On the Role of Covariates in the Synthetic Control Method

Botosaru, Irene and Ferman, Bruno

Simon Fraser University, Sao Paulo School of Economics, FGV

14 August 2017

On the Role of Covariates in the Synthetic Control Method

Irene Botosaru[*] Bruno Ferman[†]

August 13, 2017

Please click here for the most recent version

Abstract

This note revisits the role of time-invariant observed covariates in the Synthetic Control (SC) method. We first derive conditions under which the original result of Abadie et al. (2010) regarding the bias of the SC estimator remains valid when we relax the assumption of a perfect match on observed covariates and assume only a perfect match on pre-treatment outcomes. We then show that, even when the conditions for the first result are valid, a perfect match on pre-treatment outcomes does not generally imply an approximate match for all covariates. This will only be true for those that are both relevant and whose effects (over time) are not collinear with the effects of other observed and unobserved covariates. Taken together, our results show that a perfect match on covariates should not be required for the SC method, as long as there is a perfect match on a long set of pre-treatment outcomes.

JEL codes: C13, C21, C23.
Keywords: Synthetic controls, covariates, perfect match.

[*]E-mail: ibotosar@sfu.ca. Address: Simon Fraser University, 8888 University Drive, Department of Economics, Burnaby, B.C. V5A 1S6, Canada.

[†]E-mail: bruno.ferman@fgv.br. Address: Sao Paulo School of Economics, FGV, Rua Itapeva no. 474, Sao Paulo - Brazil, 01332-000.

1 Introduction

Social scientists are often interested in evaluating the effect of a policy or a treatment on an outcome of interest. To perform such an analysis, both a control and a treatment group are needed. In the absence of randomized experiments, however, it is often difficult to find a suitable control group. The synthetic control (SC) method, developed in a series of papers by Abadie and Gardeazabal (2003), Abadie et al. (2010), and Abadie et al. (2015), allows practitioners to construct a control group from a set of potential control groups. The method uses a data-driven weighted average of the selected groups to construct a synthetic control group that is *"more similar to the treatment group than any of the individual control groups,"* c.f. Athey and Imbens (2017). This fact has contributed to the method's success, earning it its distinction of being *"arguably one of the most important innovations in the policy evaluation literature in the last 15 years,"* c.f. Athey and Imbens (2017).

A crucial step in applications of the SC method is the choice of predictors used to estimate the weights. Although there is little guidance on which variables to be used as predictors, see e.g. Ferman et al. (2016), the original article on the SC method mentions using pre-treatment outcomes and other time-invariant observed covariates. In fact, when potential outcomes follow a linear factor model, Abadie et al. (2010) show that the existence of weights that achieve a perfect match on *both* pre-treatment outcomes and time-invariant covariates implies that the bias of the SC estimator is bounded by a function that approaches zero as the number of pre-treatment periods increases.

In this note, we revisit this result. First, we consider whether the results of Abadie et al. (2010) remain valid without imposing a perfect match on time-invariant covariates.[1] We show that this will be the case if we impose additional assumptions on the effects of observed covariates on potential outcomes relative to the assumptions in Abadie et al. (2010). Second, we derive conditions under which a perfect match on a long set of pre-treatment outcomes also implies an approximate match on time-invariant covariates.[2] We show that this is true for covariates that are both relevant *and* whose effects on the potential outcomes are linearly independent from the effects of other observable and unobservable covariates. Therefore, a perfect match on pre-treatment outcomes does not necessarily imply a perfect match on all covariates that are relevant in determining the potential outcomes. However, this does not invalidate the result that the bias of the SC estimator is bounded by a function that goes to zero when the number of pre-treatment periods increases.

[1] In parallel with our work, Kaul et al. (2015) claim (without formalizing it) that ignoring covariates is not expected to lead to asymptotic bias when the number of pre-treatment periods goes to infinity.

[2] We refer to "approximate match" for a covariate as the difference between the covariate for the treated unit and for the SC unit being bounded by a function that goes to zero with the number of pre-treatment periods.

Taken together, our results show that a perfect match on observed time-invariant covariates should not be required for the SC method, as long as there is a perfect match on a long set of pre-treatment outcomes.

