=1}^{L})}{1-e_{i,t+f}(\{\mathbf{U}_{i,t+f-\ell}\}_{\ell=1}^{L})} \tag{4}$$

for $i' \in \mathcal{M}_{it}$ and $w_{it}^{i'}=0$ if $i' \notin \mathcal{M}_{it}$. Finally, we apply the DiD estimator in equation (18) to obtain an estimate of the long term ATT under the specified treatment sequence as defined in equation (1).

Appendix B Proof of Theorem 1

Let $A_{it}=2X_{it}-1$. We consider the following a general definition of the weights,

$$W_{it} \;=\; \sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'}\cdot v_{it}^{i't'} \quad \text{and} \quad v_{it}^{i't'} \;=\; \begin{cases} A_{it} & \text{if } (i,t)=(i',t'+F) \\ 1 & \text{if } (i,t)=(i',t'-1) \\ -A_{it}\cdot w_{i't'}^{i} & \text{if } i\in\mathcal{M}_{i't'}, t=t'+F \\ -w_{i't'}^{i} & \text{if. } i\in\mathcal{M}_{i't'}, t=t'-1 \\ 0 & \text{otherwise.}\end{cases}$$

Note that the quantity of interest given in equation (1) implies that $A_{it} = 1$ if $(i, t) = (i', t' + F)$, and $A_{it} = -1$ if $(i, t) \in \mathcal{M}_{i't'}, t = t' + F$ as the treatment status does not change for at least F time periods once treatment is administered at time t. This gives the weights in equation (22).

We begin this proof by establishing the following algebraic equality. Specifically, we prove that for any unit-specific constant α_i^*, the following equality holds,

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=1}^{T} W_{it} A_{it} \alpha_i^* \\
= \;& \sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^{i't'} A_{it} \alpha_i^* \right) \\
= \;& \sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'} \left(1 - 1 - \sum_{i \in \mathcal{M}_{i't'}} \sum_{t=t'+F} A_{it}^2 \cdot w_{i't'}^i - \sum_{i \in \mathcal{M}_{i't'}} \sum_{t=t'-1} A_{it} \cdot w_{i't'}^i \right) \alpha_i^* \\
= \;& \sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'} \left(1 - 1 - \sum_{i \in \mathcal{M}_{i't'}} \sum_{t=t'+F} w_{i't'}^i + \sum_{i \in \mathcal{M}_{i't'}} \sum_{t=t'-1} w_{i't'}^i \right) \alpha_i^* \\
= \;& \sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'} \left(1 - 1 - 1 + 1\right) \alpha_i^* \; = \; 0
\end{aligned} \tag{5}$$

where the second equality follows from the fact that $A_{it} = 1$ if $(i, t) = (i', t' + F)$, $A_{it} = -1$ if $(i, t) = (i', t' - 1)$, and $A_{it} = -1$ if $(i, t) \in \mathcal{M}_{i't'}, t = t' - 1$ as given by equation (1). The last equality if from $\sum_{i \in \mathcal{M}_{i't'}} w_{i't'}^i = 1$.

Following the same logic, it is straightforward to show that $\sum_{i=1}^{N}\sum_{t=1}^{T} W_{it} A_{it} \gamma_t^* = 0$ for any time-specific constant γ_t^* and $\sum_{i=1}^{N}\sum_{t=1}^{T} W_{it} A_{it} K^* = 0$ for any constant K^*. This implies that

$$A_{it} - \overline{A}_i^* - \overline{A}_t^* + \overline{A}^* = A_{it} \tag{6}$$

where $\overline{A}_i^* = \sum_{t=1}^{T} W_{it} A_{it} / \sum_{t=1}^{T} W_{it}$, $\overline{A}_t^* = \sum_{i=1}^{N} W_{it} A_{it} / \sum_{i=1}^{N} W_{it}$, $\overline{A}^* = \sum_{i=1}^{N}\sum_{t=1}^{T} W_{it} A_{it} / \sum_{i=1}^{N}\sum_{t=1}^{T} W_{it}$.

Second, we show the following algebraic equality,

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=1}^{T} W_{it} \\
= \;& \sum_{i=1}^{N}\sum_{t=1}^{T} \left(\sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'} \cdot v_{it}^{i't'} \right) \\
= \;& \sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^{i't'} \right) \\
= \;& \sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'} \left(1 + 1 + \sum_{i \in \mathcal{M}_{i't'}} \sum_{t=t'+F} w_{i't'}^i - \sum_{i \in \mathcal{M}_{i't'}} \sum_{t=t'-1} w_{i't'}^i \right) \\
= \;& 2 \sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'} = 2 \sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}.
\end{aligned} \tag{7}$$

Finally, we can derive the desired result,

$$\begin{aligned}
\hat{\beta}_{\mathsf{DiD}} &= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} W_{it}(A_{it} - \overline{A}_i^* - \overline{A}_t^* + \overline{A}^*)(Y_{it} - \overline{Y}_i^* - \overline{Y}_t^* + \overline{Y}^*)}{\sum_{i=1}^{N}\sum_{t=1}^{T} W_{it}(A_{it} - \overline{T}_i^* - \overline{T}_t^* + \overline{T}^*)^2} \\
&= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} W_{it}A_{it}(Y_{it} - \overline{Y}_i^* - \overline{Y}_t^* + \overline{Y}^*)}{\sum_{i=1}^{N}\sum_{t=1}^{T} W_{it}} \\
&= \frac{1}{2\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}} \sum_{i=1}^{N}\sum_{t=1}^{T} W_{it}A_{it}(Y_{it} - \overline{Y}_i^* - \overline{Y}_t^* + \overline{Y}^*) \\
&= \frac{1}{2\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}} \sum_{i=1}^{N}\sum_{t=1}^{T} W_{it}A_{it}Y_{it} \\
&= \frac{1}{2\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'} \cdot v_{it}^{i't'} A_{it}Y_{it} \\
&= \frac{1}{2\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}} \sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'} \sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^{i't'} A_{it}Y_{it} \\
&= \frac{1}{2\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}} \sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'} \left(Y_{i',t'+F} - Y_{i',t'-1} - \sum_{i \in \mathcal{M}_{i't'}} \sum_{t=t'+F} w_{i't'}^{i} Y_{it} + \sum_{i \in \mathcal{M}_{i't'}} \sum_{t=t'-1} w_{i't'}^{i} Y_{it} \right) \\
&= \frac{1}{2\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}} \sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'} \left(Y_{i',t'+F} - Y_{i',t'-1} - \sum_{i \in \mathcal{M}_{i't'}} w_{i't'}^{i} Y_{i,t'+F} + \sum_{i \in \mathcal{M}_{i't'}} w_{i't'}^{i} Y_{i,t'-1} \right) \\
&= \frac{1}{2\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}} \sum_{i=1}^{N}\sum_{t=L+1}^{T-F} D_{it} \left\{ (Y_{i,t+F} - Y_{i,t-1}) - \sum_{i' \in \mathcal{M}_{it}} w_{it}^{i'} \left(Y_{i',t+F} - Y_{i',t-1} \right) \right\} \\
&= \hat{\delta}(F, L)/2
\end{aligned}$$

where the second equality follows from equation (6), the third equality follows from equation (7), and the fourth equality is implied by equation (5). The second from the last equality follows from the fact that $D_{it} = 0$ for $t < L + 1$ and $t > T - F$ for any unit i, because there will be no matched for such units by construction. This concludes the proof because $2\hat{\beta}_{\mathsf{DiD}} = \hat{\delta}(F, L)$ (see Theorem 1). Note that the multiplication by 2 is required due to the change of the variable of the original treatment variable, i.e., $A_{it} = 2X_{it} - 1$. □

Appendix C A Simulation Study

In this appendix, we describe our simulation study. We emphasize that our simulation setting is favorable to OLS as the data are generated according to a linear model. We find that the proposed methodology is more robust to model misspecification than OLS while OLS is generally more efficient.

