Matching Estimators with Few Treated and Many Control Observations

Ferman, Bruno

Sao Paulo School of Economics - FGV

4 May 2017

Matching Estimators with Few Treated and Many Control Observations.*

Bruno Ferman†

Sao Paulo School of Economics - FGV

First Draft: May, 2017

Please click here for the most recent version

Abstract

We analyze the properties of matching estimators when the number of treated observations is fixed while the number of treated observations is large. We show that, under standard assumptions, the nearest neighbor matching estimator for the average treatment effect on the treated is asymptotically unbiased, even though this estimator is not consistent. We also provide a test based on the theory of randomization tests under approximate symmetry developed in Canay et al. (2014) that is asymptotically valid when the number of control observations goes to infinity. This is important because large sample inferential techniques developed in Abadie and Imbens (2006) would not be valid in this setting.

Keywords: matching estimator, treatment effect, hypothesis testing, randomization inference
JEL Codes: C12; C13; C21

*I would like to thank Ricardo Masini for comments and suggestions.
†bruno.ferman@fgv.br

1 Introduction

Matching estimators have been widely used for the estimation of treatment effects under a conditional independence assumption (CIA).[1] In many cases, matching estimators have been applied in settings where (1) the interest is in the average treatment effect for the treated, and (2) there is a large reservoir of potential controls (Imbens and Wooldridge (2009)). Abadie and Imbens (2006) study the theoretical properties of matching estimators when the number of control observations grow at a higher rate than the number of treated observations. However, their asymptotic results still depend on both the number of treated and control observations going to infinity.

In this paper, we analyze the properties of matching estimators when the number of treated observations is fixed while the number of treated observations is large. We show that the nearest neighbor matching estimator is asymptotically unbiased for the average treatment effect on the treated under standard assumptions used in the literature on estimation of treatment effects under selection on unobservables. This result is consistent with Abadie and Imbens (2006), who show that the conditional bias of the matching estimator can be ignored provided that the number of control observations increases faster enough relative to the number of treated observations. In their setting, the matching estimator would be consistent and asymptotically normal. Differently from Abadie and Imbens (2006), since we assume that the number of treated observations is fixed, we find that, in our setting, the variance of the matching estimator does not converge to zero and that the estimator will not generally be asymptotically normal. Our theoretical results should provide a better approximation to the behavior of the matching estimator relative to Abadie and Imbens (2006) in settings in which not only there is a much larger number of control observations relative to treated observations, but also the number of treated observations are not large enough, so that we cannot rely on asymptotic results.[2]

We also provide a randomization inference procedure based on Canay et al. (2014) that provides asymptotically valid hypothesis testing when the number of control observations goes to infinity, even when the number of treated observations is fixed. The main intuition of the test is that it is possible to construct estimators of the treatment effect for each treated observation and, asymptotically, these estimators are independent and symmetric around zero under the null. Therefore, the joint distribution of these estimators is asymptotically invariant to sign changes. This result is important because inferential techniques that rely on large samples as developed in Abadie and Imbens (2006) might not be appropriate when the number

[1] see Imbens (2004), Imbens and Wooldridge (2009), and Imbens (2014) for reviews.

[2] The finite sample properties of matching estimators have also been evaluated in simulations in Frolich (2004) and Busso et al. (2014). Note that we provide theoretical results holding the number of treated observations fixed, but we rely on the number of control observations going to infinity.

of treated observations is not large. Moreover, bootstrap methods are not generally valid for matching estimators even in large samples (Abadie and Imbens (2008)). In finite samples, Rosenbaum (1984) and Rosenbaum (2002) consider permutation tests for observational studies under strong ignorability. However, these tests rely on strong assumptions.[3] In contrast, our inference method is asymptotically valid under standard assumptions in the literature on treatment effects under unconfoundedness.