The remainder of paper is organized as follows. In Section 2 we set-up the model and briefly review the results of Abadie et al. (2010), while in Section 3 we present the new results. All proofs are contained in the Appendix.

2 The model in Abadie et al. (2010)

Let $Y_{it}(1)$ and $Y_{it}(0)$ be potential outcomes in the presence and in the absence of a treatment, respectively, for unit i at time t. Consider the model:

$$\begin{cases} Y_{it}(0) = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \varepsilon_{it} \\ Y_{it}(1) = \alpha_{it} + Y_{it}(0) \end{cases} \tag{1}$$

where δ_t is an unknown common factor with constant factor loadings across units; λ_t is a $(1 \times F)$ vector of common factors; μ_i is a $(F \times 1)$ vector of unknown factor loadings; θ_t is a $(1 \times r)$ vector of unknown parameter; Z_i is a $(r \times 1)$ vector of observed covariates (not affected by the intervention), and the error terms ε_{it} are unobserved transitory shocks. As in Abadie et al. (2010), we treat θ_t and λ_t as parameters and μ_i as random. We say that a covariate Z_{ki} is relevant if its associated coefficient $\theta_{kt} \neq 0$ for some t, and we refer to μ_i as an unobserved covariate. The observed outcomes are given by

$$Y_{it} = D_{it} Y_{it}(1) + (1 - D_{it}) Y_{it}(0), \tag{2}$$

where $D_{it} = 1$ if unit i is treated at time t.

Suppose that the treatment takes place after time $t = T_0$ and let the index 1 denote the treated unit. We observe the outcomes of the treated unit and of J control units for T_0 pre-intervention periods and for $T - T_0$ post-intervention periods, that is, $(Y_{1,t}, ..., Y_{J+1,t})$ for $t = 1, ..., T_0, T_0 + 1, ..., T$.

The main goal of the SC method is to estimate the treatment effect on the treated, i.e.

$$\alpha_{1t} = Y_{1t}(1) - Y_{1t}(0), \ t > T_0. \tag{3}$$

Since $Y_{1t}(0)$ for $t > T_0$, is not observed, the main idea of the SC method is to consider a weighted average of the control units to construct a proxy for this counterfactual. That is, for a given set of weights

$$\mathbf{w} \in \{(w_2, ..., w_{J+1}) | \sum_{j=2}^{J+1} w_j = 1 \text{ and } w_j \geq 0\} \tag{4}$$

the SC estimator for $t > T_0$ is given by:

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{j \neq 1} w_j Y_{jt} = \alpha_{1t} + \theta_t \left(Z_1 - \sum_{j \neq 1} w_j Z_j \right) + \lambda_t \left(\mu_1 - \sum_{j \neq 1} w_j \mu_j \right) + \left(\varepsilon_{1t} - \sum_{j \neq 1} w_j \varepsilon_{jt} \right).$$

Abadie et al. (2010) provide conditions under which the bias of the SC estimator is bounded by a function that goes to zero as the number of pre-intervention periods grows. The authors assume the existence of weights $\mathbf{w}^* \in \mathbb{R}^J$ that satisfy (4) and such that

$$Y_{1t} = \sum_{j \neq 1} w_j^* Y_{jt}, \ t \leq T_0, \tag{5}$$

$$Z_1 = \sum_{j \neq 1} w_j^* Z_j, \tag{6}$$

where (5) is the assumption of a perfect match on pre-treatment outcomes, and (6) is the assumption of perfect match on time-invariant observed covariates. Given (5) and (6), and other additional assumptions, Abadie et al. (2010) derive bounds on the bias of the SC estimator that go to zero when T_0 increases. More precisely, letting

$$\hat{\alpha}_{1t}^* = Y_{1t} - \sum_{j \neq 1} w_j^* Y_{jt}, \tag{7}$$

they show that there exists a function $b(T_0)$ such that for all $t > T_0$:

$$|\mathbb{E}(\hat{\alpha}_{1t}^*) - \alpha_{1t}| \leq b(T_0) \ \text{ with } b(T_0) \to 0 \text{ as } T_0 \to \infty. \tag{8}$$

3 The role of covariates in the Abadie et al. (2010) method

We derive conditions under which result (8) remains valid when (6) is not assumed. We also derive conditions under which assuming (5) implies that (6) holds approximately. The main idea of our proof is to treat observed covariates (Z_i) as factor loadings and their associated time-varying effects (θ_t) as common factors.

