SCIENTIFIC LETTERS

Econometrics Letters Volume (7), 2, 2020, June.

Panel Data Model and the Synthetic Control Method: A Note on the Unbiasedness of the Synthetic Control Estimator

Orkideh Gharehgozli[1]

Abstract

The synthetic control method has been used in comparative case studies in which the existence of a counter-factual unit with high level of similarities is crucial. The method tries to find the best fitted counterfactual unit with minimizing the discrepancy between the outcome values of the treated unit and the synthetic control unit. The current underlying data generating process of the synthetic control method is considered to be a factor model with time-variant coefficients and time-invariant variables. We believe in aggregate comparative case studies this model may be too restrictive, and in this paper, we show that the synthetic control method provides an unbiased estimator if the underlying model of the outcome variable of interest is considered to be a dynamic panel data model, in which the restriction of time-invariant variables is relaxed. We believe the unbiasedness opens a door to future studies aiming to expand the method and to compare it with a regression framework.

Keywords: Unbiasedness, Synthetic Control Method, Panel Data Model
JEL Codes : C23 ; C52; C13.

[1] (Corresponding Author) Department of Economics, Feliciano School of Business, Montclair State University, 1 Normal Ave, Montclair, NJ, USA 07043. Tel: +1 973 655 7504, e-mail: gharehgozlio@montclair.edu

1. Introduction

Synthetic control method provides a point estimate of the treatment effect and the method lacks a discussion of the underlying asymptotic distribution of the estimators and the standard errors. On the other hand, we can benefit from asymptotic assumptions while using a regression framework and many studies have been done and many methods have been offered in order to overcome the potential shortages of a traditional regression framework in such comparative case studies, specifically in our interest is a panel data model. In the literature, the suggested underlying data generating process of the synthetic control method is a factor model with time-variant coefficients and time-invariant variables. We believe in aggregate comparative case studies this model may be too restrictive. The purpose of this article is to clarify that under specific assumptions the synthetic control estimator will be an unbiased estimator in a two-way error component dynamic panel data model.

2. Unbiasedness of the Synthetic Control Estimator

For unit $i = 1, \dots, J+1$ and $t = 1, \dots, T_0, \dots, T$ where T_0 is the last period before the intervention, we have:

$$y_{it}^N = \rho y_{it-1}^N + Z_{it}'\theta + X_i'\gamma + \mu_i + \delta_t + \epsilon_{it} \quad (1)$$

$$Z_{it} = A_1 Z_{it-1} + A_2 y_{it-1}^N + \mu_{zi} + \delta_{zt} + \nu_{it} \quad (2)$$

where y_{it}^N is the outcome that would be observed in the absence of the intervention, for unit $i = 1, \dots, J+1$ and $t = 1, \dots, T_0, \dots, T$.
Let y_{it}^I be the outcome that would be observed for unit i at time t if unit i is exposed to the intervention in period $T_0 + 1$ to T. δ_t stands for time fixed effects which we assume to be constant across units. $\acute{X_\iota}$ is a $(1 \times M)$ row vector of time-invariant explanatory variables, γ is $(M \times 1)$, and θ is a $(K \times 1)$ vector of parameters. μ_i is unit i fixed effect. $\acute{Z_{\iota t}}$ is a $(1 \times K)$ row vector of time-variant covariates that moves over time by a process given in Equation (2). ϵ_{it} and ν_{it} are idiosyncratic shock at unit level with mean zero. So, for $t = 1, \dots, T_0$ and all $i = 1, \dots, J+1$, we have: $y_{it}^N = y_{it}^I$.[2]
Following (Abadie , et al., 2010), let $\alpha_{it} = y_{it}^I - y_{it}^N$ be the effect of the intervention for unit i at time t, and let D_{it} be an indicator that equals 1 is unit i is exposed to the intervention at time t and equals 0 otherwise. Therefore, the observed outcome for unit i at time t is $y_{it} = y_{it}^N + \alpha_{it} D_{it}$. We assume unit $i = 1$ is exposed to the intervention for period $t = T_0 + 1, \dots, T$; therefore, we want to estimate $\alpha_{1t} = y_{1t}^I - y_{1t}^N$ for $t = T_0 + 1, \dots, T$. Note that y_{it}^I is observed, and we just need to estimate y_{it}^N.
We start at one period post- intervention, and we show that the synthetic control estimator, which uses the weighted average of the donors to represent the outcome variable of interest

