# Appendix: Bias Bound of Synthetic Control Estimator

Zijian Wang

8/2024

## 1 Common Bounding Target

We can denote our panel dataset acquired with matrix $A = \boldsymbol{\Lambda\mu}$ for $\boldsymbol{\Lambda}$ being $(p \times F)$ and $\boldsymbol{\mu}$ being $(F \times n)$ where $F$ is the number of factors, $p$ is number of time periods in both pre and post treatment, $n$ is the number of observed units, which gives us matrix $A$ as $(p \times n)$. As Abadie et al. (2010) proposed originally, they have proposed a bound on the bias of the post-treatment period when Synthetic Control Estimator is applied. Throughout their justification and following their workflow, the key task we identified is to bound the following expression:

$$\sigma^2 = \boldsymbol{\Lambda}_{\cdot t}(\boldsymbol{\Lambda}_:'\boldsymbol{\Lambda}_:)^{-1}\boldsymbol{\Lambda}_{\cdot t}', \tag{1}$$

where $\sigma^2 = Var(\tilde{\boldsymbol{\varepsilon}}_j)$, for $\tilde{\boldsymbol{\varepsilon}}_j = \boldsymbol{\Lambda}_{\cdot t}(\boldsymbol{\Lambda}_:'\boldsymbol{\Lambda}_:)^{-1}\boldsymbol{\Lambda}_:'\boldsymbol{\varepsilon}_{:j}$, $\boldsymbol{\Lambda}_:$ being $(p-1 \times F)$ matrix with $t$th row being the common factor $\boldsymbol{\Lambda}_{\cdot t}$ $(1 \times F)$, $\boldsymbol{\varepsilon}_{:j}$ is a $(p-1 \times 1)$ vector with $t$th item being unobserved transitory shocks at the region level with zero mean $\varepsilon_{it}$, $p-1$ is number of pre-treatment periods[1]. Under their workflow but with different but justifiable steps, we are able to bound $\sigma \leq \sqrt{\frac{F\bar{\Lambda}^2}{(p-1)\eta}}$, and the ultimate bound followed could be derived into:

$$\begin{aligned} Bias_{post} &\leq \tilde{c_p}\left\|\hat{\boldsymbol{\theta}}\right\|_q \sigma \\ &\leq \tilde{c_p}\left\|\hat{\boldsymbol{\theta}}\right\|_q \sqrt{\frac{F\bar{\Lambda}^2}{(p-1)\eta}} \end{aligned} \tag{2}$$

where $\tilde{c_p}$ is some arbitrary constant, $\hat{\boldsymbol{\theta}}$ be the vector of weights $\hat{\theta}_j$ for $j = 1,...,n$ being number of units and $\left\|\hat{\boldsymbol{\theta}}\right\|_q$ is its q norm, $\bar{\Lambda}$ is the a bound imposed on the common factors such that $|\boldsymbol{\Lambda}_{tf}| \leq \bar{\Lambda}$, and $\eta$ is the smallest eigenvalue of $\boldsymbol{\Lambda}_:'\boldsymbol{\Lambda}_:$. The process of derivation would be elaborate in Appendix 1. The common question all of us would like to dig in is that how to bound such bias introduced when we apply the weights obtained from pre-treatment to post-treatment in synthetic control. But we would like to achieve it from a different angle and representation. Specifically, we define:

$$A = \begin{bmatrix} A_: \\ \boldsymbol{a}_e \end{bmatrix},$$

[1] We wrote things in pre-treatment period with a subscript of ":", while in Abadie's paper, they are denoted as superscript $P$. Specifically, the authors have written $\boldsymbol{\Lambda}_:$ as $\boldsymbol{\lambda}^P$ and $\boldsymbol{\Lambda}_{\cdot t}$ as $\boldsymbol{\lambda}_t$.

where $A_{:}$ is the pre-treatment stage with $p-1$ number of periods $(p-1\times n)$, and $\boldsymbol{a}_e$ is a row of additional outcomes we obtained during the post-treatment period $(1\times n)$, and the last column of both matrices denotes the synthetically controlled units. Then, suppose the weights we obtained are still denoted as $\hat{\boldsymbol{\theta}}$, and let $\boldsymbol{x}$ denotes $\begin{bmatrix}\hat{\boldsymbol{\theta}}\\-1\end{bmatrix}$ where only the last entry is -1, then the bias of the weighted system can be written as:

$$Bias_{post} = A\boldsymbol{x} = \begin{bmatrix}A_{:}\\ \boldsymbol{a}_e\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\theta}}\\-1\end{bmatrix} \tag{3}$$

Another way to present our question is that how large the ratio, $\frac{(\boldsymbol{a}_e\boldsymbol{x})^2}{||A_{:}\boldsymbol{x}||^2}$, can possibly be. Translating the ratio into English, it means that how off we are from the pre-treatment fit by introducing such column of $\boldsymbol{a}_e$ in the post-treatment. And to achieve a upper bound of it, we would like to find its maximal eigenvalue or the largest singular value.

$$\begin{aligned}
\max_x[\frac{||\boldsymbol{a}_e\boldsymbol{x}||^2}{||A_{:}\boldsymbol{x}||^2}] &= \max_x[\frac{\boldsymbol{x}'\boldsymbol{a}_e'\boldsymbol{a}_e\boldsymbol{x}}{\boldsymbol{x}'A_{:}'A_{:}\boldsymbol{x}}]\\
&= \max_x[\frac{\boldsymbol{x}'C\boldsymbol{x}}{\boldsymbol{x}'B\boldsymbol{x}}] \text{ , where } B = A_{:}'A_{:}, C = \boldsymbol{a}_e'\boldsymbol{a}_e\\
&= \max_\lambda[\frac{\boldsymbol{v}'B^{-\frac{1}{2}}CB^{-\frac{1}{2}}\boldsymbol{v}}{\boldsymbol{v}'\boldsymbol{v}}] \text{ , where we let } \boldsymbol{v} = B^{\frac{1}{2}}\boldsymbol{x}\\
&= \max_\lambda[B^{-\frac{1}{2}}CB^{-\frac{1}{2}}]\\
&= \max_\lambda[(A_{:}'A_{:})^{-\frac{1}{2}}(\boldsymbol{a}_e'\boldsymbol{a}_e)(A_{:}'A_{:})^{-\frac{1}{2}}]
\end{aligned} \tag{4}$$

Now our question has been transformed into finding the maximum eigenvalue of the system above to obtain a reasonable bound for such ratio. Though in different form and notation, the bound we would like to approach is equivalent to what Abadie et al. (2010) had achieved in terms of the primary bounding subject. To see this, we can make some further decomposition to the expression above. Since we observe from above that $\boldsymbol{a}_e(A_{:}'A_{:})^{-\frac{1}{2}}$ is a rank-1 eigenvector, the matrix $(A_{:}'A_{:})^{-\frac{1}{2}}(\boldsymbol{a}_e'\boldsymbol{a}_e)(A_{:}'A_{:})^{-\frac{1}{2}}$ is also rank-1, and the only non-zero eigenvalue is given by $\boldsymbol{a}_e(A_{:}'A_{:})^{-1}\boldsymbol{a}_e'$, which simplified the ultimate bounding target.