Define the $1 \times (r+F)$ row vector $\gamma_t \equiv (\theta_t, \lambda_t)$, and denote by $\xi(T_0)$ the smallest eigenvalue of $\frac{1}{T_0} \sum_{t=1}^{T_0} \gamma_t' \gamma_t$.

Consider the following assumptions, which are similar to those in Abadie et al. (2010).

Assumption 1

(i) ε_{it} *are inid with* $\mathbb{E}\left(\varepsilon_{it}\right)=0$ *and* $\sigma^2_{it}=\mathbb{E}\left(\varepsilon^2_{it}\right)<\infty$ *for all* i *and* t*;*

(ii) $\mathbb{E}[\varepsilon_{it}|Z_i,\mu_i]=0$*;*

(iii) $\exists\underline{\xi}>0$ *and* $\bar{T}\in\mathbb{N}$ *such that* $\xi(T_0)>\underline{\xi}$ *for all* $T_0>\bar{T}$*;*

(iv) $|\gamma_{tm}|\leq\overline{\gamma}$ *for all* $t=1,...,T$ *and* $m=1,...,r+F$*;*

(v) $\mathbb{E}\left(|\varepsilon_{it}|^p\right)<\infty$ *for* $p=2m$ *where* $1\leq m\in\mathbb{N}$*, and for all* $t=1,...,T_0$ *and* $i=2,...,J+1$*;*

Remark 1 Assumption 1.(iii) excludes the possibility of covariates that are irrelevant in determining the potential outcome (that is, $\theta_{kt}=0$ for all t). This assumption also excludes the possibility of covariates whose effects are multicollinear with the effects of other observed or unobserved covariates. If we were considering a setting with only unobserved covariates, then we would always be able to redefine the unobserved covariates so that we have an observationally equivalent model with no covariates that are irrelevant or whose effects are multicollinear with the effects of other covariates.[3] However, this will not be the case if we have observed covariates. We show later that it is possible to relax this assumption and still provide bounds on the bias of the SC estimator.

Proposition 1 Consider the model (1) and (2). Let there be weights $\mathbf{w}^*\in\mathbb{R}^J$ such that (4) and (5) hold, and let Assumption 1 hold. Then there exists a function $b_\alpha(T_0)$ with $lim_{T_0\rightarrow\infty}b_\alpha(T_0)=0$ such that:

$$|\mathbb{E}\left(\hat{\alpha}^*_{1t}\right)-\alpha_{1t}|\leq b_\alpha(T_0)\text{ for all }t>T_0. \tag{9}$$

Additionally, there exist functions $b_{\mu,l}\left(T_0\right)$, $l=1,...,F$, and $b_{Z,k}(T_0)$, $k=1,...,r$, with $lim_{T_0\rightarrow\infty}b_{\mu,l}(T_0)=0=\lim_{T_0\rightarrow\infty}b_{Z,k}(T_0)$ such that:

$$\left|\mathbb{E}\left(Z_{k1}-\sum_{j=2}^{J+1}w^*_jZ_{kj}\right)\right| \leq b_{Z,k}(T_0)\text{ for all }k=1,...,r, \tag{10}$$

$$\left|\mathbb{E}\left(\mu_{l1}-\sum_{j=2}^{J+1}w^*_j\mu_{lj}\right)\right| \leq b_{\mu,l}(T_0)\text{ for all }l=1,...,F. \tag{11}$$

[3]Note, however, that assumption 1.(iii) may still fail in this case. For example, we may have a simple example in which $\lambda_{1,t}=1$ for $t=1$ and $\lambda_{1,t}=0$ for $t>1$. In this case, $\lambda_{1,t}$ is relevant, but $\xi(T_0)\rightarrow 0$ when $T_0\rightarrow\infty$.

Proof. We provide the proof of Proposition 1 in the Appendix.