C.1 The Setup

To make our simulation studies realistic, we use the original data from Acemoglu *et al.* (2019). For simplicity, we begin by creating a balanced TSCS data set with $N = 162$ units and $T = 51$ time periods although our method can handle missing and/or unbalanced data. Since the original data set is unbalanced, we impute missing values for continuous (binary) variables based on linear (logistic) regression models.[1] We emphasize that the proposed methodology does not require the data to be balanced. Next, we generate the binary treatment variable X_{it} and the outcome variable Y_{it}, with true data generating process given by

$$\begin{aligned} X_{it} &\sim \text{Benoulli}\left(\text{logit}^{-1}\left\{\tilde{\alpha}_i + \tilde{\gamma}_t + \sum_{\ell=1}^{L} \tilde{\beta}_\ell^\top X_{i,t-\ell} + \sum_{\ell=0}^{L}\left(\tilde{\zeta}_\ell^\top \mathbf{Z}_{i,t-\ell} + \tilde{\phi}_\ell^\top [\mathbf{Z}_{i,t-\ell}^{(1)} : \mathbf{Z}_{i,t-\ell}^{(3)}]\right)\right\}\right) && (8) \\ Y_{it} &= \alpha_i + \gamma_t + \sum_{\ell=0}^{L} \beta_\ell^\top X_{i,t-\ell} + \sum_{\ell=0}^{L}\left(\zeta_\ell^\top \mathbf{Z}_{i,t-\ell} + \phi_\ell^\top [\mathbf{Z}_{i,t-\ell}^{(1)} : \mathbf{Z}_{i,t-\ell}^{(3)}]\right) + \epsilon_{it}, && (9) \end{aligned}$$

where $\epsilon \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$, α_i and γ_t are unit and time fixed effects, and $\mathbf{Z}_{i,t-\ell} = (\mathbf{Z}_{i,t-\ell}^{\top(1)}, \mathbf{Z}_{i,t-\ell}^{\top(2)}, \mathbf{Z}_{i,t-\ell}^{\top(3)})^\top$ is the lagged covariates consisting of continuous variables $\mathbf{Z}_{i,t-\ell}^{(1)}$, binary variables $\mathbf{Z}_{i,t-\ell}^{(2)}$, and a set of other continuous variables $\mathbf{Z}_{i,t-\ell}^{(3)}$ that have interactive effects with $\mathbf{Z}_{i,t-\ell}^{(1)}$ with the interaction being represented by a colon.[2] We set the values of all the parameters to the actual estimates obtained from fitting the above treatment and outcome models to the imputed data set, including the true contemporaneous treatment effect β_0 to -7.5. Finally, we set σ to the sample variance of the outcome variable.

We use $L = 3$ such that the true dynamic data generating process includes the lagged treatment, covariates, and the interaction between the covariates across three time periods. For each of the 1,000 independent Monte Carlo replications, we generate the treatment and outcome variables according to the models above while keeping the covariates $\mathbf{Z}_{it}$ fixed. To evaluate the robustness of the proposed methodology to model misspecification, we consider four scenarios: (1) severe misspecification with only one period of lagged variables (i.e., $L = 1$), (2) moderate misspecification with only two periods of lagged variables (i.e., $L = 2$), (3) correct specification (i.e., $L = 3$), and (4) over-specification by adjusting for unnecessary lags (i.e., $L = 4$).[3]

We examine the performance of OLS against the proposed methods. To ensure that OLS estimates are also conditioned on the treatment history, we take two distinct approaches. First, we control for the lags of the treatment variable according to the degree of mis(over)-specification: i.e., L. The results based on this approach are denoted as "OLS." Second, we constructed a dummy variable for each possible combination of treatment history. The results from this specification, including the saturated treatment dummies, are denoted as "OLS with dummies." We use the following three refinement methods for the proposed methodology: (1) Mahalanobis distance matching with at most 10 maches (i.e., $J = 10$), (2) propensity score matching with $J = 10$, and (3) propensity score weighting. To make a fair comparison across different estimators, we include an identical set of covariates as well as lagged variables in OLS and the matching estimators. That is, under the scenarios of model misspecification (i.e., $L \neq 3$), the proposed matching estimator is also misspecified. Finally, we compute 90% confidence interval for each method in order to evaluate the its coverage rate.

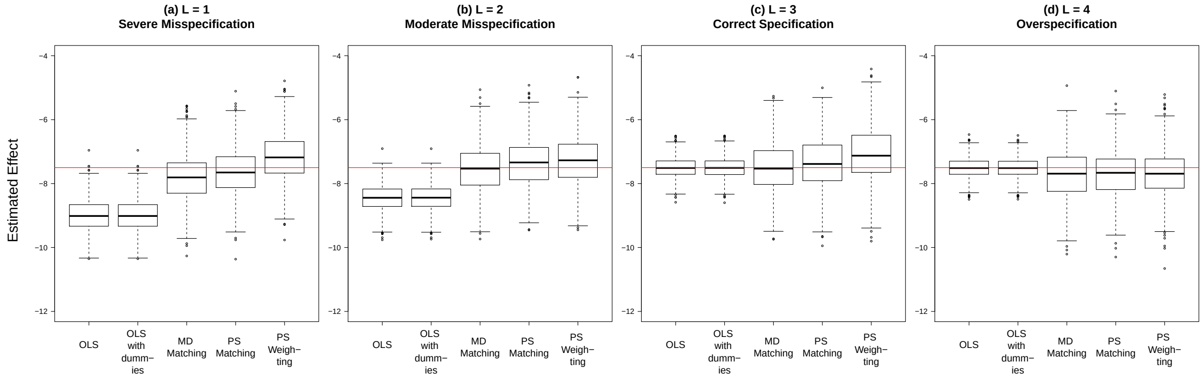

Figure C.1: **Robustness of the Proposed Methodology to Model Misspecification with $N = 162$.** This figure summarizes the simulation studies across four levels of misspecification. Panels (a) and (b) show that the ordinary least squares (OLS) estimator with unit and time fixed effects yields a significant bias as the severity of misspecification increases. So does OLS with dummies for treatment history. In contrast, the proposed methodology, based on three different refinement methods (Mahalanobis (MD) matching , Propensity score (PS) matching, and Propensity score (PS) weighting), returns similar estimates of the quantity of interest under the two different model misspecifications. Panel (c) and (d) shows that when the model is correctly specified or overspecified, the OLS estimators is unbiased and most efficient.

C.2 Results

Simulation results with $N = 162$. Figure C.1 presents the distribution of the estimated effects across 1,000 Monte Carlo simulations under each scenario for each estimator. Panels (a) and (b) show that the proposed methodology performs significantly better than the OLS estimator when the model is misspecified. Specifically, the OLS estimator tends to substantially underestimate the true contemporaneous treatment effect (the horizontal red line at -7.5) whereas our matching estimators yield relatively unbiased estimates regardless of the degree of model misspecification. Although OLS estimators are generally more efficient than the matching estimators, their variances also increase as the degree of model misspecification increases. As expected, when the model is correctly specified and overspecified, OLS is unbiased and most efficient as shown in panels (c) and (d). Note that the matching estimators tend to have similar variances across all scenarios as they are less sensitive to model misspecification in terms of both bias and variance.

Table C.1 further investigates the bias-variance tradeoff across the estimators. We find that under the two model misspecification scenarios considered here (i.e., $L = 1, 2$), the root mean squared error (RMSE) of the OLS estimator exceeds those of the matching estimators. This suggests that although the OLS estimators are more efficient than the matching estimators, their biases increase quickly once the model becomes misspecified. In contrast, the RMSE of the matching estimators stays relatively stable across model (mis)specifications considered here. Finally, the 90% confidence intervals of the proposed methodology maintain a reasonable coverage rate (Cov.) whereas the corresponding confidence intervals of the OLS

[1] We use a linear time trend for variables when 35% or less of the data is missing, while including a quadratic time trend when the missingness is more severe. We added a small error term from a normal distribution by setting the standard error to the standard deviation of the difference in the variable between two consecutive time periods.

[2] For the simulation analysis, $\mathbf{Z}_{it}$ includes the log population, log population of age below 16 years, the log population of age above 64 years, net financial flow as a fraction of GDP, trade volume as a fraction of GDP, and a dichotomous measure of social unrest. We use the log population as the single variable $\mathbf{Z}_{it}^{(3)}$ for the interaction term.

[3] We also conducted a simulation with further over-specification $L = 5$, which leads to essentially the same result as $L = 4$.

Method	L=1			L=2			L=3			L=4		
	Severe Misspecification			**Moderate Misspecification**			**Correct Specification**			**Overspecification**		
	Bias	RMSE	Cov	Bias	RMSE	Cov	Bias	RMSE	Cov	Bias	RMSE	Cov
OLS	-1.50	1.58	0.08	-0.95	1.03	0.27	-0.00	0.32	0.89	-0.01	0.30	0.91
OLS with dummies	-1.50	1.58	0.08	-0.94	1.03	0.27	-0.00	0.32	0.89	-0.01	0.30	0.91
MD. Matching	-0.30	0.80	0.96	-0.04	0.73	0.96	0.00	0.79	0.94	-0.21	0.84	0.94
PS. Matching	-0.13	0.74	0.94	0.12	0.74	0.95	0.13	0.80	0.94	-0.20	0.78	0.98
PS. Weighting	0.33	0.82	0.97	0.21	0.78	0.95	0.40	0.94	0.94	-0.20	0.76	0.96

Table C.1: **Results of Simulation Studies with $N = 162$.** The proposed methodology is denoted by "MD. Matching" for Mahalanobis distance matching, "PS. Matching (Weighting)" for propensity score matching (weighting). The OLS estimators are "OLS" for directly controlling for L lags of treatment and "OLS with dummies" for controlling for treatment history matrices. When the model is misspecified, the proposed methodology exhibits a smaller bias and Root Mean Square Error (RMSE), compared to the ordinary least squares OLS estimators. The OLS estimators are generally more biased but less variable than the matching and weighting methods, as shown by the smaller standard deviation (SD). The 90% confidence intervals of the proposed methodology produces reasonable coverage rates (Cov.) under all simulation scenarios whereas the OLS estimators result in substantial under-coverage unless the model is correctly specified or overspecified.

estimators have a poor coverage unless the model is correctly specified or overspecified. Overall, we find that the matching estimators outperform the OLS estimators unless the model is correctly specified.