One last derivation would help us better see the commonality between two bounding targets. Following the factorization presented above, we denote $A_{:} = \boldsymbol{\Lambda}_{:}\boldsymbol{\mu}$ for $\boldsymbol{\Lambda}_{:}$ with the post-treatment column excluded, and $\boldsymbol{a}_e = \boldsymbol{\Lambda}_{\cdot t}\boldsymbol{\mu}$ for $\boldsymbol{\Lambda}_{\cdot t}$, representing the post-treatment column. Then we can write and make further derivation for the expression above to be:

$$\begin{aligned}
\boldsymbol{a}_e(A_{:}'A_{:})^{-1}\boldsymbol{a}_e' &= (\boldsymbol{\Lambda}_{\cdot t}\boldsymbol{\mu})[(\boldsymbol{\Lambda}_{:}\boldsymbol{\mu})'(\boldsymbol{\Lambda}_{:}\boldsymbol{\mu})]^{-1}(\boldsymbol{\Lambda}_{\cdot t}\boldsymbol{\mu})'\\
&= (\boldsymbol{\Lambda}_{\cdot t}\boldsymbol{\mu})[(\boldsymbol{\mu}'\boldsymbol{\Lambda}_{:}')(\boldsymbol{\Lambda}_{:}\boldsymbol{\mu})]^{-1}(\boldsymbol{\mu}'\boldsymbol{\Lambda}_{\cdot t}')
\end{aligned} \tag{5}$$

To be more generalized, we need to consider cases when $A_{:}$ is non-invertible. A way to get around with it is to apply pseudo-inverse, or Moore–Penrose inverse, in replacing the common inverse above. A useful property of pseudo-inverse we apply here is that for two matrices $A$ and $B$, $(AB)^+ = B^+A^+$ when $A$ has linearly independent columns and $B$ has linearly independent rows. Using this property, we can eventually obtain[2]:

$$\boldsymbol{a}_e(A_{:}^{*}A_{:})^{+}\boldsymbol{a}_e^{*} = \boldsymbol{\Lambda}_{\cdot t}(\boldsymbol{\Lambda}_{:}^{*}\boldsymbol{\Lambda}_{:})^{+}\boldsymbol{\Lambda}_{\cdot t}^{*} \tag{6}$$

Detailed use of such property in derivation is discussed in Appendix 2. Comparing equation (1) and equation (6) above, we noticed that, with our expression being a more general description, ultimately the bounding object in our derivation and workflow, despite notations, is the same as the aim Abadie had presented before, and in the following sections, we would like to derive a bound using our approach.

# 2 Our Bound and Justifications

## 2.1 The Bound

Another way to express the bounding target is by saying how smaller than 1 the ratio $\frac{\|A_{:}\boldsymbol{x}\|^2}{\|A\boldsymbol{x}\|^2}$ can be, as it is describing the deficit of $A$ from $A_{:}$ when the extra post-treatment column is incorporated. This is the same as the description of our previous ratio with respect to the effect of post-treatment units on the entire system. Given the commonality of our bounding target, we would like to derive a bound with same goal on post-treatment bias using our matrix $A$ defined previously and a way to lower bound the ratio, $\min_x[\frac{\|A_{:}\boldsymbol{x}\|^2}{\|A\boldsymbol{x}\|^2}]$, and its final bound forms as follows:

$$\begin{aligned} Bias_{ratio} &= 1 - \sum_k \sigma_k^{-2}(\boldsymbol{v}_k'\boldsymbol{a}_e)^2 \\ &\geq 1 - \frac{1}{cnp} n \cdot \max \|\boldsymbol{a}_e\| \\ &\approx 1 - \frac{1}{p} \text{ if we ignore the constants} \end{aligned} \tag{7}$$

And our target bound on the post-treatment period can be written as:

$$Bias_{post} \leq RMSE_{pre} * \frac{1}{(1-\frac{1}{p})} \tag{8}$$

In our bound expression and derivation process in section 2.3, it primarily involves the notation of two concepts or approaches. First, we've expressed matrix $A$ in terms of its Singular Value Decomposition (SVD), then $A = \sum_k \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k'$, where $\sigma_k$'s are singular values of matrix $A$, and $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$ are orthonormal basis. Also we incorporate the Strong Factor Assumption here to bound the singular values of $A$, that we suppose $\sigma_k \geq \sqrt{cnp}$ for $\sigma_k$, the $k$th singular value of matrix A, some constant $c$, number of units $n$, and number of time periods $p$.

[2] $*$ denoting conjugate transpose in context of pseudo inverse

## 2.2 Inferences

As our bound applied the Strong Factor Assumption stating that $\sigma_k \geq \sqrt{cnp}$ for $\sigma_k$, the kth singular value of matrix A, some constant $c$, number of units $n$, and number of time periods pre-treatment $p$. More general forms of the assumption can be written as $\sigma_k \geq (cnp)^{\alpha}$ for any $\alpha \in [0, \frac{1}{2}]$, or $\sigma_k \geq cn^{\alpha_n} p^{\alpha_p}$, where both $\alpha_n, \alpha_p \in [0, \frac{1}{2}]$ can be even more generalized. And in this most generalized form, our bound on the ratio becomes:

$$
\begin{aligned}
\min_{\lambda}[\frac{\|A_{:}\boldsymbol{x}\|^2}{\|A\boldsymbol{x}\|^2}] &= 1 - \sum_k \sigma_k^{-2}(\boldsymbol{v}_k'\boldsymbol{a}_e)^2 \\
&\geq 1 - \sum_k \frac{1}{cn^{2\alpha_n}p^{2\alpha_p}}(\boldsymbol{v}_k'\boldsymbol{a}_e)^2 \\
&= 1 - \frac{1}{cn^{2\alpha_n}p^{2\alpha_p}}\|P_{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_k}(\boldsymbol{a}_e)\|^2 \\
&\geq 1 - \frac{1}{cn^{2\alpha_n}p^{2\alpha_p}} n \cdot \max\|\boldsymbol{a}_e\| \\
&\geq 1 - \frac{n^{1-2\alpha_n}}{p^{2\alpha_p}} \text{ if we ignore the constants}
\end{aligned}
\tag{9}
$$