Proposition 1 provides conditions under which perfect matching on pre-treatment outcomes implies that the bias of the SC estimator converges to zero with the number of pre-treatment periods, e.g. result (9). Note that assumptions 1.(iii) and 1.(iv) refer to both the effects of observed and unobserved covariates (θ_t and λ_t), while the equivalent result in Abadie et al. (2010) only requires conditions on the effects of unobserved covariates (λ_t). Therefore, while we relax the assumption of perfect match on covariates, we require additional assumptions on the effects of unobserved covariates relative to Abadie et al. (2010).[4]

The proposition also provides conditions under which a perfect match on pre-treatment outcomes is sufficient for an approximate match on observed covariates, e.g. result (10), and an approximate match on unobserved covariates, e.g. result (11). This will be the case if observed and unobserved covariates are relevant and their effects on the potential outcomes are not linearly dependent. The intuition behind this result is that it would not be possible to match on a large number of pre-treatment outcomes without matching on both observed and unobserved relevant covariates.

We now relax assumption 1.(iii). We allow for covariates that are irrelevant or whose effects are multicolinear with the effects of other observed and unobserved covariates. That is, we allow for $\gamma_t \mathbf{b} = 0$ for all t for some $\mathbf{b} \in \mathbb{R}^{r+F} \backslash \{0\}$.[5] Without loss of generality, suppose that the first $\tilde{r}$ covariates are relevant and have effects that are not multicolinear ($0 \leq \tilde{r} \leq r$), and let $\tilde{\theta}_t$ be a $1 \times \tilde{r}$ vector with the first $\tilde{r}$ components of θ_t and $\tilde{Z}_i$ be a $\tilde{r} \times 1$ vector with the first $\tilde{r}$ components of Z_i. Also, let $\tilde{a}$ be the dimension of the complement of the space $\{\mathbf{b} \in \mathbb{R}^{r+F} \backslash \{0\} | \gamma_t \mathbf{b} = 0\}$. Then we can always find a $1 \times \tilde{a}$ vector $\tilde{\gamma}_t$ with first $\tilde{r}$ components equal to $\tilde{\theta}_t$ such that, for any $b \in \mathbb{R}^{r+F}$, there will be a $\tilde{b} \in \mathbb{R}^{\tilde{a}}$ such that $\gamma_t b = \tilde{\gamma}_t \tilde{b}$ for all t. Moreover, the first $\tilde{r}$ components of b will be the same as the first $\tilde{r}$ components of $\tilde{b}$. Therefore, we can find a $\tilde{a} \times 1$ vector $\tilde{X}_i$ with first $\tilde{r}$ components equal to $\tilde{Z}_i$, such that model 1 can be rewritten as $Y_{it}(0) = \delta_t + \tilde{\gamma}_t \tilde{X}_i + \varepsilon_{it}$.

Therefore, if we assume that $\frac{1}{T_0}\sum_{t=1}^{T_0} \tilde{\gamma}_t' \tilde{\gamma}_t$ satisfies the conditions from assumption 1.(iii), then we can apply Proposition 1. In this case, the bias of the SC estimator is bounded, and the first $\tilde{r}$ covariates are approximately matched. However, in this case, it is not be possible to guarantee an approximate match for all covariates if $r > \tilde{r}$. There are two reasons for this. First, some covariates may be irrelevant in determining the potential outcomes. In this case, it is clear that a perfect match on pre-treatment outcomes may be achieved even in the presence of a mismatch in such covariates. More interestingly, there may be a mismatch even for covariates that are relevant. For example, imagine that there is a time-invariant common factor

[4]See Ferman and Pinto (2016) for the implications for the SC estimator when the effects of covariates are allowed to increase without bounds, so that assumption 1.(iv) is violated.

[5]For example, this allows for irrelevant covariates or for two or more covariates with time-invariant effects.

$\lambda_{1t} = 1$ with associated factor loading μ_{1i}, and a covariate Z_{1i} with time-invariant effects $\theta_{1t} = \theta_1$. In this case, we would guarantee an approximate match for $(\mu_{1i} + Z_{1i}\theta_1)$, but we would not be able to guarantee an approximate match for μ_{1i} and for Z_{1i} separately. Intuitively, this multicollinearity implies that there would be weighted averages of the control units that may provide a perfect match for the treated unit even if there is a mismatch in these covariates. Importantly, these results suggest that a mismatch in observed covariates does not necessarily imply an (asymptotically) biased SC estimator, even if such covariates are relevant in determining potential outcomes.