Our setting is also related to the Synthetic Control (SC) method, which is an alternative to estimate treatment effects in comparative case studies (Abadie and Gardeazabal (2003), Abadie et al. (2010), and Abadie et al. (2015)). Similar to the matching estimator, the SC method aims to match the pre-treatment outcomes of the treated unit using a convex combination of the control units. Daz et al. (2015) explore the idea of a matching estimator that considers convex combinations of the characteristics of the individuals in the corresponding counterfactual. In this sense, the SC estimator would be a matching estimator as in Daz et al. (2015) using the pre-treatment outcomes as covariates. Our results show that, if treatment assignment is "as good as random" conditional on the pre-treatment outcomes, then a SC estimator that assigns weights using Daz et al. (2015) procedure should be asymptotically unbiased when the number of control units goes to infinity, even if the number of pre-treatment periods is fixed.[4] Moreover, we provide a test that is asymptotically valid when the number of control units goes to infinity, provided that the number of treated units is not too small.[5,6] In contrast, Ferman and Pinto (2017) show that the placebo tests proposed in Abadie et al. (2010) will not, in general, satisfy the approximate symmetry property required in Canay et al. (2014).[7] Therefore, our test can be a valid alternative to the placebo tests proposed in Abadie et al. (2010) for the SC estimator when there are multiple treated units and a large number of control units.

[3] Rosenbaum (1984) assumes that the propensity score follows a logit model, while Rosenbaum (2002) assumes that observations are matched in pairs such that the probability of treatment assignment is the same conditional on the pair.

[4] If however, treatment assignment is only "as good as random" conditional on common factors (which allows for some correlation between treatment assignment and post-treatment potential outcomes), then this would not be necessarily true. Gobillon and Magnac (2016) show that the SC estimator can be asymptotically unbiased if the number of control units and the number of pre-treatment periods go to infinity, while Abadie et al. (2010) show that, conditional on a perfect pre-treatment match, the bias of the SC estimator is bounded by a function that goes to zero when the number of pre-treatment periods increases, even if the number of control units is fixed. See also Ferman and Pinto (2016) for a discussion on the conditions for asymptotic unbiasedness for the SC estimator when the number of control units is fixed.

[5] While still valid, our test would have poor power when the number of treated units is too small. The same is true for the differences-in-differences inference method proposed in Canay et al. (2014).

[6] Note that this test should only be asymptotically valid if we use Daz et al. (2015) procedure to calculate the SC weights. Their procedure will guarantee that the SC unit for each treated unit will assign positive weights to only few donors, which will imply that the treatment effect estimators for each treated unit will be independent. If we use Abadie et al. (2010) original procedure, it is not clear that control units will, asymptotically, receive positive weights to only one treated unit.

[7] See also Firpo and Possebom (2016) and Hahn and Shi (2016) for considerations on the placebo tests proposed in Abadie et al. (2010).

2 Setting and Notation

We observe a sample $\{Y_i, X_i\}_{i=1}^{N_1}$ that receives treatment ($W_i = 1$) and a sample $\{Y_i, X_i\}_{i=N_1+1}^{N}$ that does not receive treatment ($W_i = 0$), where Y_i is the observed outcome of observation i, and X_i is a set of covariates. We assume that X_i is a continuous random vector of dimension k in $\mathbb{R}^k$.[8] Following Rubin (1973), let $Y_i(1)$ denote the potential outcome if observation i received treatment and $Y_i(0)$ denote the potential outcome if observation i did not receive treatment. We consider the case in which the number of treated observations, N_1, is finite, while the number of control observations, $N_0 = N - N_1$, is large. Let $\mathcal{I}_w$ denote the set of indexes for observations with $W_i = w$. We aim to estimate the treatment effect on the treated, which we denote by:

$$\tau = \frac{1}{N_1} \sum_{i \in \mathcal{I}_1} \mathbb{E}\left[Y_i(1) - Y_i(0) | X_i\right] \tag{1}$$

Note that we focus on the estimation of the treatment effect on the treated because, given our setting with N_1 finite and N_0 large, there is no hope in constructing a counterfactual for the control observations using only a finite set of treated observations. Also, for most of our results we will consider the properties of the matching estimator conditional on the realization of $\{X_i\}_{i \in \mathcal{I}_1}$.[9] We consider the unconditional case in remark 2.

We present our main assumptions in a slightly different way relative to Abadie and Imbens (2006) in order to consider the case with the number of control observations going to infinity while the number of treated observations is fixed. The main intuition behind our assumptions, however, remain the same.

We assume that the sample we observe for the treated (control) observations consists of i.i.d. observations of individuals with $W_i = 1$ ($W_i = 0$), and that treated and control observations are independent.

Assumption 1 (Sample) *$\{Y_i(0), Y_i(1), X_i\}_{i \in \mathcal{I}_w}$ consists of N_w i.i.d. observations with $W_i = w$. Furthermore, we assume that individuals in the treated and control samples are independent.*

The following assumption restricts the way in which the distributions of the treatment and control observations may differ.