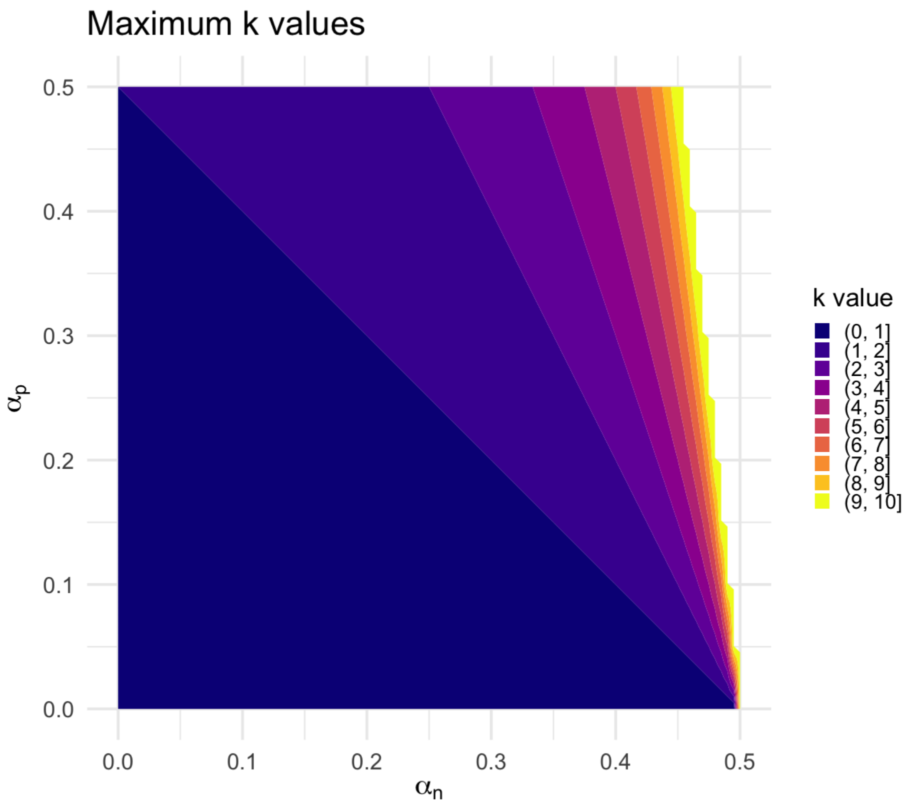
Consider the most generalized format. Now we suppose $n = p^{\kappa}$ for any $\kappa > 0$. Then the bound can be written as

$$
\begin{aligned}
\min_{\lambda}[\frac{\|A_{:}\boldsymbol{x}\|^2}{\|A\boldsymbol{x}\|^2}] &\geq 1 - \frac{n^{1-2\alpha_n}}{p^{2\alpha_p}} \\
&= 1 - \frac{p^{\kappa(1-2\alpha_n)}}{p^{2\alpha_p}} \\
&= 1 - p^{\kappa - 2\kappa\alpha_n - 2\alpha_p}
\end{aligned}
\tag{10}
$$

By our settings, a small post-treatment error would be achieved when $p^{\kappa-2\kappa\alpha_n-2\alpha_p}$ is small, and it's also reasonable to believe that $\kappa - 2\kappa\alpha_n - 2\alpha_p \leq 0$ since post-treatment error is generally no less than the pre-treatment error. With that we see $\kappa \leq \frac{2\alpha_p}{1-2\alpha_n}$ for $\alpha_p, \alpha_n \in [0, \frac{1}{2}]$. In figure 1, we visualize the maximum $\kappa$ value required for a reasonable bound with corresponding $\alpha$'s, and lower $\kappa$ would render smaller bias in post-treatment period. The discrete regions colored differently represent combinations of $\alpha_p$ and $\alpha_n$ inducing the corresponding $\kappa$ value. For example, when $\alpha_p = 0.3$ and $\alpha_n = 0.2$, $\kappa$ needs to be smaller than 1 to achieve a reasonable bound, or further, a small bias in post treatment.

One thing to notice is the behavior of $\kappa$ at the 4 corners of the square:

1. The Upper Left: when $\alpha_p = 0.5$ and $\alpha_n = 0$, we require $\kappa \leq 1$
2. The Upper Right: when $\alpha_p = 0.5$ and $\alpha_n = 0.5$, $\kappa$ can be arbitrary
3. The Lower Left: when $\alpha_p = 0$ and $\alpha_n = 0$, it's impossible as $\kappa \leq 0$
4. The Lower Right: when $\alpha_p \to 0$ and $\alpha_n \to 0.5$, we require $\kappa \leq 1$




Figur 1: Small Bias with Different *k* Values Regarding Strong Factor Assumption

At the same time, we can look at the behavior of $\kappa$ on the 4 sides of the square and its 2 diagonals:

1. Left side: when $\alpha_n = 0$, we require $\kappa < 2\alpha_p$
2. Upper side: when $\alpha_p = 0.5$, we require $\kappa \leq \frac{1}{1-2\alpha_n}$
3. Right side: when $\alpha_n = 0.5$, $\kappa$ can be arbitrary
4. Lower side: when $\alpha_p = 0$, it's impossible as $\kappa \leq 0$
5. Diagonal from lower left to upper right: when $\alpha_n = \alpha_p = \alpha$, we require $\kappa \leq \frac{2\alpha}{1-2\alpha}$ for $\alpha \in [0, \frac{1}{2}]$
6. Diagonal from upper left to lower right: when $\alpha_p = -\alpha_n + \frac{1}{2}$, we require $\kappa \leq 1$

A further inference can be drawn from the plot that when the sum of $\alpha_p + \alpha_n$ becomes larger, or when the factors become relatively stronger, the maximum $\kappa$ values would also grow until $\kappa$ being arbitrary when both $\alpha$'s are $\frac{1}{2}$ to achieve a reasonable bound and possibly a smaller bias, which infers that we are able to incorporate more the number of units $n$ when we have relatively stronger factors such that $n = p^\kappa$ for pre-treatment periods $p$.