4 Conclusion

We revisit the role of time-invariant covariates in the SC method. We formally derive two results. First, we provide conditions under which the result in Abadie et al. (2010) regarding the bias of the SC estimator remains valid when we relax the assumption of perfect match on covariates and assume only a perfect match on pre-treatment outcomes. Second, we provide conditions under which a perfect match on pre-treatment outcomes also provide an approximate match for the covariates. We show that an approximate match for covariates may not be achieved even under conditions in which the bias of the SC estimator is bounded. This may be the case even for relevant covariates. Taken together, our results show that a perfect match on covariates should not be required for the SC method, as long as there is a perfect match on a long set of pre-treatment outcomes.

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

Athey, Susan and Guido W. Imbens, "The State of Applied Econometrics: Causality and Policy Evaluation," *Journal of Economic Perspectives*, May 2017, *31* (2), 3–32.

Ferman, Bruno and Cristine Pinto, "Revisiting the synthetic control estimator," 2016.

_ , _ , and Vitor Possebom, "Cherry Picking with Synthetic Controls," 2016. Working Paper.

Kaul, Ashok, Stefan Klobner, Gregor Pfeifer, and Manuel Schieler, "Synthetic Control Methods: Never Use All Pre-Intervention Outcomes as Economic Predictores," 2015.

A Appendix

Proof of Proposition 1

The proof follows closely Abadie et al. (2010). We first prove result (9) of Proposition 1. First, notice that

$$Y_{1t}(0) - \sum_{i=2}^{J+1} w_i Y_{it}(0) = \gamma_t \left(X_1 - \sum_{i=1}^{J+1} w_i X_i \right) + \sum_{i=2}^{J+1} w_i \left(\varepsilon_{1t} - \varepsilon_{it} \right), \tag{12}$$

where $X_i = (Z_i, \mu_i)'$ is a $(r+F) \times 1$ vector.

Stacking pre-treatment variables, i.e. $Y_i^P \equiv (Y_{i1}, ..., Y_{iT_0})'$, we have that:

$$Y_1^P - \sum_{i=2}^{J+1} w_i Y_i^P \quad = \quad \Gamma^P \left(X_1 - \sum_{i=1}^{J+1} w_i X_i \right) + \sum_{i=2}^{J+1} w_i \left(\varepsilon_1^P - \varepsilon_i^P \right) \tag{13}$$

where Y_i^P and ε_i^P are $T_0 \times 1$ vectors, and $\Gamma^P = \left[\gamma_1', ..., \gamma_{T_0}' \right]'$ is a $T_0 \times (r+F)$ matrix.

We solve (13) for $\left(X_1 - \sum_{i=1}^{J+1} w_i X_i \right)$ to obtain

$$\left(X_1 - \sum_{i=1}^{J+1} w_i X_i \right) = \left(\Gamma^{P\prime} \Gamma^P \right)^{-1} \Gamma^{P\prime} \left(Y_1^P - \sum_{i=2}^{J+1} w_i Y_i^P \right) - \left(\Gamma^{P\prime} \Gamma^P \right)^{-1} \Gamma^{P\prime} \sum_{i=2}^{J+1} w_i \left(\varepsilon_1^P - \varepsilon_i^P \right) \tag{14}$$

where we used assumption 1(iii), which guarantees that $\left(\Gamma^{P\prime} \Gamma^P \right)^{-1}$ exists if T_0 is large enough. Plugging this into (12) obtains

$$\begin{aligned} Y_{1t}(0) - \sum_{i=2}^{J+1} w_i Y_{it}(0) &= \gamma_t \left(\Gamma^{P\prime}\Gamma^P\right)^{-1} \Gamma^{P\prime} \left(Y_1^P - \sum_{i=2}^{J+1} w_i Y_i^P\right) \\ &\quad - \gamma_t \left(\Gamma^{P\prime}\Gamma^P\right)^{-1} \Gamma^{P\prime} \sum_{i=2}^{J+1} w_i \left(\varepsilon_1^P - \varepsilon_i^P\right) \\ &\quad + \sum_{i=2}^{J+1} w_i \left(\varepsilon_{1t} - \varepsilon_{it}\right). \end{aligned}$$