[8] We abstract from the case in which components of X_i is discrete because, as argued in Abadie and Imbens (2006), discrete covariates with a finite number of support points can be easily dealt with by analyzing estimation of average treatment effects within subsamples defined by their values.

[9] Note that our analysis is a mixture of finite sample (N_1 is finite) and large sample ($N_0 \to \infty$), which is similar to the setting considered in Ferman and Pinto (2015). We consider our results conditional on the realization of the treated covariates in an analogy to what is usually done in the study of finite sample properties of estimators.

Assumption 2 ("Conditional Independence Assumption") *Conditional on X_i, the distribution of $Y_i(0)$ is the same for i in the treated and in the control groups.*

Assumption 2 is equivalent to the conditional independence assumption (CIA).[10] While in assumption 1 we allow for different distributions of $(Y_i(0), Y_i(1), X_i)$ whether i is treated or control, assumption 2 restricts that the conditional distribution of $Y_i(0)$ given X_i is the same for both treatment and control observations. However, the density $f_1(X_i)$ for $i \in \mathcal{I}_1$ can potentially be different from the density $f_0(X_i)$ for $i \in \mathcal{I}_0$. This is what generates potential bias in a simple comparison between treated and control groups, without taking into account that these groups might have different distributions of covariates X_i.

The next assumption states that possible values of X_i for the treated observations are in the support of X_i for the control observations.

Assumption 3 (Overlap) $\mathbb{X}_1 \subset \mathbb{X}_0$, *where* $\mathbb{X}_w$ *is the support of* $f_w(X_i)$, *for* $w \in \{0, 1\}$

Assumption 3 replaces the standard assumption that $Pr(W = 1|X = x) < 1 - \eta$ for some $\eta > 0$. This assumption will guarantee that, for each i in the treated group, we will be able to find an observation j in the control group with covariates X_j arbitrarily close to X_i when $N_0 \to \infty$.

The main identification problem arises from the fact that we observe either $Y_i(1)$ or $Y_i(0)$ for each observation i. Note that, if we had two observations, $i \in \mathcal{I}_1$ and $j \in \mathcal{I}_0$, with $X_i = X_j = x$, then, under assumption 2, $\mathbb{E}[Y_i|W_i = 1, X_i = x] - \mathbb{E}[Y_j|W_j = 0, X_j = x] = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x]$. That is, we would be able estimate the average treatment effect conditional on each value of the covariates $X_i = x$. Then we would be able to aggregate these effects to construct the average treatment effect or the average treatment effect on the treated. The main challenge is that, with a continuous random variable X_i, the probability of finding observations with exactly the same X_i is zero. The idea of the nearest neighbor matching estimator is to input the missing potential outcomes of a treated observation $i \in \mathcal{I}_1$ with observations in the control group $j \in \mathcal{I}_0$ that are as close as possible in terms of covariates X_i. More specifically, for a given metric $d(a, b)$ in $\mathbb{R}^k$, let $\mathcal{J}_M(i)$ be the set of M nearest neighbors in the control group of observation $i \in \mathcal{I}_1$. Then the matching estimator is given by:

$$\hat{\tau} = \frac{1}{N_1} \sum_{i \in \mathcal{I}_1} \left[Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j \right] \tag{2}$$

[10] Since we treat W_i as non-random, we cannot talk about independence between potential outcomes and W_i.

3 Asymptotic Unbiasedness

Following Abadie and Imbens (2006), for $w \in \{0,1\}$, we define $\mu(x,w) = \mathbb{E}[Y|X=x, W=w]$, $\mu_w(x) = \mathbb{E}[Y(w)|X=x]$, and $\epsilon_i = Y_i - \mu_{W_i}(X_i)$. Note that, under assumption 2, we have that $\mu(x,w) = \mu_w(x)$.

Using this notation, we have that:

$$\hat{\tau} = \frac{1}{N_1} \sum_{i \in \mathcal{I}_1} \left[\left(\mu_1(X_i) - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} \mu_0(X_j) \right) + \left(\epsilon_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} \epsilon_j \right) \right] \tag{3}$$

We show that $\hat{\tau}$ is an asymptotically unbiased estimator for the treatment effect on the treated.