## 2.3 Derivation

To elaborate the derivation of our bound, under similar steps we've achieved from equation (4), we can see a similar resulted equation for our updated target on obtaining a lower bound on the ratio: $\min_x[\frac{\|A_{\cdot}\boldsymbol{x}\|^2}{\|A\boldsymbol{x}\|^2}] = \min_\lambda[(A'A)^{-\frac{1}{2}}(A'_{\cdot}A_{\cdot})(A'A)^{-\frac{1}{2}}]$. To continue, we would like to simplify the expression in terms of Singular Value Decomposition (SVD) of matrix $A$ that $A = \sum_k \sigma_k \boldsymbol{u}_k \boldsymbol{v}'_k$, then we write $A'A = \sum_k \sigma_k^2 \boldsymbol{v}_k \boldsymbol{v}'_k$, where $\sigma_k$'s are singular values of matrix $A$; $\boldsymbol{u}_k$ is orthonormal basis such that $\boldsymbol{u}'_k \boldsymbol{u}_k = 1$, cancelled when we break the parenthesis with multiplication; and $\boldsymbol{v}_k \boldsymbol{v}'_k$ forms the resulting matrix with the outer product. At the same time, we are also able to write $(A'A)_{ij} = (A'_{:i}A_{:j}) - \boldsymbol{a}_{e.}\boldsymbol{a}'_{e_j}$ by definition. Therefore, we can see the following transformation:

$$
\begin{aligned}
\min_x[\frac{\|A_{\cdot}\boldsymbol{x}\|^2}{\|A\boldsymbol{x}\|^2}] &= \min_\lambda[(A'A)^{-\frac{1}{2}}(A'_{\cdot}A_{\cdot})(A'A)^{-\frac{1}{2}}] \\
&= \min_\lambda[(\sum_k \sigma_k^{-1}\boldsymbol{v}_k\boldsymbol{v}'_k)(\sum_l \sigma_l^2 \boldsymbol{v}_l\boldsymbol{v}'_l - \boldsymbol{a}_e\boldsymbol{a}'_e)(\sum_m \sigma_m^{-1}\boldsymbol{v}_m\boldsymbol{v}'_m)] \\
&= \min_\lambda[(\sum_k \sigma_k^{-1}\boldsymbol{v}_k\boldsymbol{v}'_k)(\sum_l \sigma_l^2 \boldsymbol{v}_l\boldsymbol{v}'_l)(\sum_m \sigma_m^{-1}\boldsymbol{v}_m\boldsymbol{v}'_m) \\
&- (\sum_k \sigma_k^{-1}\boldsymbol{v}_k\boldsymbol{v}'_k)(\boldsymbol{a}_e\boldsymbol{a}'_e)(\sum_m \sigma_m^{-1}\boldsymbol{v}_m\boldsymbol{v}'_m)] \\
&= \min_\lambda[(\sum_{k=m=l} \boldsymbol{v}_k\boldsymbol{v}'_m) - (\boldsymbol{w}\boldsymbol{w}')] \text{ , where } \boldsymbol{w} = \sum_k \sigma_k^{-1}(\boldsymbol{a}'_e\boldsymbol{v}_k)\boldsymbol{v}_k
\end{aligned}
\tag{11}
$$

The cancellation above is made through utilizing the property of orthonormal basis such that only when $i = j$, the inner product $\boldsymbol{v}'_i\boldsymbol{v}_j = 1$, otherwise it's 0.

Now suppose scalar $\tilde{\sigma}_k^{-1} = \sigma_k^{-1}(\boldsymbol{a}_e \boldsymbol{v}_k)$ and $M = (\sum_{k=m=l} \boldsymbol{v}_k \boldsymbol{v}_m') - (\boldsymbol{w}\boldsymbol{w}')$, then we are able to rephrase our minimum eigenvalue of $M$ as $\min_\lambda(M) = \min_{\boldsymbol{z}:\|\boldsymbol{z}\|=1}(\boldsymbol{z}'M\boldsymbol{z})$, where $\boldsymbol{z}$ is the eigenvector of $M$. Then continuing our decomposition process:

$$\begin{aligned}
\min_\lambda(M) &= \min_{\boldsymbol{z}:\|\boldsymbol{z}\|=1}(\boldsymbol{z}'M\boldsymbol{z}) \\
&= \min_{\boldsymbol{z}:\|\boldsymbol{z}\|=1}[\boldsymbol{z}'[(\sum_k \boldsymbol{v}_k\boldsymbol{v}_k') - (\sum_{k,l} \tilde{\sigma}_k^{-1}\tilde{\sigma}_l^{-1}\boldsymbol{v}_k\boldsymbol{v}_l')]\boldsymbol{z}] \\
&= \min_{\boldsymbol{z}:\|\boldsymbol{z}\|=1}[\sum_k (\boldsymbol{z}'\boldsymbol{v}_k)^2 - \sum_{k,l} \tilde{\sigma}_k^{-1}\tilde{\sigma}_l^{-1}(\boldsymbol{z}'\boldsymbol{v}_k)(\boldsymbol{z}'\boldsymbol{v}_l)] \\
&= \min_{\boldsymbol{z}:\|\boldsymbol{z}\|=1}[\sum_{k,l} (\boldsymbol{z}'\boldsymbol{v}_k)(\boldsymbol{z}'\boldsymbol{v}_l)\mathbf{1}_{k=l} - \sum_{k,l} \tilde{\sigma}_k^{-1}\tilde{\sigma}_l^{-1}(\boldsymbol{z}'\boldsymbol{v}_k)(\boldsymbol{z}'\boldsymbol{v}_l)] \\
&= \min_{\boldsymbol{z}:\|\boldsymbol{z}\|=1}[\sum_{k,l}[\mathbf{1}_{k=l} - \tilde{\sigma}_k^{-1}\tilde{\sigma}_l^{-1}](\boldsymbol{z}'\boldsymbol{v}_k)(\boldsymbol{z}'\boldsymbol{v}_l)]
\end{aligned} \tag{12}$$

Now, we would like to write $\boldsymbol{z}$ in orthonormal basis format: $\boldsymbol{z} = \sum_r \alpha_r \boldsymbol{v}_r$. And by the previous property of orthonormal basis, we see that $\|\boldsymbol{z}\|^2 = (\sum_r \alpha_r \boldsymbol{v}_r)(\sum_s \alpha_s \boldsymbol{v}_s) = \sum_{r,s} \alpha_r\alpha_s(\boldsymbol{v}_r'\boldsymbol{v}_s) = \sum_{r=s} \alpha_r^2 = 1$. Then we further rewrite the expression where again we apply the property of orthonormal basis:

$$\begin{aligned}
\min_{\boldsymbol{z}:\|\boldsymbol{z}\|=1}(\boldsymbol{z}'M\boldsymbol{z}) &= \min_{\|\alpha\|=1}[\sum_{k,l,r,s}[\mathbf{1}_{k=l} - \tilde{\sigma}_k^{-1}\tilde{\sigma}_l^{-1}](\alpha_r \boldsymbol{v}_r'\boldsymbol{v}_k)(\alpha_s \boldsymbol{v}_s'\boldsymbol{v}_l)] \\
&= \min_{\|\alpha\|=1}[\sum_{k=l}[\mathbf{1}_{k=l} - \tilde{\sigma}_k^{-1}\tilde{\sigma}_l^{-1}]\alpha_k\alpha_l] \\
&= \min_{\|\alpha\|=1}[\boldsymbol{\alpha}'\{I - \tilde{\boldsymbol{\sigma}}^{-1}\tilde{\boldsymbol{\sigma}}^{-1}\}\boldsymbol{\alpha}] \text{ in matrix format}
\end{aligned} \tag{13}$$