[2] We consider our model to be a specific case of (Abadie , et al., 2010)'s factor model. In the paper, the authors show that the synthetic control estimator is an unbiassed estimator for the intervention effect in a factor model such as:

$$y_{it}^N = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{it} \quad (3)$$

where δ_t is an unknown common factor with constant factor loadings across units, Z_i is a $(r \times 1)$ vector of observed covariates (not affected by the intervention), θ_t is a $(1 \times r)$ vector of unknown parameters, λ_t is a $(1 \times F)$ vector of unobserved common factors, μ_i is an $(F \times 1)$ vector of unknown factor loadings, and the error term ϵ_{it} are unobserved transitory shocks with mean zero.

when unobservable, is an unbiased estimator if the outcome variable of interest follows the Dynamic Panel Data process mentioned in Equation 1. Later, we will provide the same derivation for the second post-intervention period, and we conclude that, under the same assumptions, the synthetic control provides an unbiased estimator if the outcome variable follows a DPD process.

Period $T_0 + 1$:

When we substitute Equation 2 in Equation 1, we obtain:

$$y_{it}^N = \rho y_{it-1}^N + Z'_{it-1}A'_1\theta + y_{it-1}^N A'_2\theta + \mu'_{zi}\theta + \delta'_{zt}\theta + \nu'_{it}\theta + X'_i\gamma + \mu_i + \delta_t + \epsilon_{it} \quad (4)$$

Now we use $(J \times 1)$ vector of weights $W = (w_2, w_3, w_4, \dots, w_{J+1})'$ such that $w_j \geq 0$ for $j = 2, \dots, J+1$ and $w_2 + w_3 + w_4 + \cdots + w_{J+1} = 1$. Therefore, we have:

$$\sum_{j=2}^{J+1} w_j y_{jt}^N = [\rho + A'_2\theta] \sum_{j=2}^{J+1} w_j\, y_{jt-1}^N + \sum_{j=2}^{J+1} w_j\, Z'_{jt-1}A'_1\theta + \sum_{j=2}^{J+1} w_j\, \mu'_{zj}\theta + \delta'_{zt}\theta + \sum_{j=2}^{J+1} w_j\, \nu'_{jt}\theta + \sum_{j=2}^{J+1} w_j\, X'_j\gamma + \sum_{j=2}^{J+1} w_j\, \mu_j + \delta_t + \sum_{j=2}^{J+1} w_j\, \epsilon_{jt} \quad (5)$$

So, in the first post-intervention period we have:

$$y_{1T_0+1}^N - \sum_{j=2}^{J+1} w_j\, y_{jT_0+1}^N = A\left(y_{1T_0}^N - \sum_{j=2}^{J+1} w_j\, y_{jT_0}^N\right) + (Z_{1T_0} - \sum_{j=2}^{J+1} w_j Z_{jT_0})' A'_1\,\theta + \left(X_1 - \sum_{j=2}^{J+1} w_j\, X_j\right)'\gamma + \left(\mu_{z1} - \sum_{j=2}^{J+1} \mu_{zj}\right)'\theta + \left(\mu_1 - \sum_{j=2}^{J+1} w_j \mu_j\right) + \sum_{j=2}^{J+1} w_j (\nu_{1t} - \nu_{jt})'\theta + \sum_{j=2}^{J+1} w_j (\epsilon_{1t} - \epsilon_{jt}) \quad (6)$$