Proposition 1 *Under assumptions 1, 2, and 3, if $\mu_0(x)$ is continuous and bounded, then:*

1. $\mathbb{E}[\hat{\tau}|\{X_i\}_{i \in \mathcal{I}_1}] \to \frac{1}{N_1} \sum_{i \in \mathcal{I}_1} [\mu_1(X_i) - \mu_0(X_i)]$
2. *If we add that $\tilde{f}(x) = \mathbb{E}[f(Y(0))|X=x]$ is continuous and bounded for any $f(y)$ continuous and bounded, then, conditional on $\{X_i\}_{i \in \mathcal{I}_1}$:*

$$\hat{\tau} \xrightarrow{d} \frac{1}{N_1} \sum_{i \in \mathcal{I}_1} (\mu_1(X_i) - \mu_0(X_i)) + \frac{1}{N_1} \sum_{i \in \mathcal{I}_1} \left(\epsilon_i - \frac{1}{M} \sum_{m=1}^{M} \epsilon_m(X_i) \right) \tag{4}$$

where $\epsilon_m(X_i) \overset{d}{=} Y_i(0)|X_i - \mu_0(X_i)$ for $i \in \mathcal{I}_1$, and $\epsilon_m(X_i)$ is independent across m and i.

Proof. Let $X^i_{(m)}$ be the covariate value of the m-closest match to observation i. The main intuition of this result is that, for a fixed $X_i = \bar{x}$, $X^i_{(m)} \xrightarrow{p} \bar{x}$ when $N_0 \to \infty$, because we will always be able to find M observations in the control group that are arbitrarily close to $\bar{x}$. Independence of $\epsilon_m(X_i)$ across m and i follows from the fact that the probability of two treated observations sharing the same nearest neighbor converges to zero. See details in Appendix A. ■

Proposition 1 shows that, conditional on the realization of $\{X_i\}_{i \in \mathcal{I}_1}$, the expected value of the matching estimator converges to $\tau = \frac{1}{N_1} \sum_{i \in \mathcal{I}_1} (\mu_1(X_i) - \mu_0(X_i))$, which is the average treatment effect on the treated. With an additional assumption on the conditional distribution of $Y(0)$ given X, we can derive the asymptotic distribution of the matching estimator, which is centered on $\mathbb{E}[\hat{\tau}|\{X_i\}_{i \in \mathcal{I}_1}]$. This result is important for the construction of the inference method we propose in Section 4.

Remark 1 The condition that $\mu_0(x)$ is continuous and bounded would be satisfied if we assume that $\mu_0(x)$ is continuous and $\mathbb{X}_0$ is compact, as is assumed in Abadie and Imbens (2006).

Remark 2 We focus on the properties of the matching estimator conditional on $\{X_i\}_{i\in\mathcal{I}_1}$. We might be interested, however, on the unconditional properties of the matching estimator. For example, we may think that $\{Y_i, X_i\}_{i\in\mathcal{I}_1}$ is a finite sample from a super population.[11] Under the assumptions from Proposition 1, we also have that $\mathbb{E}[\hat{\tau}] = \mathbb{E}\left\{\mathbb{E}[\hat{\tau}|\{X_i\}_{i\in\mathcal{I}_1}]\right\}$ converges to $\mathbb{E}[\mu_1(X_i) - \mu_0(X_i)|i \in \mathcal{I}_1]$, which is the average treatment effect on the treated *population*. Alternatively, we may think that there is indeed a finite N_1 population of treated individuals, but these individuals were selected to receive treatment from a larger population. See details in Appendix A.

Remark 3 With N_1 fixed, the estimator is not consistent. This happens because, with a fixed number of treated observations, we cannot apply a law of large numbers to the average of the error of the treated observations. Also, the matching estimator will not be asymptotically normal, unless we assume that the error ϵ_i is normal.

Remark 4 The nearest-neighbor matching estimator is not, in general, unbiased for a fixed N_0. This happens because, for a fixed N_0, it is not possible to guarantee a perfect match in terms of covariates. As shown in Abadie and Imbens (2006) and Abadie and Imbens (2011), in a setting in which the number of treated and control observations grow (even if the number of control observations grows at a faster rate), nearest-neighbor matching estimators include a conditional bias term that converges to zero at a rate that may be slower than $N^{1/2}$. In our setting, however, since the variance of the estimator does not go to zero when $N_0 \to \infty$, this bias will always be of a lower order relative to the variance of the estimator. For this reason, we are also able to consider slightly less restrictive assumptions when we derive the asymptotic properties of the estimator in our setting.