Afterwards, we know that if $\tilde{\boldsymbol{\sigma}}$ and $\boldsymbol{\alpha}$ are in the same direction, then the minimizing vector $\boldsymbol{\alpha}$ is proportional to $\tilde{\boldsymbol{\sigma}}$, the singular values of the original matrix and the projection of $\boldsymbol{a}_e$ onto the right singular vectors, which means that $\boldsymbol{\alpha} = c\tilde{\boldsymbol{\sigma}}$ for some scalar $c$. Then, since $\|\boldsymbol{\alpha}\| = 1$, we see that $1 = \|\boldsymbol{\alpha}\|^2 = c^2\|\tilde{\boldsymbol{\sigma}}\|^2$, $c = \frac{1}{\|\tilde{\boldsymbol{\sigma}}\|}$, and $\boldsymbol{\alpha} = \frac{\tilde{\boldsymbol{\sigma}}}{\|\tilde{\boldsymbol{\sigma}}\|}$

$$\begin{aligned}
\min_{\|\alpha\|=1}[\boldsymbol{\alpha}'\{I - \tilde{\boldsymbol{\sigma}}^{-1}\tilde{\boldsymbol{\sigma}}^{-1}\}\boldsymbol{\alpha}] &= \frac{\tilde{\boldsymbol{\sigma}}'}{\|\tilde{\boldsymbol{\sigma}}\|}\{I - \tilde{\boldsymbol{\sigma}}^{-1}\tilde{\boldsymbol{\sigma}}^{-1}\}\frac{\tilde{\boldsymbol{\sigma}}}{\|\tilde{\boldsymbol{\sigma}}\|} \\
&= \frac{\tilde{\boldsymbol{\sigma}}'\tilde{\boldsymbol{\sigma}}}{\|\tilde{\boldsymbol{\sigma}}\|^2} - \frac{\tilde{\boldsymbol{\sigma}}'\tilde{\boldsymbol{\sigma}}^{-1}\tilde{\boldsymbol{\sigma}}^{-1}\tilde{\boldsymbol{\sigma}}}{\|\tilde{\boldsymbol{\sigma}}\|^2} \\
&= \frac{\|\tilde{\boldsymbol{\sigma}}\|^2}{\|\tilde{\boldsymbol{\sigma}}\|^2} - \frac{\left\|\tilde{\boldsymbol{\sigma}}^{-1}\right\|^2\tilde{\boldsymbol{\sigma}}'\tilde{\boldsymbol{\sigma}}}{\|\tilde{\boldsymbol{\sigma}}\|^2} \\
&= 1 - \left\|\tilde{\boldsymbol{\sigma}}^{-1}\right\|^2 \\
&= 1 - \sum_k \sigma_k^{-2}(\boldsymbol{v}_k'\boldsymbol{a}_e)^2
\end{aligned} \tag{14}$$

Now we would like to use our Strong Factor Assumption to further specify such bound, that we suppose $\sigma_k \geq \sqrt{cnp}$ for $\sigma_k$, the kth singular value of matrix A, some constant $c$, number of units $n$, and number of time periods in pre-treatment $p$. Considering an orthonormal basis

$\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ that spans the entire space in which $\boldsymbol{a}_e$ exists, we see that $||P_{\boldsymbol{v}_1,...,\boldsymbol{v}_n}(\boldsymbol{a}_e)||^2 = ||\boldsymbol{a}_e||^2$. Using all the justifications above we see:

$$
\begin{aligned}
1 - \sum_k \sigma_k^{-2}(\boldsymbol{v}_k' \boldsymbol{a}_e)^2 &\geq 1 - \sum_k \frac{1}{cnp}(\boldsymbol{v}_k' \boldsymbol{a}_e)^2 \\
&= 1 - \frac{1}{cnp}\sum_k (\boldsymbol{v}_k' \boldsymbol{a}_e)^2 \\
&= 1 - \frac{1}{cnp}\,||P_{\boldsymbol{v}_1,...,\boldsymbol{v}_k}(\boldsymbol{a}_e)||^2 \\
&= 1 - \frac{1}{cnp}\,||\boldsymbol{a}_e||^2 \\
&\geq 1 - \frac{1}{cnp} n \cdot \max ||\boldsymbol{a}_e||
\end{aligned} \tag{15}
$$

So eventually the lower bound we give to the original ratio becomes:

$$
\begin{aligned}
\min_\lambda [\frac{||A_{:}\boldsymbol{x}||^2}{||A\boldsymbol{x}||^2}] &\geq 1 - \frac{1}{cnp} n \cdot \max ||\boldsymbol{a}_e|| \\
&= 1 - \frac{1}{cp} \cdot \max ||\boldsymbol{a}_e|| \\
&\approx 1 - \frac{1}{p} \text{ if we ignore the constants}
\end{aligned} \tag{16}
$$

One immediate observation we see from the lower bound is that as $p$, number of time periods in pre-treatment, increases, the minimal value of such ratio would approach to 1, representing a small post-treatment error introduced by $\boldsymbol{a}_e$, which is close to the pre-treatment error. When $p$ is small, the ratio decreases, demonstrating the portion of error added from $\boldsymbol{a}_e$ is increasingly larger than the pre-treatment error.

We would like to continue to bound on the post-treatment bias as our ultimate goal, and we can derive it from the bounded relationship between the Post-Treatment Error and Root Mean Squared Error (RMSE) in Pre-Treatment Stage, where the ratio between the two can be written as $\frac{Bias_{post}}{RMSE_{pre}} = \frac{(\boldsymbol{a}_e\boldsymbol{x})^2}{\frac{1}{p}||A_{:}\boldsymbol{x}||^2}$:

$$
\begin{aligned}
\max_x [\frac{(\boldsymbol{a}_e\boldsymbol{x})^2}{\frac{1}{p}||A_{:}\boldsymbol{x}||^2}] &= \max_x [p \cdot \frac{(\boldsymbol{a}_e\boldsymbol{x})^2}{||A_{:}\boldsymbol{x}||^2}] \\
&= \max_x [p \cdot \frac{||A\boldsymbol{x}||^2 - ||A_{:}\boldsymbol{x}||^2}{||A_{:}\boldsymbol{x}||^2}] \\
&= \max_x [p \cdot [\frac{||A\boldsymbol{x}||^2}{||A_{:}\boldsymbol{x}||^2} - 1]] \\
&\leq p \cdot [\frac{1}{1 - \frac{1}{p}} - 1] \\
&= \frac{1}{1 - \frac{1}{p}}
\end{aligned} \tag{17}
$$