Why does the matching estimator perform well even under the model misspecification? To examine this question, we investigate the covariate balance achieved by our matching methods. Figure C.2 shows the covariate balance for each refinement method (Mahalanobis distance matching in the top row, propensity score matching in the middle row, and propensity score weighting in the bottom row) under the four scenarios (columns). In each plot, we present the distribution of the average absolute standardized mean difference across covariates for contemporaneous ("L0") and each of the three lags ("L1" through "L4"). We find similar covariate balance across moderate misspecification, correct specification, and overspecification.The covariate balance under severe misspecification, on the other hand, is worse than the other three scenarios. This is consistent with the performance of the matching estimators presented in Figure C.1 and Table C.1.

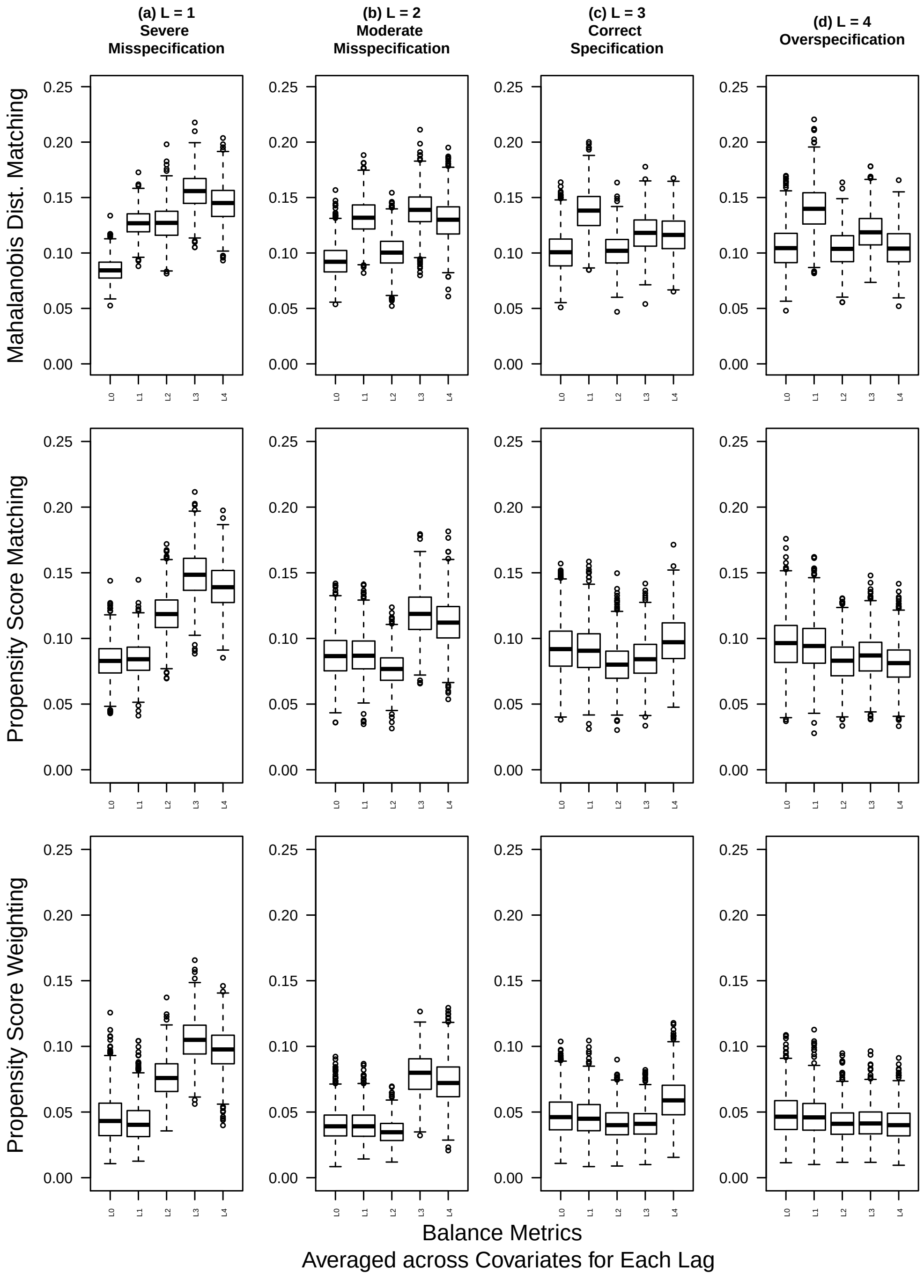

Figure C.2: **Covariate Balance Achieved by the Proposed Methodology under Model Misspecification with $N = 162$.** This figure summarizes the simulation studies of Mahalanobis distance matching, propensity score matching, and propensity score weighting in first through third row across four levels of misspecification in columns (a) through (d). For each method under each scenario, a plot shows the distribution of the average absolute standardized mean difference across all covariates for the contemporaneous period ("L0") and four lag periods ("L1" through "L4").

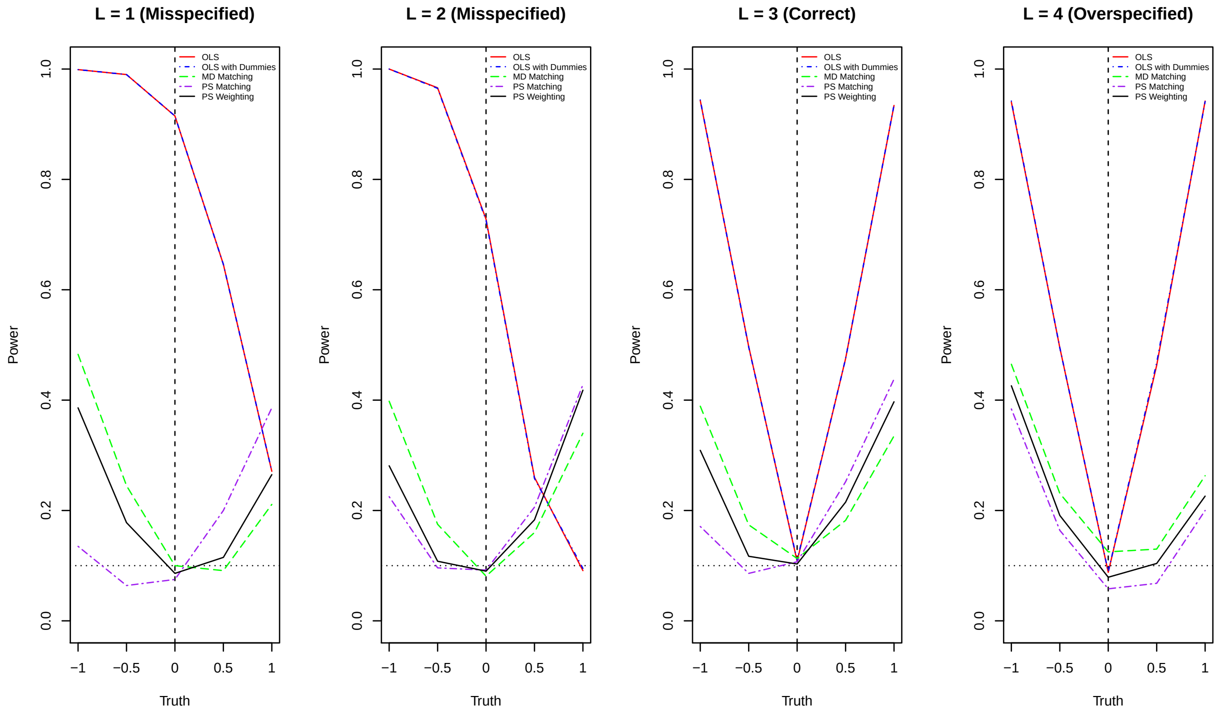

Figure C.3: **Statistical Power of the Proposed Methodology to Model Misspecification with $N = 162$.** This figure summarizes the simulation studies across four levels of misspecification. It shows the statistical power of each estimator with the 90% confidence level across five different truths, $-1, -0.5, 0, 0.5, 1$. The results indicate that the proposed methodology is more conservative than OLS.