Using (4) and (5) obtains:

$$\begin{aligned} & Y_{1t}(0) - \sum_{i=2}^{J+1} w_i^* Y_{it}(0) && (15) \\ =\; & \gamma_t \left(\Gamma^{P\prime}\Gamma^P\right)^{-1} \Gamma^{P\prime} \sum_{i=2}^{J+1} w_i^* \varepsilon_i^P && (16) \\ & - \gamma_t \left(\Gamma^{P\prime}\Gamma^P\right)^{-1} \Gamma^{P\prime} \varepsilon_1^P && (17) \\ & + \sum_{i=2}^{J+1} w_i^* \left(\varepsilon_{1t} - \varepsilon_{it}\right). && (18) \end{aligned}$$

Noting that the $(r+F) \times (r+F)$ matrix $\Gamma^{P\prime}\Gamma^P = \sum_{j=1}^{T_0} \gamma_j' \gamma_j$, we write the right hand side of (16) as:

$$\begin{aligned} \gamma_t \left(\Gamma^{P\prime}\Gamma^P\right)^{-1} \Gamma^{P\prime} \sum_{i=2}^{J+1} w_i^* \varepsilon_i^P &= \sum_{i=2}^{J+1} w_i^* \gamma_t \left(\sum_{j=1}^{T_0} \gamma_j' \gamma_j\right)^{-1} \sum_{s=1}^{T_0} \gamma_s' \varepsilon_{is} \\ &= \sum_{i=2}^{J+1} w_i^* \sum_{s=1}^{T_0} \psi_{ts} \varepsilon_{is} \qquad (19) \end{aligned}$$

where

$$\psi_{ts} \equiv \gamma_t \left(\sum_{j=1}^{T_0} \gamma_j' \gamma_j\right)^{-1} \gamma_s'$$

Taking expectations on both sides of (15) and using expression (19) obtains for $t > T_0$:

$$\mathbb{E}\left(Y_{1t}(0) - \sum_{i=2}^{J+1} w_i^* Y_{it}(0)\right) = \mathbb{E}\left(\sum_{i=2}^{J+1} w_i^* \sum_{s=1}^{T_0} \psi_{ts} \varepsilon_{is}\right) \qquad (20)$$

where (17) and (18) equal to zero by assumption and since w_i^* is independent of ε_{it} for $t > T_0$.

We show below that there exists a positive function $b_\alpha(T_0)$ such that

$$\left|\mathbb{E}\left(\sum_{i=2}^{J+1} w_i^* \sum_{s=1}^{T_0} \psi_{ts}\varepsilon_{is}\right)\right| \leq b_\alpha(T_0) \text{ with } \lim_{T_0\to\infty} b_\alpha(T_0) = 0.$$

First, consider the following string of inequalities:

$$\psi_{ts}^2 \leq \psi_{tt}\psi_{ss} \leq \left(\frac{(r+F)\,\overline{\gamma}^2}{T_0\xi}\right)^2$$

where the first inequality follows by the Cauchy Schwarz inequality and by the fact that $\sum_{j=1}^{T_0}\gamma_j'\gamma_j$ is positive definite and symmetric, while the second inequality follows since $\left(\frac{1}{T_0}\sum_{j=1}^{T_0}\gamma_j'\gamma_j\right)^{-1}$ is symmetric positive definite with its largest eigenvalue given by ξ^{-1}. Then

$$\psi_{tt} \leq \frac{\gamma_t\gamma_t'}{T_0\xi} = \frac{\sum_{m=1}^{r+F}\gamma_{tm}^2}{T_0\xi} \leq \frac{(r+F)\,\overline{\gamma}^2}{T_0\xi}$$