Then we are able to conclude the bound on the post-treatment bias in terms of the pre-treatment RMSE with some similar inferences we can observe: $Bias_{post} \leq RMSE_{pre} * \frac{1}{(1-\frac{1}{p})}$

# 3 Appendix 1: Modified Derivation of Bound in Abadie et al. (2010)

For any synthetic control (SC), we consider a weighted combination of the outcomes of units in absence of intervention, and suppose $\theta_j$ to be the weight assigned for $j$th unit, then $Y_{jt}$ denote the outcome of $j$th unit at time period $t$. Thus, the SC estimator for the treated is $\hat{Y}_{jt} = \sum_{j=1}^{n-1} \theta_j Y_{jt}$. Now consider the factor model for the outcomes: $Y_{jt} = A_{jt} + \varepsilon_{jt} = \mathbf{\Lambda}_{\cdot t} \boldsymbol{\mu}_j + \varepsilon_{jt}$, for $\mu_j$ being vector of loadings, $\Lambda_{\cdot t}$ being vector of factors, and $\varepsilon_{jt}$ being error term. Following the original paper, we see the bias terms with covariance terms ignored temporarily for our argument by plugging in the SC estimator and factor model as we defined:

$$\begin{aligned} Y_{nt} - \hat{Y}_{jt} &= Y_{nt} - \sum_{j=1}^{n-1} \theta_j Y_{jt} \\ &= (\mathbf{\Lambda}_{\cdot t} \boldsymbol{\mu}_n + \varepsilon_{nt}) - \sum_{j=1}^{n-1} \theta_j (\mathbf{\Lambda}_{\cdot t} \boldsymbol{\mu}_j + \varepsilon_{jt}) \\ &= \mathbf{\Lambda}_{\cdot t} (\boldsymbol{\mu}_n - \sum_{j=1}^{n-1} \theta_j \boldsymbol{\mu}_j) + \sum_{j=1}^{n-1} \theta_j (\varepsilon_{nt} - \varepsilon_{jt}) \end{aligned} \tag{18}$$

Now consider the pre-treatment period factor model: $Y_{nt}^P = \mathbf{\Lambda}_{:} \boldsymbol{\mu}_n + \varepsilon_{:_n}$. Based on a similar derivation above, we see that

$$\begin{aligned} Y_{nt}^P - \hat{Y}_{jt}^P &= Y_{nt}^P - \sum_{j=1}^{n-1} \theta_j Y_{jt}^P \\ &= \mathbf{\Lambda}_{:} (\boldsymbol{\mu}_n - \sum_{j=1}^{n-1} \theta_j \boldsymbol{\mu}_j) + \sum_{j=1}^{n-1} \theta_j (\boldsymbol{\varepsilon}_{:_n} - \boldsymbol{\varepsilon}_{:_j}) \end{aligned} \tag{19}$$

Then we are able to combine the two equations into one to better represent the bias. With operations of multiplying $(\mathbf{\Lambda}_{:}' \mathbf{\Lambda}_{:})^{-1} \mathbf{\Lambda}_{:}'$, we are able to express the general bias term by incorporating the pre-treatment fit term $Y_{nt}^P - \sum_{j=1}^{n-1} \theta_j Y_{jt}^P$:

$$\begin{aligned} Y_{nt} - \sum_{j=1}^{n-1} \theta_j Y_{jt} = \mathbf{\Lambda}_{\cdot t} (\mathbf{\Lambda}_{:}' \mathbf{\Lambda}_{:})^{-1} \mathbf{\Lambda}_{:}' (Y_{nt}^P - \sum_{j=1}^{n-1} \theta_j Y_{jt}^P) \\ - \mathbf{\Lambda}_{\cdot t} (\mathbf{\Lambda}_{:}' \mathbf{\Lambda}_{:})^{-1} \mathbf{\Lambda}_{:}' (\boldsymbol{\varepsilon}_{:_n} - \sum_{j=1}^{n-1} \theta_j \boldsymbol{\varepsilon}_{:_j}) \\ + \sum_{j=1}^{n-1} \theta_j (\varepsilon_{nt} - \varepsilon_{jt}) \end{aligned} \tag{20}$$

The above equation, when we plug in Equation (2), would be equal to the original general bias equation specified in Equation (1) with cancellations of $\mathbf{\Lambda}_{:}$'s and $\boldsymbol{\varepsilon}$'s. By plugging in

Equation (2), we see $\boldsymbol{\Lambda}_{:}$'s cancelled when $(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}\boldsymbol{\Lambda}_{:}'$ times $\boldsymbol{\Lambda}_{:}$. We also see that by our constraint on weights, $\sum_j \theta_j = 1$, so then, $\boldsymbol{\varepsilon}_{:n} = \sum_j \theta_j \boldsymbol{\varepsilon}_{:n}$, and $\sum_j \theta_j(\boldsymbol{\varepsilon}_{:n} - \boldsymbol{\varepsilon}_{:j}) = \boldsymbol{\varepsilon}_{:n} - \sum_j \theta_j \boldsymbol{\varepsilon}_{:j}$, from which we are able to see the ultimate cancellation and the recovery of original bias expression with matrix expression in pre-treatment fit, $Y_{nt}^P - \sum_{j=1}^{n-1} \theta_j Y_{jt}^P$, introduced.

Stated above, we would like to only discuss the case of perfect pre-treatment fit specifically, so we suppose the solution to our optimization problem of satisfying weights is $\hat{\boldsymbol{\theta}} = \{\hat{\theta}_1, ..., \hat{\theta}_{n-1}\}$ such that :

$$\sum_{j=1}^{n-1} \hat{\theta}_j Y_{j1} = Y_{n1}, \qquad \sum_{j=1}^{n-1} \hat{\theta}_j Y_{j2} = Y_{n2}, ..., \qquad \sum_{j=1}^{n-1} \hat{\theta}_j Y_{jp-1} = Y_{np-1} \tag{21}$$

Then under such solution, we see the term $Y_{nt}^P - \sum_{j=1}^{n-1} \theta_j Y_{jt}^P = 0$ under perfect pre-treatment fit assumption and now we are left with:

$$Y_{nt} - \sum_{j=1}^{n-1} \theta_j Y_{jt} = -\boldsymbol{\Lambda}_{\cdot t}(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}\boldsymbol{\Lambda}_{:}'(\boldsymbol{\varepsilon}_{:n} - \sum_{j=1}^{n-1} \hat{\theta}_j \boldsymbol{\varepsilon}_{:j}) + \sum_{j=1}^{n-1} \hat{\theta}_j(\varepsilon_{nt} - \varepsilon_{jt}) \tag{22}$$