4 Inference

The fact that the matching estimator is not asymptotically normal in our setting poses an important challenge when it comes to inference. In particular, the analytic asymptotic variance estimator derived in Abadie and Imbens (2006) should not provide a good approximation in our setting. Moreover, Abadie and Imbens (2008) show that bootstrap procedures may be invalid for matching estimators, even when the number of observations of both treatment and control groups is large. We show that it is possible to construct a test

[11] See Imbens and Wooldridge (2009) and Abadie et al. (2014) for a discussion on defining the estimand of interest as the treatment effect on the finite population at hand versus the from a super-population.

that is asymptotically valid when $N_0 \to \infty$ based on the theory of randomization tests under an approximate symmetry assumption developed in Canay et al. (2014).

Consider a function of the data given by:

$$S_{N_0} = (\hat{\tau}_1, ..., \hat{\tau}_{N_1})' \tag{5}$$

where $\hat{\tau}_i = Y_i - \frac{1}{M}\sum_{j \in \mathcal{J}_M(i)} Y_j$. Note that each $\hat{\tau}_i$ depends on the M nearest neighbors of observation i, so it implicitly depends on N_0.

We want to test a sharp null hypothesis that the treatment effect is zero for all individuals. That is, $Y_i(1) = Y_i(0)$ for all i. Following Canay et al. (2014), we consider a test statistic given by:

$$T_{Wald}(S_{N_0}) = \frac{|\hat{\tau}|}{\sqrt{\frac{1}{N_1 - 1}\sum_{i=1}^{N_1}(\hat{\tau}_i - \hat{\tau})^2}} \tag{6}$$

where note that $\hat{\tau} = \frac{1}{N_1}\sum_{i \in \mathcal{I}_1} \hat{\tau}_i$.

We consider group of transformations given by $\mathbf{G} = \{-1, 1\}^{N_1}$, where $gS_{N_0} = (g_1\hat{\tau}_1, ..., g_{N_1}\hat{\tau}_{N_1})'$. Let $K = |\mathbf{G}|$ and denote by:

$$T^{(1)}(S_{N_0}) \leq T^{(2)}(S_{N_0}) \leq ... \leq T^{(K)}(S_{N_0}) \tag{7}$$

the ordered values of $\{T(gS_{N_0}) : g \in \mathbf{G}\}$. Let $k = \lceil K(1-\alpha) \rceil$, where α is the significance level of the test, and define:

$$\begin{aligned} K^+(S_{N_0}) &= |\{1 \leq j \leq K : T^{(j)}(S_{N_0}) > T^{(k)}(S_{N_0})\}| \\ K^0(S_{N_0}) &= |\{1 \leq j \leq K : T^{(j)}(S_{N_0}) = T^{(k)}(S_{N_0})\}| \end{aligned} \tag{8}$$

The randomization test is given by:

$$\phi(S_{N_0}) = \begin{cases} 1 & \text{if } T(S_{N_1}) > T^{(k)}(S_{N_1}) \\ a(S_{N_0}) & \text{if } T(S_{N_1}) = T^{(k)}(S_{N_1}) \\ 0 & \text{if } T(S_{N_1}) < T^{(k)}(S_{N_1}) \end{cases} \tag{9}$$

where:

$$a(S_{N_0}) = \frac{K\alpha - K^+(S_{N_1})}{K^0(S_{N_1})}$$

In words, we calculate the test statistic $T(gS_{N_0})$ for all possible $gS_{N_0} = (g_1\hat{\tau}_1, ..., g_{N_1}\hat{\tau}_{N_1})'$, and then we compare the actual test statistic $T(S_{N_0})$ with the distribution $\{T(gS_{N_0}) : g \in \mathbf{G}\}$.

Proposition 2 *Under the same assumptions used in part 2 of Proposition 1, if we further assume that ϵ_i is symmetric around zero for all $i = 1, ..., N$, then testing a null hypothesis that $\mu_1(X_i) = \mu_0(X_i)$ for all $i \in \mathcal{I}_1$ using the decision rule defined in 9 satisfies, under the null, $\mathbb{E}[\phi(S_{N_1})] \to \alpha$ for any $\alpha \in (0, 1)$.*

Proof.