The bias reduction of our matching methods does not come free. Indeed, as can be seen from Figure C.1 and Table C.1, the variance of the proposed matching estimators is typically greater than that of the least squares estimator. We illustrate this point by computing the statistical power of each estimator with the 90% confidence level as shown in Figure C.3 while varying the true value across various values from -1 to 1. We find that although the proposed matching estimators are less powerful than the least squares estimator across three scenarios, the statistical power of the latter depends on the true value in an asymmetrical fashion when the model is misspecified. Taken together, the strong parametric assumption of OLS reduces the variance and may increase statistical power. However, this results in the sensitivity to model misspecification. In contrast, matching has less bias even in the presence of misspecification though this comes at the cost of increased variance.

Simulation results with $N = 50$. We find similar results when the sample size is smaller, $N = 50$. As shown in Figure C.4 and Table C.2, the least squares are much more sensitive to the model misspecification than the proposed matching estimators. Figure C.5 shows that the covariate balance is reasonable so long as the model is not severely misspecified. As before, a better balance leads to a better performance of matching estimator. Finally, when the sample size is small, the statistical power of the proposed estimators further deteriorates (see Figure C.6) and the coverage of the confidence interval diverges from the nominal coverage even under correct model specification (see also Table C.2). This suggests that a large sample size is necessary for obtaining better uncertainty estimates.

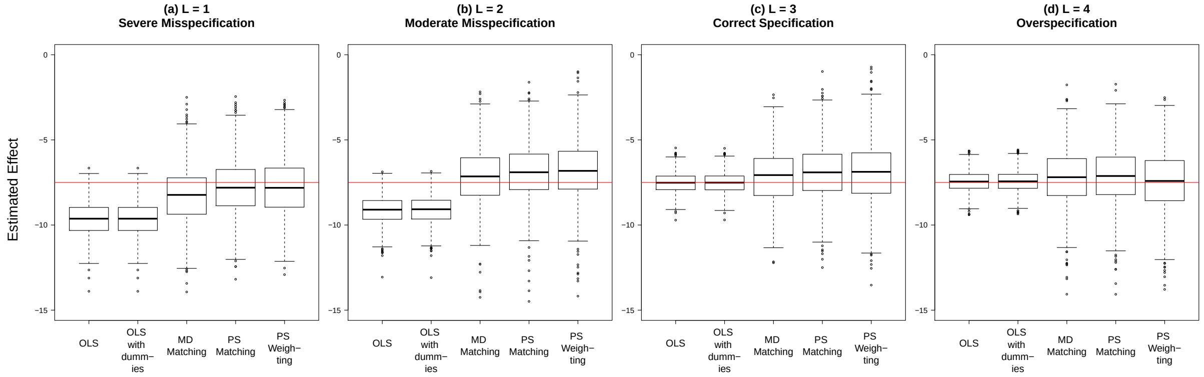

Figure C.4: **Robustness of the Proposed Methodology to Model Misspecification with $N = 50$.**

Method	L=1			L=2			L=3			L=4		
	Severe Misspecification			**Moderate Misspecification**			**Correct Specification**			**Overspecification**		
	Bias	RMSE	Cov	Bias	RMSE	Cov	Bias	RMSE	Cov	Bias	RMSE	Cov
OLS	-2.15	2.36	0.36	-1.64	1.85	0.50	-0.01	0.58	0.92	0.05	0.62	0.91
OLS with dummies	-2.15	2.36	0.36	-1.61	1.82	0.51	-0.01	0.59	0.91	0.06	0.62	0.90
MD. Matching	-0.74	1.83	0.95	0.34	1.65	0.90	0.35	1.65	0.89	0.27	1.73	0.90
PS. Matching	-0.29	1.66	0.94	0.59	1.72	0.88	0.57	1.74	0.91	0.33	1.74	0.93
PS. Weighting	-0.27	1.78	0.95	0.69	1.87	0.88	0.55	1.91	0.89	0.08	1.85	0.92

Table C.2: **Results of Simulation Studies with $N = 50$.**

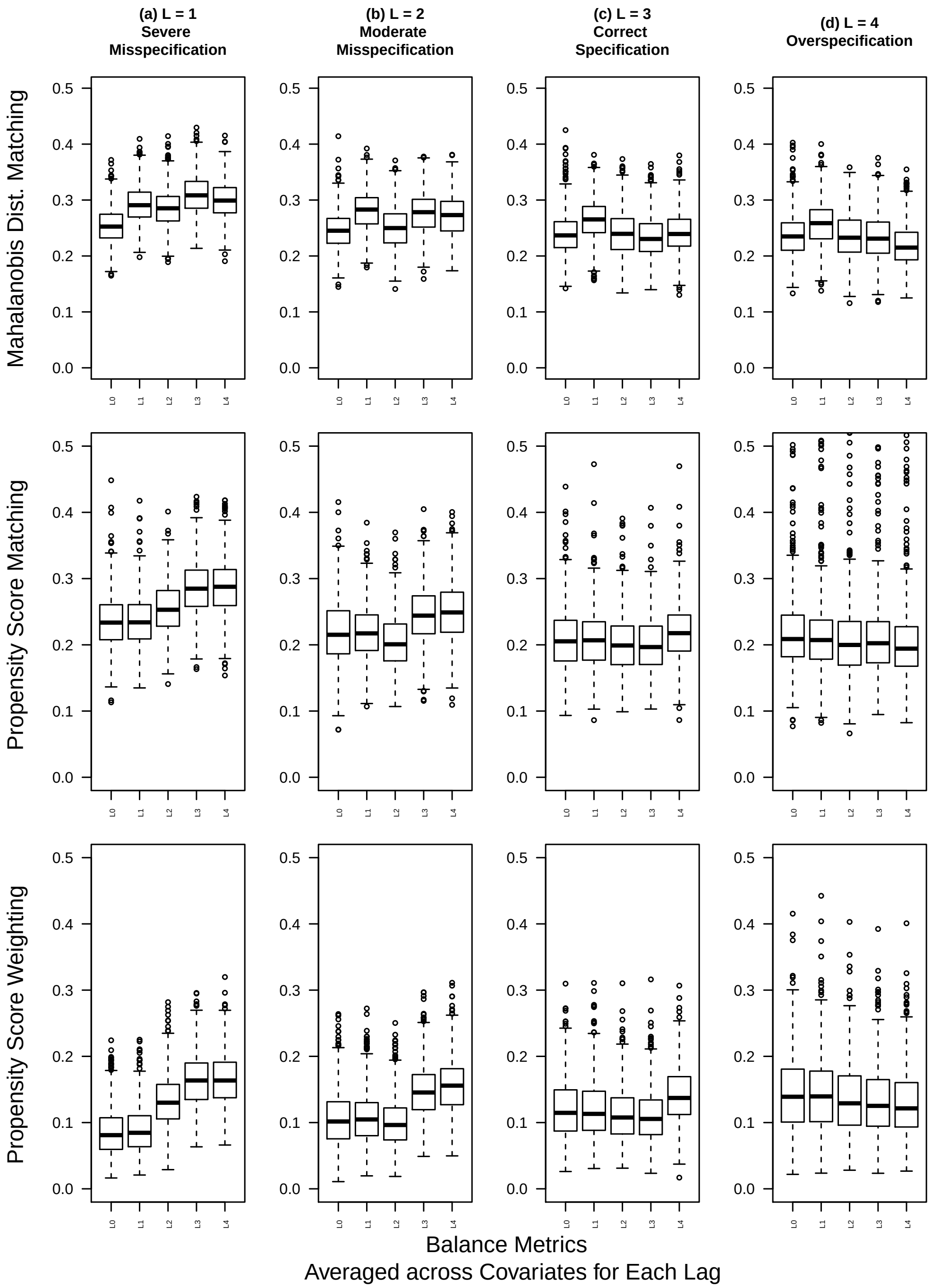

Figure C.5: **Covariate Balance Achieved by the Proposed Methodology under Model Misspecification, TRUTH = -7.5 and N = 50.** This figure summarizes the simulation studies of Mahalanobis distance matching, propensity score matching, and propensity score weighting in first through third row across four levels of misspecification in panels (a) through (d), respectively. For each method under each misspecification scenario, there is a graph showing the absolute standardized mean difference across all covariates at each l, summarized across simulations in a boxplot.