To better analyze the bias, we split the entire expression into three parts, $R_{1t}, R_{2t}, R_{3t}$, where:

$$\begin{aligned} R_{1t} &= \boldsymbol{\Lambda}_{\cdot t}(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}\boldsymbol{\Lambda}_{:}' \sum_{j=1}^{n-1} \hat{\theta}_j \boldsymbol{\varepsilon}_{:j} \\ R_{2t} &= -\boldsymbol{\Lambda}_{\cdot t}(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}\boldsymbol{\Lambda}_{:}'\boldsymbol{\varepsilon}_{:n} \\ R_{3t} &= \sum_{j=1}^{n-1} \hat{\theta}_j(\varepsilon_{jt} - \varepsilon_{nt}) \end{aligned} \tag{23}$$

When we consider each components in the three parts above, we see that in $R_{2t}$, $\boldsymbol{\varepsilon}_{:n}$ is random error from the pre-treatment period that would not introduce the systematic bias. And in $R_{3t}$, the weights $\hat{\theta}_j$ is uncorrelated with the noise terms $\varepsilon_{jt}$ and $\varepsilon_{nt}$). Thus we conclude that $E[R_{2t}]$ and $E[R_{3t}]$ are both zero, and our bias term would be shortened to only $E[R_{1t}]$, where $\varepsilon_{:j}$ is correlated with the weights $\hat{\theta}_j$ that would bring the bias into our system,

$$Bias_{post} = Y_{nt} - \sum_{j=1}^{n-1} \hat{\theta}_j Y_{jt} = E[R_{1t}] = E[\boldsymbol{\Lambda}_{\cdot t}(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}\boldsymbol{\Lambda}_{:}' \sum_{j=1}^{n-1} \hat{\theta}_j \boldsymbol{\varepsilon}_{:j}] \tag{24}$$

In the original paper, scalar multiplication form is used to denote the bias in a simpler way, whereas in our justification, matrix multiplication is preferred to use for notation. Now we would like to denote $\hat{\boldsymbol{\theta}}$ to be the aggregate version of weights for each unit $\hat{\theta}_j$ and $\tilde{\boldsymbol{\varepsilon}}_j = \boldsymbol{\Lambda}_{\cdot t}(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}\boldsymbol{\Lambda}_{:}'\boldsymbol{\varepsilon}_{:j}$. Then $R_{1t} = \hat{\boldsymbol{\theta}}^T \tilde{\boldsymbol{\varepsilon}}_j$. Now we can apply a general version of Hölder's Inequality:

$$
\begin{aligned}
E[R_{1t}] &= E[\hat{\boldsymbol{\theta}}^T \tilde{\boldsymbol{\varepsilon}}_j] \\
&\leq \left\|\hat{\boldsymbol{\theta}}\right\|_q E[\left\|\tilde{\boldsymbol{\varepsilon}}_j\right\|_p] \\
&\leq \left\|\hat{\boldsymbol{\theta}}\right\|_q \{E[(\left\|\tilde{\boldsymbol{\varepsilon}}_j\right\|_p)^p]\}^{1/p}, \text{ for } \frac{1}{p}+\frac{1}{q}=1 \text{ by Jensen's Inequality} \\
&= \left\|\hat{\boldsymbol{\theta}}\right\|_q \cdot \{\sigma^p(p-1)!!\}^{1/p} \cdot c_p, \text{ for constant } c_p = \begin{cases} \sqrt{2/\pi}, & \text{if } p \text{ is odd.} \\ 1, & \text{if } p \text{ is even.} \end{cases} \\
&= c_p(p-1)!!^{1/p} \left\|\hat{\boldsymbol{\theta}}\right\|_q \sigma \\
&= \tilde{c_p} \left\|\hat{\boldsymbol{\theta}}\right\|_q \sigma, \text{ if we denote } c_p(p-1)!!^{1/p} \text{ as some constant } \tilde{c_p}
\end{aligned}
\tag{25}
$$

With $\left\|\hat{\boldsymbol{\theta}}\right\|_q$ being fixed and multiplied by a constant $\tilde{c_p}$, the ultimate task is to find out $\sigma$, where $\sigma^2 = Var(\tilde{\boldsymbol{\varepsilon}}_j)$, for $\tilde{\boldsymbol{\varepsilon}}_j = \boldsymbol{\Lambda}_{\cdot t}(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}\boldsymbol{\Lambda}_{:}'\boldsymbol{\varepsilon}_{:j}$ defined previously. Now we would like to additionally define a row vector $\boldsymbol{v}^T = \boldsymbol{\Lambda}_{\cdot t}(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}\boldsymbol{\Lambda}_{:}'$, so the definition of $\tilde{\boldsymbol{\varepsilon}}_j$ can be simplified into $\tilde{\boldsymbol{\varepsilon}}_j = \boldsymbol{v}^T\boldsymbol{\varepsilon}_{:j}$.

$$
\begin{aligned}
\sigma^2 &= Var[(\tilde{\boldsymbol{\varepsilon}}_j)^2] \\
&= Var[(\boldsymbol{v}^T\boldsymbol{\varepsilon}_{:j})] \\
&= (\boldsymbol{v}^T)\Sigma_{\boldsymbol{\varepsilon}_{:j}}(\boldsymbol{v}) \\
&= (\boldsymbol{v}^T)(\boldsymbol{v}), \text{ if we assume } \boldsymbol{\varepsilon}_{:j} \text{ is standard normal} \\
&= (\boldsymbol{\Lambda}_{\cdot t}(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}\boldsymbol{\Lambda}_{:}')(\boldsymbol{\Lambda}_{\cdot t}(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}\boldsymbol{\Lambda}_{:}')^T \\
&= (\boldsymbol{\Lambda}_{\cdot t}(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}\boldsymbol{\Lambda}_{:}')(\boldsymbol{\Lambda}_{:}(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}\boldsymbol{\Lambda}_{\cdot t}') \\
&= \boldsymbol{\Lambda}_{\cdot t}[(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})](\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}\boldsymbol{\Lambda}_{\cdot t}' \\
&= \boldsymbol{\Lambda}_{\cdot t}(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}\boldsymbol{\Lambda}_{\cdot t}'
\end{aligned}
\tag{26}
$$