We apply Theorem 3.1 from Canay et al. (2014). We only need to show that, when $N_0 \to \infty$, the limiting distribution of S_{N_0} under the null is invariant to sign changes. This will be true if, asymptotically, $\hat{\tau}_i$ and $\hat{\tau}_j$ are independent for $i \neq j$, and the distribution of $\hat{\tau}_i$ is symmetric around zero. Note that it is not required that $\hat{\tau}_i$ has the same distribution across i.

From the results in Proposition 1, we know that, under the null, the asymptotic distribution of $\hat{\tau}_i$ conditional on $\{X\}_{i\in\mathcal{I}_1}$ is given by $\epsilon_i - \frac{1}{M}\sum_{m=1}^{M}\epsilon_m(X_i)$, which is symmetric around zero given the assumption that ϵ_i is symmetric around zero for all $i = 1, ..., N$. Moreover, we also know from Proposition 1 that $\hat{\tau}_i$ are independent across i. Therefore, the assumptions for Theorem 3.1 from Canay et al. (2014) are satisfied. ■

Remark 5 Note that we can test the null hypothesis that the average treatment effect is equal to zero conditional on each covariate value in $\{X_i\}_{i\in\mathcal{I}_1}$. This null hypothesis is implied by more narrowly defined null hypotheses that are usually considered in Fisher-type tests, such as $Y_i(0)|X_i \overset{d}{=} Y_i(1)|X_i$ or $Y_i(0) = Y_i(1)$ with probability one.

Remark 6 The assumption that ϵ_i is symmetric around zero for all $i = 1, ..., N$ is symmetric around zero could be replaced by ϵ_i is i.i.d. and $M = 1$, even if ϵ_i is not symmetric.

Remark 7 In the case $M = 1$ the randomization test we propose is equivalent to a permutation test conditional on the matched pair. In this case, $\hat{\tau}_i = Y_i - Y^i_{(1)}$ so a sign transformation $-\hat{\tau}_i = Y^i_{(1)} - Y_i$ is equivalent to permute the treatment assignment within each pair. Rosenbaum (2002) considers Fisher exact tests in observational studies with matched pairs. They show that, if the probability of treatment assignment is the same for both observations in each pair, then a permutation test conditional on the pair is valid even in finite samples. With a finite N_0, however, it is not possible to guarantee this condition

even under assumption 2, since we will not have a perfect match in terms of covariates. We show that this condition can be approximately satisfied when $N_0 \to \infty$.

Remark 8 The randomization induced by $a(S_{N_0})$ when $T(S_{N_1}) = T^{(k)}(S_{N_1})$ guarantees an asymptotic rejection rate of α despite the discreteness of the randomization distribution. As stated in Canay et al. (2014), a non-randomized test that rejects if $T(S_{N_1}) > T^{(k)}(S_{N_1})$ is level α and, unless N_1 is very small, this should not be a problem.

5 Conclusion

We consider the asymptotic properties of matching estimators when the number of control observations is large, but the number of treated observations is fixed. We show that the nearest neighbor matching estimator is asymptotically unbiased for the average treatment effect on the treated under standard assumptions used in the literature on estimation of treatment effects under selection on unobservables. Moreover, we provide a test based on the theory of randomization tests under approximate symmetry developed in Canay et al. (2014) that is asymptotically valid when the number of control observations goes to infinity. Our theoretical results should provide a better approximation to the behavior of the matching estimator and more reliable hypothesis testing relative to Abadie and Imbens (2006) in settings in which not only there is a much larger number of control observations relative to treated observations, but also that the number of treated observations are not large enough so that we can rely on asymptotic results.

Finally, our results can be extended to the SC estimator in settings in which the number of control units is large. Asymptotic unbiasedness is guaranteed under the assumption that treatment assignment is not correlated with potential outcomes in the post-period once we condition on the pre-treatment outcomes. Moreover, our inference method can be applied in applications with multiple treated units, provided that the number of treated units is not too small.