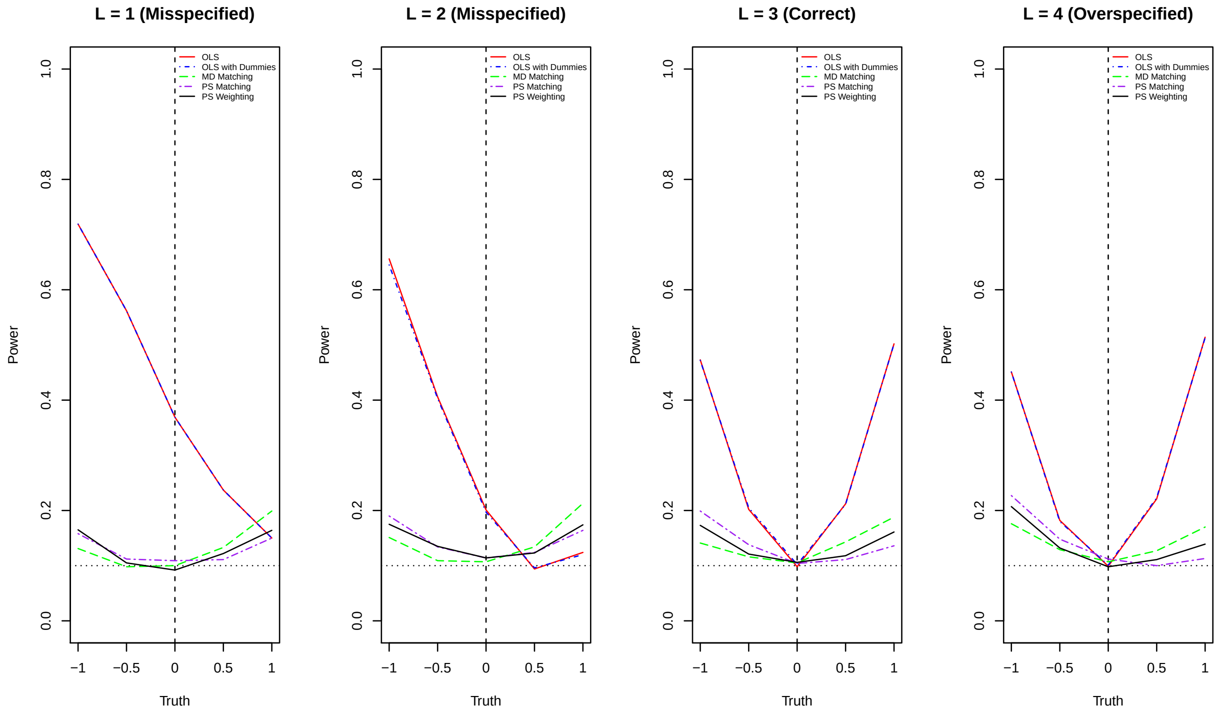

Figure C.6: **Statistical Power of the Proposed Methodology to Model Misspecification, N = 50.** This figure summarizes the simulation studies across three levels of misspecification. It shows the rate of Type II error with 90% confidence interval across five different truths, -1, -0.5, 0, 0.5, 1. The results indicate that the proposed methodology is more conservative than OLS.

Appendix D Covariate Balance when the Treatment Reversal is Not Allowed

This appendix presents the covariate balance for the two empirical applications in the case where treatment reversal is not allowed (as opposed to the cases in the main text, which allow for the treatment reversal). That is, we present the covariate balance for "stable policy change" as described in equation (1). Below, we find that the covariate balance for stable policy change is far from satisfactory. As such, the resulting causal estimates are likely to be less credible than those presented in the main text where the treatment reversal is allowed.

First, we present scatter plots for $F = 1, 2, 3, 4$ in Figures D.1— D.4.[4] Notice that the covariate balance for stable policy change in the case of $F = 1$ (shown in Figures D.1) already slightly deteriorates relative to its counterpart that allows for the treatment reversal in Figure 4. Notably, for the Scheve and Stasavage (2012) study, Figure D.1 shows that the off-diagonal post-refinement covariate balance (those above the 45-degree line) with "Four Year Lags" tends to be further away from the 45-degree line compared to its counterpart in Figure 4, while the balance in general gets worse with propensity score matching (second row) with "One Year Lag."

Moverover, the covariate balance for the same study in the case of $F = 4$ (see Figure D.4) exhibits

[4] Note that when $F = 0$, we are estimating the contemporaneous effects and hence the treatment reversal does not matter. In this case, therefore, the covariate balance figures (both scatter and line plots) are identical to Figures 4 and 5.

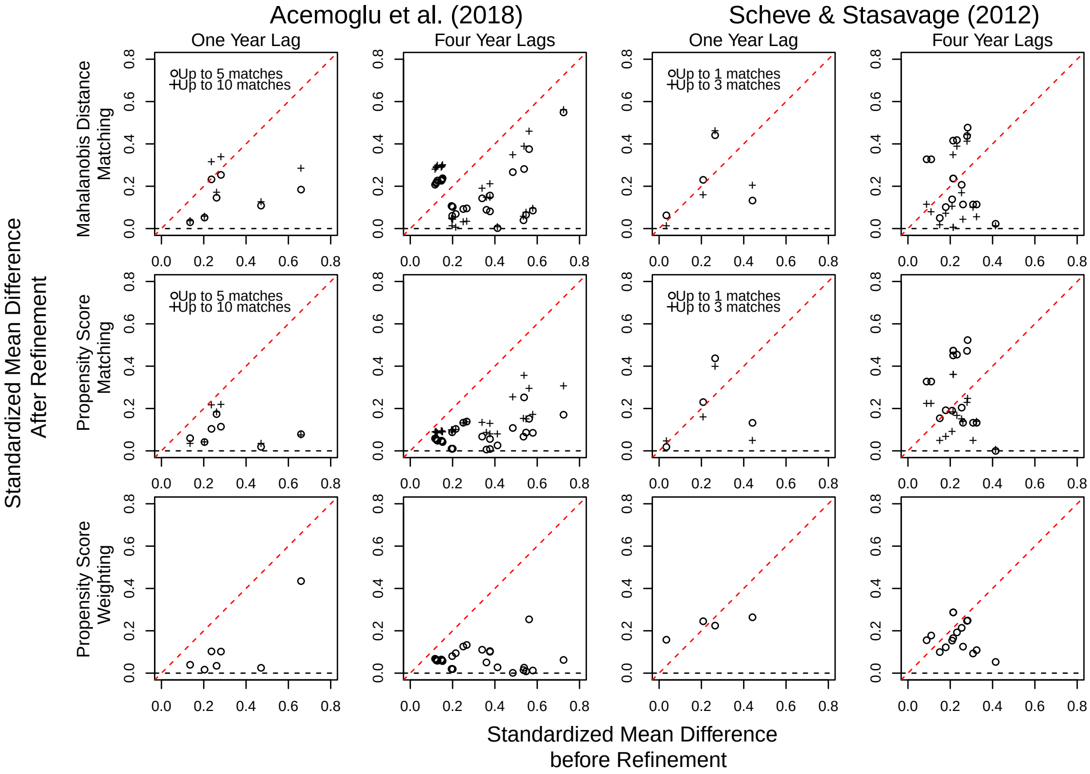

Figure D.1: **Covariate Balance due to the Refinement of Matched Sets when Estimating the Average Effects of Stable Policy Change, with** $F = 1$. See the caption of Figure 4.

further deterioration regardless of the matching methods and the choice of lags. For instance, several covariates have balances that are outside the range of the graph for all three methods with "One Year Lag." As for "Four Year Lags," the deterioration is also clear across methods for the Scheve and Stasavage (2012) study. Note that in contrast, balances for the Acemoglu *et al.* (2019) study show deterioration that is less severe across methods and the choices of lag.

Next, we present the line plots in Figures D.5— D.8. To begin, a comparison between Figures D.5 and 5 demonstrates that the covariate balance based on Mahalanobis matching for the authoritarian reversal treatment in the Acemoglu *et al.* (2019) study is substantially worse. Notice that covariate imbalance exacerbates further as F increases to 2, 3, and 4. For example, when $F = 4$ (see Figure D.8), the covariate balance lines for Mahalanobis distance matching, propensity score matching, and propensity score weighting are much further away from zero when compared to their counterparts in Figure 5 for the case of authoritarian reversal (second row).

Similarly, we observe a clear deterioration in covariate balance for the Scheve and Stasavage (2012) study (third row) when $F = 4$ for Mahalanobis distance matching and propensity score weighting. In addition, the number of unmatched treated observations increases when we do not allow for treatment reversal because a unit with treatment reversal no longer qualifies as a control unit. In the authoritarian reversal scenario, for example, the number of unmatched treated observations is 11, 15, 19, 22 for F = 1, 2, 3, 4, respectively. In the case of starting war as the treatment, the number of unmatched treated observations is 7, 8, 9, 19 for F = 1, 2, 3, 4 using four year lags.[5]

[5] Using one year lag, the numbers are 6, 8, 9, 19 for the four post-treatment periods, respectively.