and, similarly,

$$\psi_{ss} \leq \frac{(r+F)\,\overline{\gamma}^2}{T_0\xi}.$$

Define

$$\overline{\varepsilon}_{it} \equiv \sum_{s=1}^{T_0}\psi_{ts}\varepsilon_{is},\ i = 2, ..., J+1$$

and consider

$$\begin{aligned}\left|\sum_{i=2}^{J+1} w_i^*\overline{\varepsilon}_{it}\right| &\leq \sum_{i=2}^{J+1} w_i^*\left|\overline{\varepsilon}_{it}\right| \\ &\leq \left(\sum_{i=2}^{J+1}(w_i^*)^q\right)^{1/q}\left(\sum_{i=2}^{J+1}\left|\overline{\varepsilon}_{it}\right|^p\right)^{1/p} \quad (21)\\ &\leq \left(\sum_{i=2}^{J+1}\left|\overline{\varepsilon}_{it}\right|^p\right)^{1/p} \quad (22)\end{aligned}$$

where (21) follows by Holder's inequality with $p, q > 1$ and $\frac{1}{p}+\frac{1}{q} = 1$ and (22) follows by norm monotonicity and (4). Hence, applying Holder's again obtains:

$$\mathbb{E}\left(\sum_{i=2}^{J+1} w_i^*\left|\overline{\varepsilon}_{it}\right|\right) \leq \left[\mathbb{E}\left(\sum_{i=2}^{J+1}\left|\overline{\varepsilon}_{it}\right|^p\right)\right]^{1/p}$$

Applying Rosenthal's inequality to

$$\mathbb{E}\left(|\bar{\varepsilon}_{it}|^p\right)=\mathbb{E}\left(\left|\sum_{s=1}^{T_0}\psi_{ts}\varepsilon_{is}\right|^p\right)$$

obtains

$$\begin{aligned}\mathbb{E}\left(\left|\sum_{s=1}^{T_0}\psi_{ts}\varepsilon_{is}\right|^p\right) &\leq C(p)\max\left\{\left(\frac{(r+F)\,\overline{\gamma}^2}{T_0\xi}\right)^p\sum_{s=1}^{T_0}\mathbb{E}\,|\varepsilon_{is}|^p,\left(\frac{(r+F)\,\overline{\gamma}^2}{T_0\xi}\right)^p\left(\sum_{s=1}^{T_0}\mathbb{E}\,(\varepsilon_{is})^2\right)^{p/2}\right\}\\ &= C(p)\left(\frac{(r+F)\,\overline{\gamma}^2}{\xi}\right)^p\max\left\{\frac{1}{T_0^p}\sum_{s=1}^{T_0}\mathbb{E}\,|\varepsilon_{is}|^p,\left(\frac{1}{T_0^2}\sum_{s=1}^{T_0}\mathbb{E}\,(\varepsilon_{is})^2\right)^{p/2}\right\}\end{aligned}$$

which follows since

$$\begin{aligned}\sum_{s=1}^{T_0}\mathbb{E}\,|\psi_{ts}\varepsilon_{is}|^p &= \sum_{s=1}^{T_0}|\psi_{ts}|^p\,\mathbb{E}\left(|\varepsilon_{is}|^p\right)\leq\left(\frac{(r+F)\,\overline{\gamma}^2}{T_0\xi}\right)^p\sum_{s=1}^{T_0}\mathbb{E}\,|\varepsilon_{is}|^p\\ \sum_{s=1}^{T_0}\mathbb{E}\,|\psi_{ts}\varepsilon_{is}|^2 &= \sum_{s=1}^{T_0}(\psi_{ts})^2\,\mathbb{E}\,(\varepsilon_{is})^2\leq\left(\frac{(r+F)\,\overline{\gamma}^2}{T_0\xi}\right)^2\sum_{s=1}^{T_0}\mathbb{E}\,(\varepsilon_{is})^2\end{aligned}$$