A Appendix

Proof of Proposition 1

For a given realization of $X_i = \bar{x}$ for an observation in the treated group and for a given $\epsilon > 0$, consider the probability that the M-closest realizations of $\{X_j\}_{j \in \mathcal{I}_0}$ are such that $d(X_j, \bar{x}) < \epsilon$. Let $X^i_{(M)}$ be the M-closest match of observation i. Then:

$$\begin{aligned} \Pr\left(d(X^i_{(M)}, \bar{x}) > \epsilon\right) &= \sum_{m=0}^{M-1} \Pr\left(d(X_j, \bar{x}) < \epsilon \text{ for exactly } m \text{ observations}\right) \\ &= \sum_{m=0}^{M-1} \begin{pmatrix} N_0 \\ m \end{pmatrix} [\Pr(d(X_j, \bar{x}) < \epsilon)]^m [\Pr(d(X_j, \bar{x}) > \epsilon)]^{N_0 - m} \end{aligned} \tag{10}$$

Since $\bar{x} \in \mathbb{X}_0$, we have that $Pr(d(X_j, \bar{x}) < \epsilon) > 0$, which implies that $\Pr(d(X_j, \bar{x}) > \epsilon) < 1$. Therefore, we have that $\Pr\left(d(X^i_{(M)}, \bar{x}) > \epsilon\right) \to 0$. By analogy, the m-nearest neighbor of i for $m < M$ will also converge in probability to $\bar{x}$.

Now consider:

$$\mathbb{E}[\hat{\tau} | \{X_i\}_{i \in \mathcal{I}_1}] = \frac{1}{N_1} \sum_{i \in \mathcal{I}_1} \left(\mu_1(X_i) - \mathbb{E}\left[\frac{1}{M} \sum_{m=1}^{M} \mu_0(X^i_{(m)}) \right] \right) \tag{11}$$

Since $\mu_0(x)$ is continuous and bounded and $X^i_{(m)} \xrightarrow{p} X_i$, then we have that $\mathbb{E}[\mu_0(X^i_{(m)}) | X_i] \to \mu_0(X_i)$, which proves of proposition 1.

For part 2, assume that $\tilde{f}(x) = \mathbb{E}[f(Y(0)) | X = x]$ is continuous and bounded for any $f : \mathbb{R} \to \mathbb{R}$ continuous and bounded. Let $Y^i_{(m)}$ be the outcome of the m-nearest neighbor of treated observation i. Therefore, for any $f(y)$ continuous and bounded, and for a given $X_i = \bar{x}$, we have that:

$$\mathbb{E}[f(Y^i_{(m)})] = E\left\{ \mathbb{E}[f(Y^i_{(m)}) | X^i_{(m)}] \right\} = E\left\{ \tilde{f}(X^i_{(m)}) \right\} \to \tilde{f}(\bar{x}) = E[f(Y(0)) | X = \bar{x}] \tag{12}$$

By the Portmanteau Lemma, we have that $Y^i_{(m)} \xrightarrow{d} Y(0) | \{X = \bar{x}\}$. Under assumption 2, $Y^i_{(m)} \xrightarrow{d} \mu_0(X_i) + e_m(X_i)$, where $e_m(X_i) \stackrel{d}{=} Y_i(0) | X_i - \mu_0(X_i)$. Therefore, conditional on $\{X_i\}_{i \in \mathcal{I}_1}$:

$$\hat{\tau} = \frac{1}{N_1} \sum_{i \in \mathcal{I}_1} \left[Y_i - \frac{1}{M} \sum_{m=1}^{M} Y^i_{(m)} \right] \xrightarrow{d} \frac{1}{N_1} \sum_{i \in \mathcal{I}_1} \left[(\mu_1(X_i) - \mu_0(X_i)) + \left(\epsilon_i - \frac{1}{M} \sum_{m=1}^{M} \epsilon_m(X_i) \right) \right] \tag{13}$$

Now we just have to show that $\epsilon_m(X_i)$ is independent across m and i. Since X_i is a continuous random variable, then $X_i \neq X_j$ with probability one for $i \neq j$ with $i, j \in \mathcal{I}_1$. Since there is a finite number of treated observations, then it must be that, conditional on $\{X_i\}_{i=1}^{N_1}$, there is an $\eta > 0$ such that $d(X_i, X_j) > \eta$ for all $i, j \in \mathcal{I}_1$ with $i \neq j$. However, we know that $Pr(d(X_i, X^i_{(m)}) > \epsilon) \to 0$ for all $\epsilon > 0$. Therefore, the probability that $k \in \mathcal{I}_0$ belongs to $\mathcal{J}_M(i)$ and $\mathcal{J}_M(j)$ converges to zero. Therefore, under the assumption that the errors ϵ_i are independent across i (which is guaranteed from assumption 1), we have that $\epsilon_m(X_i)$ is independent across m and i.