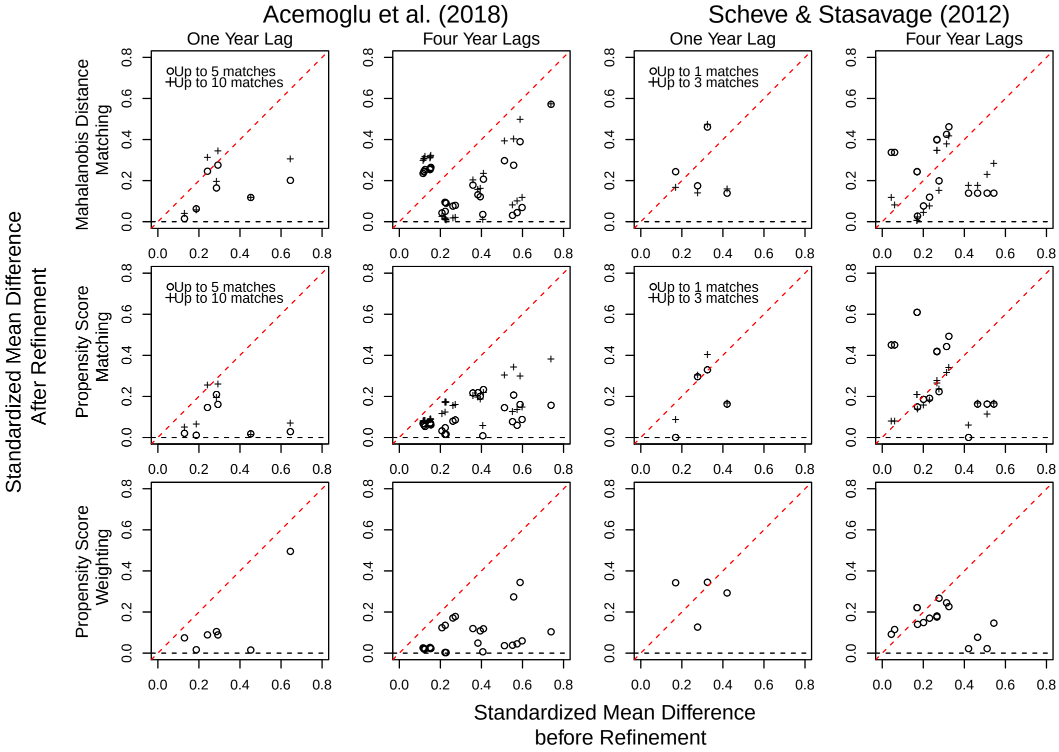

Figure D.2: **Covariate Balance due to the Refinement of Matched Sets when Estimating the Average Effects of Stable Policy Change, with** $F = 2$. See the caption of Figure 4.

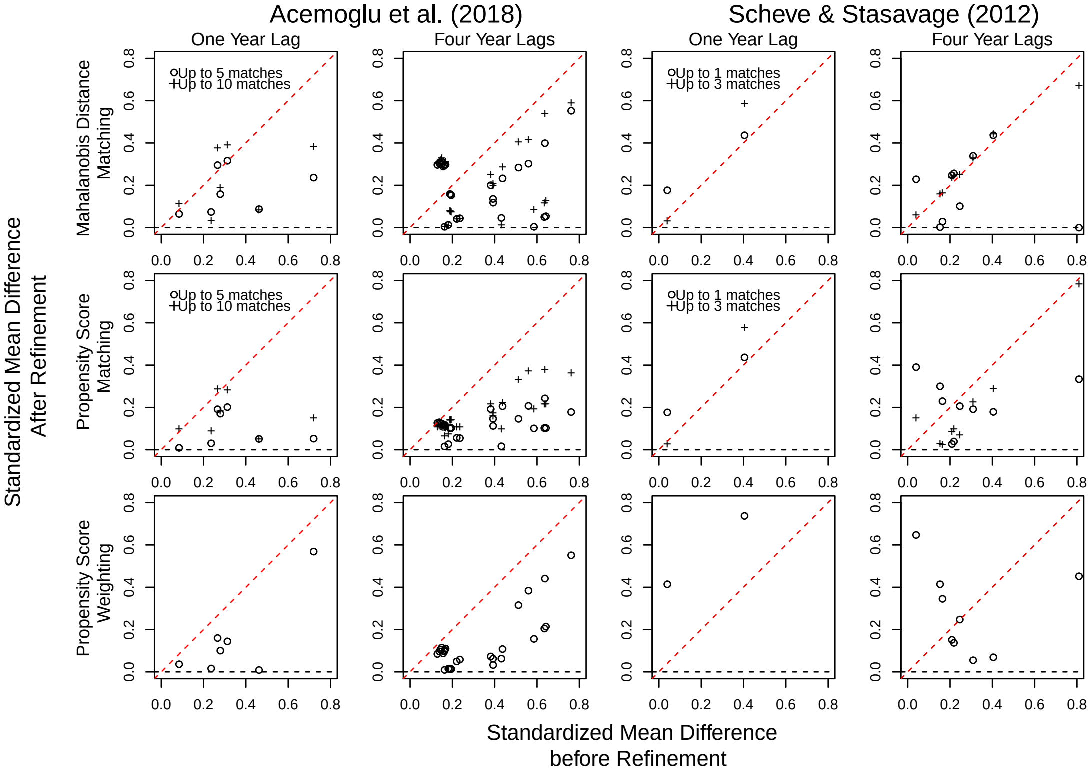

Figure D.4: **Covariate Balance due to the Refinement of Matched Sets when Estimating the Average Effects of Stable Policy Change, with** $F = 4$. See the caption of Figure 4.

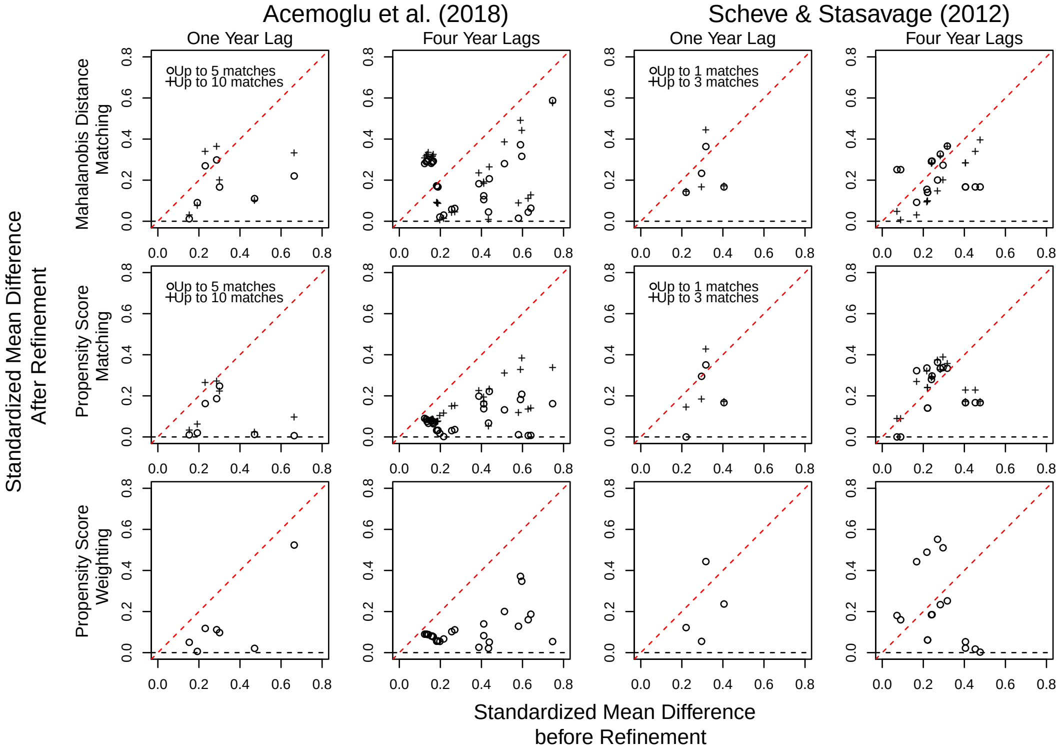

Figure D.3: **Covariate Balance due to the Refinement of Matched Sets when Estimating the Average Effects of Stable Policy Change, with** $F = 3$. See the caption of Figure 4.