To further provide a concrete bound for the variance, we suppose $\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:} = \boldsymbol{M}$ and apply Min-Max Theorem, which states that for a symmetric matrix $\boldsymbol{M}$ with eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, the quadratic form $\boldsymbol{x}'\boldsymbol{M}\boldsymbol{x}$ for any $\boldsymbol{x}$ satisfies $\lambda_{min}\boldsymbol{M}\left\|\boldsymbol{x}\right\|^2 \leq \boldsymbol{x}'\boldsymbol{M}\boldsymbol{x} \leq \lambda_{max}\boldsymbol{M}\left\|\boldsymbol{x}\right\|^2$. Now we suppose $\eta$ is the smallest eigenvalue of $\boldsymbol{M}$, then since the inverse of $\boldsymbol{M}$ has eigenvalues $\frac{1}{\lambda_i}$, the largest eigenvalue of $\boldsymbol{M}^{-1}$ is $\frac{1}{\eta}$. Therefore our bound on variance can be derived,

$$
\begin{aligned}
\sigma^2 &= \boldsymbol{\Lambda}_{\cdot t}(\boldsymbol{\Lambda}_{:}'\boldsymbol{\Lambda}_{:})^{-1}\boldsymbol{\Lambda}_{\cdot t}' \\
&\leq \lambda_{max}(\boldsymbol{M}^{-1})\left\|\boldsymbol{\Lambda}_{\cdot t}\right\|^2 \\
&= \frac{\left\|\boldsymbol{\Lambda}_{\cdot t}\right\|^2}{\eta}
\end{aligned}
\tag{27}
$$

Now lastly, we suppose $dim(\boldsymbol{\Lambda}_{\cdot t}) = F$, and suppose there exists a constant $\bar{\Lambda}$ such that $|\Lambda_{tf}| \leq \bar{\Lambda}$, we see that for $k$th factor, $\left\|\boldsymbol{\Lambda}_{\cdot t}\right\|^2 = \sum_{k=1}^F (\Lambda_{tk})^2 \leq \sum_{j=1}^F \bar{\Lambda}^2 = F\bar{\Lambda}^2$. And $\sigma$ would eventually have a bound,

$$\sigma \leq \sqrt{\frac{\|\mathbf{\Lambda}_{\cdot t}\|^2}{\eta}} \leq \sqrt{\frac{F\bar{\Lambda}^2}{\eta}} \tag{28}$$

Combining all partial results above, our final bound to our bias term, $Y_{nt} - \sum_{j=1}^{n-1} \theta_j Y_{jt}$, is

$$Bias_{post} \leq \tilde{c_p} \left\| \hat{\boldsymbol{\theta}} \right\|_q \sqrt{\frac{F\bar{\Lambda}^2}{\eta}} \tag{29}$$

# 4 Appendix 2: Justification of inherent commonality of two bounding approaches

To elaborate steps we obtain the equality: $\boldsymbol{a}_e(A^*_{:}A_{:})^+\boldsymbol{a}_e^* = \mathbf{\Lambda}_{\cdot t}(\mathbf{\Lambda}^*_{:}\mathbf{\Lambda}_{:})^+\mathbf{\Lambda}^*_{\cdot t}$, we need to incorporate the property of pseudo-inverse mentioned above that for two matrices $A$ and $B$, $(AB)^+ = B^+A^+$ when $A$ has linearly independent columns and $B$ has linearly independent rows. We can assume that, in the factorization described in Abadie, the rows of $\boldsymbol{\mu}$ and the columns of $\lambda$ linearly independent[3]. Achieving $F$ rows or columns would actually render the linear independence of the corresponding rows or columns for an item or the product of items in decomposed expression. Now we suppose $M = \boldsymbol{\mu}^*\mathbf{\Lambda}^*_{:}\mathbf{\Lambda}_{:}$ being $(n \times F)$ and $N = \boldsymbol{\mu}$ being $(F \times n)$. Therefore the columns of $M$ and the rows of $N$ are linearly independent, meaning we can apply such property that $(MN)^+ = N^+M^+$. Therefore, the following cancellation can be made:

$$\begin{aligned}
\boldsymbol{a}_e(A'_{:}A_{:})^{-1}\boldsymbol{a}'_e &= (\mathbf{\Lambda}_{\cdot t}\boldsymbol{\mu})[(\boldsymbol{\mu}^*\mathbf{\Lambda}^*_{:}\mathbf{\Lambda}_{:})\boldsymbol{\mu})]^+(\boldsymbol{\mu}^*\mathbf{\Lambda}^*_{\cdot t}) \\
&= (\mathbf{\Lambda}_{\cdot t}\boldsymbol{\mu})[MN]^+(\boldsymbol{\mu}^*\mathbf{\Lambda}^*_{\cdot t}) \\
&= (\mathbf{\Lambda}_{\cdot t}\boldsymbol{\mu})N^+M^+(\boldsymbol{\mu}^*\mathbf{\Lambda}^*_{\cdot t}) \\
&= (\mathbf{\Lambda}_{\cdot t}\boldsymbol{\mu})(\boldsymbol{\mu})^+(\boldsymbol{\mu}^*\mathbf{\Lambda}^*_{:}\mathbf{\Lambda}_{:})^+(\boldsymbol{\mu}^*\mathbf{\Lambda}^*_{\cdot t}) \\
&= \mathbf{\Lambda}_{\cdot t}(\boldsymbol{\mu}^*\mathbf{\Lambda}^*_{:}\mathbf{\Lambda}_{:})^+(\boldsymbol{\mu}^*\mathbf{\Lambda}^*_{\cdot t})
\end{aligned} \tag{30}$$

Applying our reasoning again, we suppose that $M_1 = \boldsymbol{\mu}^*$ being $(n \times F)$ and $N_1 = \mathbf{\Lambda}^*_{:}\mathbf{\Lambda}_{:}$ being $(F \times F)$. Similarly, since the columns of $M_1$ and the rows of $N_1$ are linearly independent, we apply the property again and see that:

$$\begin{aligned}
\boldsymbol{a}_e(A'_{:}A_{:})^{-1}\boldsymbol{a}'_e &= \mathbf{\Lambda}_{\cdot t}(\boldsymbol{\mu}^*\mathbf{\Lambda}^*_{:}\mathbf{\Lambda}_{:})^+(\boldsymbol{\mu}^*\mathbf{\Lambda}^*_{\cdot t}) \\
&= \mathbf{\Lambda}_{\cdot t}(\mathbf{\Lambda}^*_{:}\mathbf{\Lambda}_{:})^+(\boldsymbol{\mu}^*)^+(\boldsymbol{\mu}^*\mathbf{\Lambda}^*_{\cdot t}) \\
&= \mathbf{\Lambda}_{\cdot t}(\mathbf{\Lambda}^*_{:}\mathbf{\Lambda}_{:})^+\mathbf{\Lambda}^*_{\cdot t}
\end{aligned} \tag{31}$$

[3]Otherwise if these columns were not linearly independent, it would imply redundancy in the factorization. Such redundancy would mean that we could find an equivalent factorization using fewer than $F$ factors, where $F$ is the rank of the factorization. By choosing $F$ to be the minimal number of factors needed to represent the original matrix, we ensure that the columns of $\boldsymbol{\mu}$ and the columns of $\mathbf{\Lambda}$ are linearly independent.