Unconditional Expectation

Now we consider the unconditional expectation of $\hat{\tau}$:

$$\mathbb{E}[\hat{\tau}] = \mathbb{E}\{\mathbb{E}[\hat{\tau}|\{X_i\}_{i\in\mathcal{I}_1}]\} = \frac{1}{N_1}\sum_{i\in\mathcal{I}_1}\mathbb{E}\left[\mu_1(X_i) - \frac{1}{M}\sum_{m=1}^{M}\mu_0(X^i_{(m)})\right] \tag{14}$$

We need that $\mathbb{E}[\mu_0(X^i_{(m)})] \to \mathbb{E}[\mu_0(X_i)]$. We know that $\mathbb{E}[\mu_0(X^i_{(m)})|X_i] \to \mu_0(X_i)$ for all X_i. Again using the fact that $\mu_0(x)$ is continuous and bounded, we have that $\mathbb{E}[\mu_0(X^i_{(m)})] = \mathbb{E}\{\mathbb{E}[\mu_0(X^i_{(m)})|X_i]\} \to \mathbb{E}[\mu_0(X_i)]$. Therefore:

$$\mathbb{E}[\hat{\tau}] \to \mathbb{E}\left[\mu_1(X_i) - \mu_0(X_i)\right] \tag{15}$$

where this expectation is taken according to $f_1(x)$, the density function of the treated units.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller, "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statiscal Association*, 2010, *105* (490), 493–505.

_ , _ , and _ , "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, 2015, *59* (2), 495–510.

_ and Guido W. Imbens, "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 2006, *74* (1), 235–267.

_ and _ , "On the Failure of the Bootstrap for Matching Estimators," *Econometrica*, 2008, *76* (6), 1537–1557.

_ and _ , "Bias-Corrected Matching Estimators for Average Treatment Effects," *Journal of Business & Economic Statistics*, 2011, *29* (1), 1–11.

_ and Javier Gardeazabal, "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, 2003, *93* (1), 113–132.

_ , Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge, "Finite Population Causal Standard Errors," Working Paper 20325, National Bureau of Economic Research July 2014.

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

Canay, Ivan A., Joseph P. Romano, and Azeem M. Shaikh, "Randomization Tests under an Approximate Symmetry Assumption?," 2014.

Daz, Juan, Toms Rau, and Jorge Rivera, "A Matching Estimator Based on a Bilevel Optimization Problem," *The Review of Economics and Statistics*, October 2015, *97* (4), 803–812.

Ferman, Bruno and Cristine Pinto, "Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity," MPRA Paper 67665, University Library of Munich, Germany November 2015.

_ and _ , "Revisiting the Synthetic Control Estimator," MPRA Paper 73982, University Library of Munich, Germany September 2016.

_ **and** _ , "Placebo Tests for Synthetic Controls," MPRA Paper 78079, University Library of Munich, Germany April 2017.

Firpo, Sergio and Vitor Possebom, "Synthetic Control Estimator: A Generalized Inference Procedure and Confidence Sets," April 2016.

Frolich, Markus, "Finite-Sample Properties of Propensity-Score Matching and Weighting Estimators," *The Review of Economics and Statistics*, 2004, *86* (1), 77–90.

Gobillon, Laurent and Thierry Magnac, "Regional Policy Evaluation: Interative Fixed Effects and Synthetic Controls," *Review of Economics and Statistics*, 2016. Forthcoming.

Hahn, Jinyong and Ruoyao Shi, "Synthetic Control and Inference," 2016.

Imbens, Guido, "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *Review of Economics and Statistics*, 2004.

_ , "Matching Methods in Practice: Three Examples," NBER Working Papers 19959, National Bureau of Economic Research, Inc March 2014.

Imbens, Guido W. and Jeffrey M. Wooldridge, "Recent Developments in the Econometrics of Program Evaluation," Technical Report 1 2009.

Rosenbaum, Paul R., "Conditional Permutation Tests and the Propensity Score in Observational Studies," *Journal of the American Statistical Association*, 1984, *79* (387), 565–574.

_ , "Covariance Adjustment in Randomized Experiments and Observational Studies," *Statist. Sci.*, 08 2002, *17* (3), 286–327.

Rubin, Donald B., "Matching to Remove Bias in Observational Studies," *Biometrics*, 1973, *29* (1), 159–183.