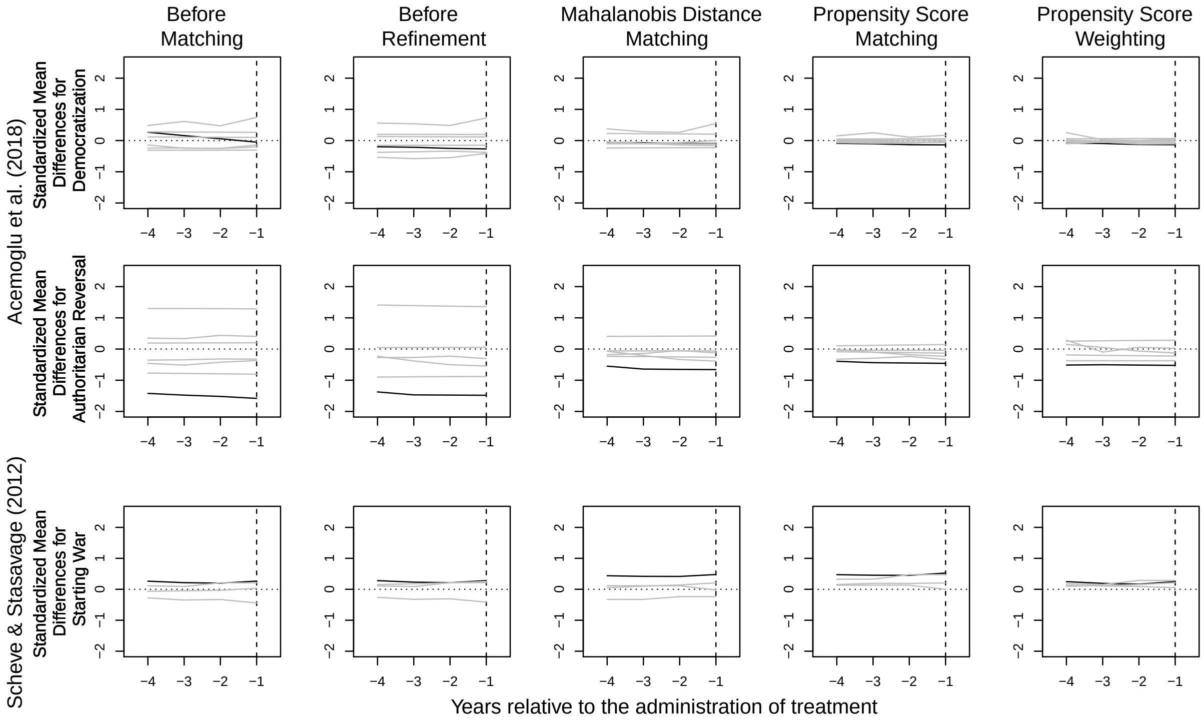

Figure D.5: **Improved Covariate Balance due to Matching over the Pre-Treatment Time Period when Estimating the Average Effects of Stable Policy Change,** $F = 1$. See the caption of Figure 5.

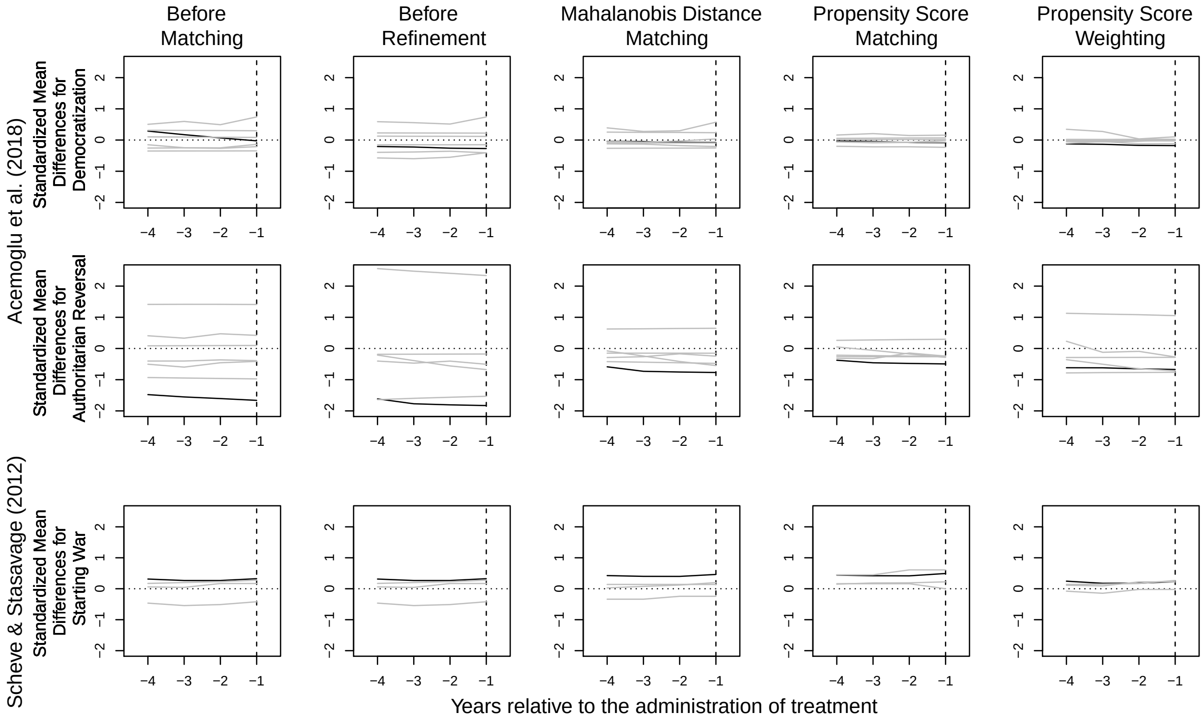

Figure D.6: **Improved Covariate Balance due to Matching over the Pre-Treatment Time Period when Estimating the Average Effects of Stable Policy Change,** $F = 2$. See the caption of Figure 5.

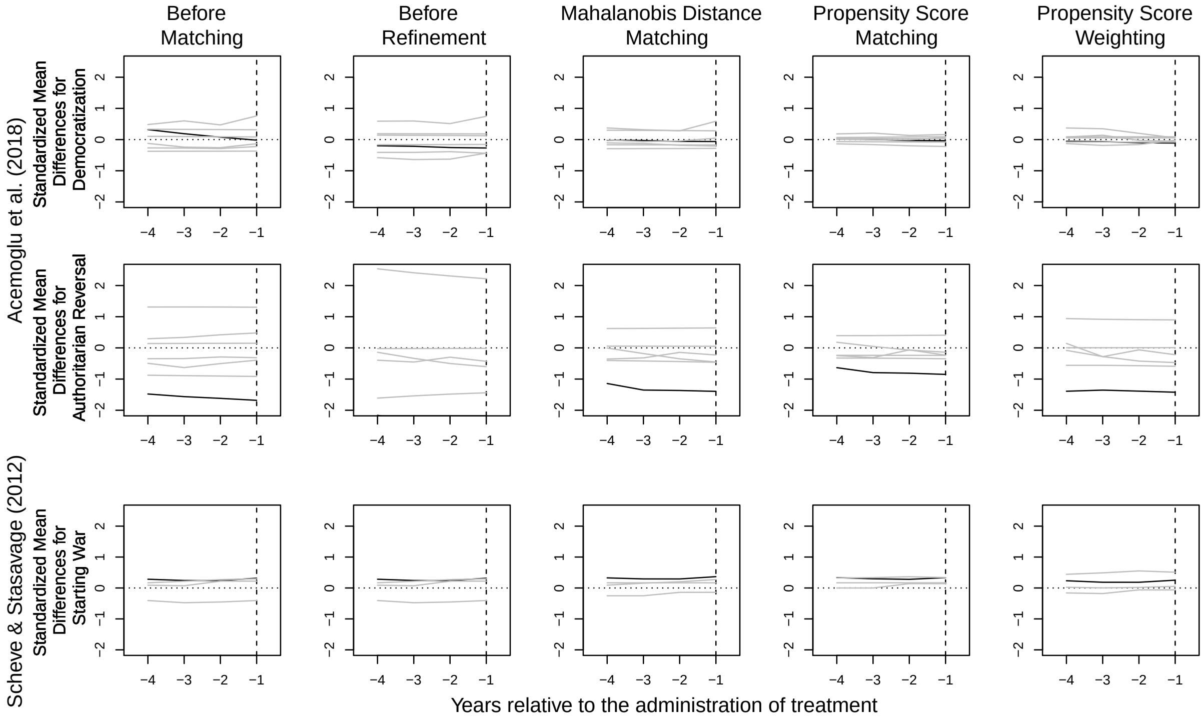

Figure D.7: **Improved Covariate Balance due to Matching over the Pre-Treatment Time Period when Estimating the Average Effects of Stable Policy Change,** $F = 3$. See the caption of Figure 5.