Finally,

$$\begin{aligned}\left|\mathbb{E}\left(\widehat{\alpha}_{1t}^*-\alpha_{1t}\right)\right| &\leq \mathbb{E}\left(\sum_{i=2}^{J+1}w_i^*\,|\bar{\varepsilon}_{it}|\right)\\ &\leq \left[\mathbb{E}\left(\sum_{i=2}^{J+1}|\bar{\varepsilon}_{it}|^p\right)\right]^{1/p}\\ &\leq \left[\sum_{i=2}^{J+1}\mathbb{E}\left(\left|\sum_{s=1}^{T_0}\psi_{ts}\varepsilon_{is}\right|^p\right)\right]^{1/p}\\ &\leq C^{1/p}(p)\left(\frac{(r+F)\,\overline{\gamma}^2}{\xi}\right)\left[\sum_{i=2}^{J+1}\max\left\{\frac{1}{T_0^p}\sum_{s=1}^{T_0}\mathbb{E}\,|\varepsilon_{is}|^p,\left(\frac{1}{T_0^2}\sum_{s=1}^{T_0}\mathbb{E}\,(\varepsilon_{is})^2\right)^{p/2}\right\}\right]^{1/p}\\ &\leq (J\times C(p))^{1/p}\left(\frac{(r+F)\,\overline{\gamma}^2}{\xi}\right)\max\left\{\frac{\overline{m}_p^{1/p}}{T_0^{1-1/p}},\frac{\overline{\sigma}}{T_0^{1/2}}\right\}\\ &\leq (J\times C(p))^{1/p}\left(\frac{(r+F)\,\overline{\gamma}^2}{\underline{\xi}}\right)\max\left\{\frac{\overline{m}_p^{1/p}}{T_0^{1-1/p}},\frac{\overline{\sigma}}{T_0^{1/2}}\right\}\equiv b_\alpha(T_0)\end{aligned}$$

where, using the same notation as in Abadie et al. (2010), we let:

$$\begin{aligned}
\sigma_{is}^2 &= \mathbb{E}\left(\varepsilon_{is}\right)^2, \; \sigma_i^2 = \frac{1}{T_0}\sum_{s=1}^{T_0}\sigma_{is}^2, \; \overline{\sigma} = \left(\max_{i=2,\ldots,J+1}\sigma_i^2\right)^{1/2} \\
m_{p,is} &= \mathbb{E}\left|\varepsilon_{is}\right|^p, \; m_{p,i} = \frac{1}{T_0}\sum_{s=1}^{T_0} m_{p,is}, \; \overline{m}_p = \max_{i=2,\ldots,J+1} m_{p,i} < \infty
\end{aligned}$$

The proof for results (10) and (11) follow by similar arguments. First, define the $1 \times (r+F)$ vector $\rho_k \equiv [0, 0, ..., 1, ..., 0]$ where only the k^{th} element equals to 1. Consider k such that $1 \leq k \leq r$. From equation (14), we have that:

$$\begin{aligned}
\left(Z_{k,1} - \sum_{i=1}^{J+1} w_i Z_{k,i}\right) &= \rho_k\left(X_1 - \sum_{i=1}^{J+1} w_i X_i\right) \\
&= \rho_k\left(\Gamma^{P\prime}\Gamma^P\right)^{-1}\Gamma^{P\prime}\left(Y_1^P - \sum_{i=2}^{J+1} w_i Y_i^P\right) \\
&\quad -\rho_k\left(\Gamma^{P\prime}\Gamma^P\right)^{-1}\Gamma^{P\prime}\sum_{i=2}^{J+1} w_i\left(\varepsilon_1^P - \varepsilon_i^P\right)
\end{aligned}$$

If we define $\bar{\gamma}' = \max\{\bar{\gamma}, 1\} > 0$, then the proof of result (10) follows exactly the same steps as the proof of result (9) if we use $\bar{\gamma}'$ instead of $\bar{\gamma}$, so we have that $\left|\mathbb{E}\left(Z_{k1} - \sum_{i=2}^{J+1} w_i^* Z_{ki}\right)\right| \leq b_{Z,k}\left(T_0\right)$ with $b_{Z,k}\left(T_0\right) \to \infty$. Similarly, if we consider $l > r$ we have that $\left|\mathbb{E}\left(\mu_{l1} - \sum_{i=2}^{J+1} w_i^* \mu_{li}\right)\right| \leq b_{\mu,l}\left(T_0\right)$ with $b_{\mu,l}\left(T_0\right) \to \infty$.