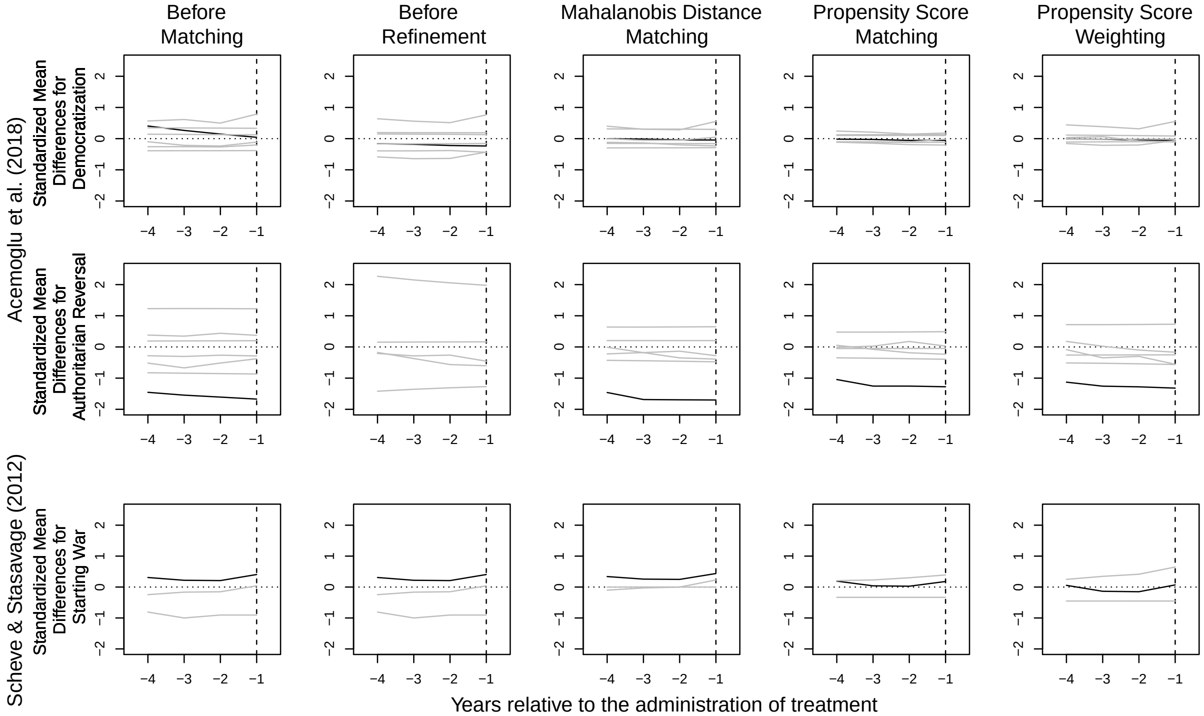

Figure D.8: **Improved Covariate Balance due to Matching over the Pre-Treatment Time Period when Estimating the Average Effects of Stable Policy Change,** $F = 4$. See the caption of Figure 5.

Appendix E The Results based on One Year Lag

This section presents the estimated effects of democracy on development using one year lag instead of four year lags as shown in Figure 6). Our findings remain substantively unchanged. We find that democracy has a positive effect on economic development not because transitioning to democracy improves a country's economic prospects, but because backsliding into autocracy worsens a country's economic development.

Appendix F The Estimated Effects of Democratization and Authoritarian Reversal based on the Linear Regression Models

This Appendix presents the estimated effects of democratic transition and authoritarian reversal, using the linear regression approach of the original analysis (Acemoglu *et al.*, 2019). We begin by replicating the results reported in Section A6.3 of the appendix of the original study. In that analysis, democratization and authoritarian reversal are coded as follows. For each country, both variables take the value of zero in the beginning period. Throughout the subsequent periods, whenever X_{it} changes from 0 to 1, the value of the democratization variable will increase by one and stays at that value until X_{it} changes from 0 to 1 again. Similarly, the authoritarian reversal variable starts with the value of zero and increases by one whenever X_{it} changes from 0 to 1.

In the original study, the authors then fit the two-way fixed effects linear regression models using the least squares and GMM estimation. They include the lagged outcome variables but no other covariate. The estimated coefficients for the democratization and authoritarian reversal variables are then interpreted as their respective average causal effects. One issue with this approach is the assumption that the effects

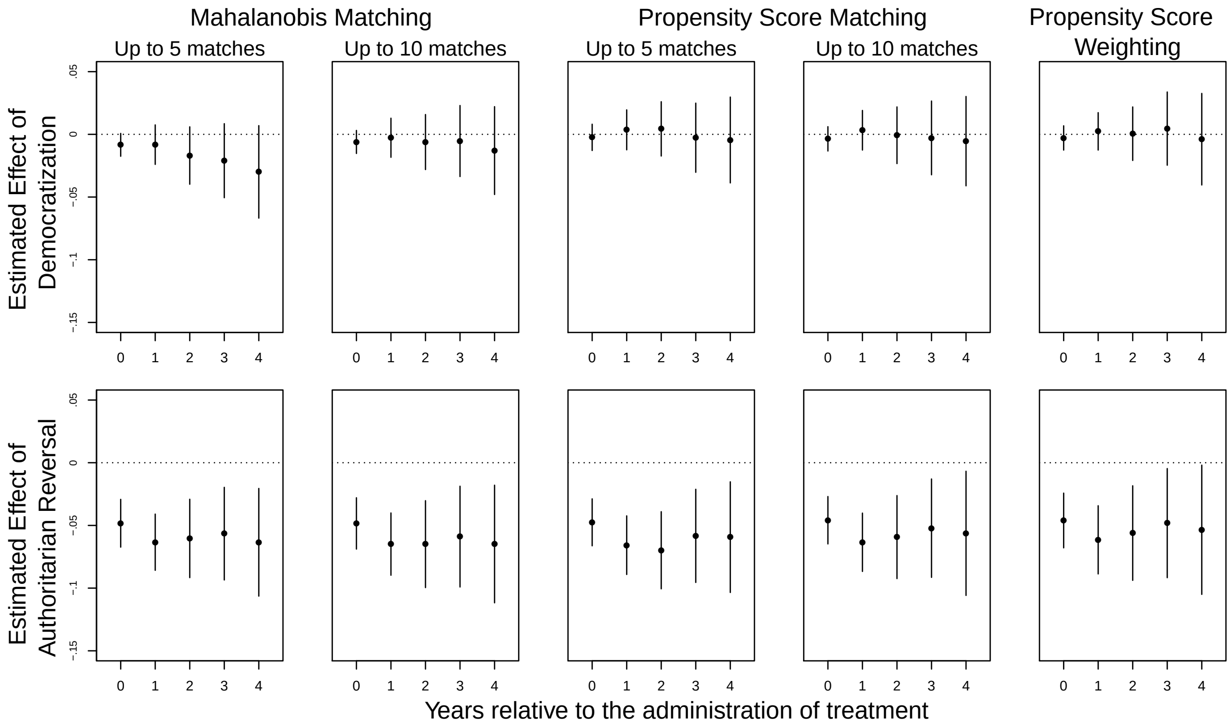

Figure E.1: **Estimated Average Effects of Democracy on Logged GDP per Capita when the Treatment Reversal is Allowed and Adjusting for One Year Pre-treatment Period.** The matching method adjusts for the treatment and covariate histories during the one year period prior to the treatment, i.e., $L = 1$. See the caption of Figure 6.

of democratization as well as those of authoritarian reversal are additive, regardless of the past history of regime changes. This contrasts with our approach where we nonparametrically adjust for the past treatment history.

The first and second columns of Table F.1 reproduces the estimates reported in the original study whereas the third and fourth columns report the estimates based on the models that include additional covariates. The results suggest that the effects of democratization are similar, in their magnitude, to those of authoritarian reversal (their signs are opposite as expected). However, the former is more precisely estimated than the latter. These results are qualitatively different from those based on our approach. We find that the economic effects of democracy are largely driven by the negative effects of authoritarian reversal rather than the positive effects of democratization. In contrast, the original analysis shows that the positive effects of democratization plays a more significant role.

	(1)	(2)	(3)	(4)
ATT	0.8033^{***}	1.470^{***}	0.6706^{**}	0.955^{**}
	(0.2381)	(0.543)	(0.3139)	(0.472)
ART	-0.7054^{**}	-1.313	-0.6459	-0.825
	(0.3398)	(0.957)	(0.4487)	(0.571)
$\hat{\rho}_1$	1.2380^{***}	1.204^{***}	1.0980^{***}	1.046^{***}
	(0.0381)	(0.0463)	(0.0416)	(0.0426)
$\hat{\rho}_2$	-0.2065^{***}	-0.192^{***}	-0.1331^{***}	-0.121^{***}
	(0.0464)	(0.0469)	(0.0405)	(0.0379)
$\hat{\rho}_3$	-0.0261	-0.0276	0.0053	0.0139
	(0.0286)	(0.0276)	(0.0296)	(0.0286)
$\hat{\rho}_4$	-0.0424^{**}	-0.0382^{*}	-0.0311	-0.0175
	(0.0176)	(0.0210)	(0.0239)	(0.0232)
country FE	Yes	Yes	Yes	Yes
time FE	Yes	Yes	Yes	Yes
covariates	No	No	Yes	Yes
estimation	OLS	GMM	OLS	GMM
N	6,336	6,161	4,416	4,245

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
robust standard errors clustered by prefecture in parentheses

Table F.1: **The Effects of Democracy on Growth with Lagged Treatments**: This table presents the estimated effects of transition to democracy from authoritarian regime (labeled as "ATT") and vice versa (labeled as "ART"). The control variables in Columns (2) and (4) are those in column (3) of Table 1

References

Acemoglu, D., Naidu, S., Restrepo, P., and Robinson, J. A. (2019). Democracy does cause growth. *Journal of Political Economy* **127**, 1, 47–100.

Robins, J. M., Hernán, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 5, 550–560.

Scheve, K. and Stasavage, D. (2012). Democracy, war, and wealth: lessons from two centuries of inheritance taxation. *American Political Science Review* **106**, 1, 81–102.