where $A = \rho + A'_2\theta$ is a scalar. Note that δ_t and $\delta'_{zt}\theta$ will be dropped out of the equation because sum of the weights is equal to1. Following (Abadie , et al., 2010), let Y_i^{pre} be a $(T_0 \times 1)$ vector of pre-intervention values of y for unit i. Therefore, for the first unit $(i = 1)$ we have:

$$Y_1^{pre} = \begin{bmatrix} y_{11} \\ y_{12} \\ . \\ . \\ . \\ y_{1T_0} \end{bmatrix}, \text{ and } Y_{1_{(-1)}}^{pre} = \begin{bmatrix} y_{10} \\ y_{11} \\ . \\ . \\ . \\ y_{1T_0-1} \end{bmatrix}, Z_1^{pre\prime} = \begin{bmatrix} Z'_{11} \\ Z'_{12} \\ . \\ . \\ . \\ Z'_{1T_0} \end{bmatrix}, \text{ and } Z_{1_{(-1)}}^{pre\ \prime} = \begin{bmatrix} Z'_{10} \\ Z'_{11} \\ . \\ . \\ . \\ Z'_{1T_0-1} \end{bmatrix}.$$

So:

$$\begin{aligned} Y_1^{pre} - \sum_{j=2}^{J+1} w_j Y_j^{pre} \\ &= A\left[Y_{1_{(-1)}}^{pre} - \sum_{j=2}^{J+1} w_j Y_{j_{(-1)}}^{pre}\right] + \left[\left(Z_{1_{(-1)}}^{pre} - \sum_{j=2}^{J+1} w_j Z_{j_{(-1)}}^{pre}\right)\right]' A_1'\theta \\ &+ \iota_{T_0}\left(X_1 - \sum_{j=2}^{J+1} w_j X_j\right)' \gamma + \iota_{T_0}\left[\left(\mu_{z1} - \sum_{j=2}^{J+1} w_j \mu_{zj}\right)\right]' \theta + \iota_{T_0}\left(\mu_1 - \sum_{j=2}^{J+1} w_j \mu_j\right) \\ &+ \left(\nu_1^{pre} - \sum_{j=2}^{J+1} w_j \nu_j^{pre}\right)' \theta + \left(\epsilon_1^{pre} - \sum_{j=2}^{J+1} w_j \epsilon_j^{pre}\right) \end{aligned} \tag{7}$$

Note that by construction $\left(Z_{1_{(-1)}}^{pre} - \sum_{j=2}^{J+1} w_j Z_{j_{(-1)}}^{pre}\right)'$ is a $(T_0 \times K)$ matrix. ι_{T_0} is a vector of ones of length T_0. Since $\left(X_1 - \sum_{j=2}^{J+1} w_j X_j\right)'\gamma$ is a scalar, $\iota_{T_0}\left(X_1 - \sum_{j=2}^{J+1} w_j X_j\right)'\gamma$ is a $(T_0 \times 1)$ vector with identical elements equal to $\left(X_1 - \sum_{j=2}^{J+1} w_j X_j\right)'\gamma$.

We multiply both sides of Equation 7 by $(\iota_{T_0}'\iota_{T_0})^{-1}\iota_{T_0}'$, and we add and subtract the resulting expression from Equation 6. So, we can write:

$$
\begin{aligned}
y_{1T_0+1}^{N} \sum_{j=2}^{J+1} w_j y_{jT_0+1}^{N} = & \left[\left(\iota_{T_0}' \iota_{T_0}\right)^{-1} \iota_{T_0}' \left(Y_1^{pre} - \sum_{j=2}^{J+1} w_j Y_j^{pre} \right) \right] \\
& + A \left[\left(y_{1T_0} - \sum_{j=2}^{J+1} w_j y_{jT_0} \right) - \left(\iota_{T_0}' \iota_{T_0}\right)^{-1} \iota_{T_0}' \left(Y_{1_{(-1)}}^{pre} - \sum_{j=2}^{J+1} w_j Y_{j_{(-1)}}^{pre} \right) \right] \\
& + \left[\left(Z_{1T_0} - \sum_{j=2}^{J+1} w_j Z_{jT_0} \right)' A_1' \theta - \left(\iota_{T_0}' \iota_{T_0}\right)^{-1} \iota_{T_0}' \left(Z_{1_{(-1)}}^{pre} - \sum_{j=2}^{J+1} w_j Z_{j_{(-1)}}^{pre} \right)' A_1' \theta \right] \\
& + \left[\left(X_1 - \sum_{j=2}^{J+1} w_j X_j \right)' \gamma - \left(\iota_{T_0}' \iota_{T_0}\right)^{-1} \iota_{T_0}' \iota_{T_0} \left(X_1 - \sum_{j=2}^{J+1} w_j X_j \right)' \gamma \right] \\
& + \left[\left(\mu_{z1} - \sum_{j=2}^{J+1} w_j \mu_{zj} \right)' \theta - \left(\iota_{T_0}' \iota_{T_0}\right)^{-1} \iota_{T_0}' \iota_{T_0} \left(\mu_{z1} - \sum_{j=2}^{J+1} w_j \mu_{zj} \right)' \theta \right] \\
& + \left[\left(\mu_1 - \sum_{j=2}^{J+1} w_j \mu_j \right) - \left(\iota_{T_0}' \iota_{T_0}\right)^{-1} \iota_{T_0}' \iota_{T_0} \left(\mu_1 - \sum_{j=2}^{J+1} w_j \mu_j \right) \right] \\
& + \left[\left[\sum_{j=2}^{J+1} w_j \left(\nu_{1T_0+1} - \nu_{jT_0+1} \right) - \left(\iota_{T_0}' \iota_{T_0}\right)^{-1} \iota_{T_0}' \left(\nu_1^{pre} - \sum_{j=2}^{J+1} w_j \nu_j^{pre} \right) \right]' \theta \right] \\
& + \left[\sum_{j=2}^{J+1} w_j \left(\epsilon_{1T_0+1} - \epsilon_{jT_0+1} \right) - \left(\iota_{T_0}' \iota_{T_0}\right)^{-1} \iota_{T_0}' \left(\epsilon_1^{pre} - \sum_{j=2}^{J+1} w_j \epsilon_j^{pre} \right) \right]
\end{aligned}
\tag{8}
$$

Note that in this equation we cancel out $\left(\mu_1 - \sum_{j=2}^{J+1} w_j \mu_j\right)$, $\left(\mu_{z1} - \sum_{j=2}^{J+1} w_j \mu_{zj}\right)$, and $\left(X_1 - \sum_{j=2}^{J+1} w_j X_j\right)' \gamma$. Noting that $\left(\iota_{T_0}' \iota_{T_0}\right)^{-1} = \frac{1}{T_0}$, we may rewrite Equation 8 as:

$$
\begin{aligned}
y_{1T_0+1}^{N} \sum_{j=2}^{J+1} w_j y_{jT_0+1}^{N} \\
&= \frac{1}{T_0}\sum_{t=1}^{T_0}\left(y_{it}^{pre} - \sum_{j=2}^{J+1} w_j y_{jt}^{pre}\right) \\
&+ A\left[\left(y_{1T_0} - \sum_{j=2}^{J+1} w_j y_{jT_0}\right) - \frac{1}{T_0}\sum_{t=1}^{T_0}\left(y_{1t_{(-1)}}^{pre} - \sum_{j=2}^{J+1} w_j y_{jt_{(-1)}}^{pre}\right)\right] \\
&+ \left[\left(Z_{1T_0} - \sum_{j=2}^{J+1} w_j Z_{jT_0}\right)' A_1'\theta - \frac{1}{T_0}\sum_{t=1}^{T_0}\left(\left(Z_{1t_{(-1)}}^{pre} - \sum_{j=2}^{J+1} w_j Z_{jt_{(-1)}}^{pre}\right)' A_1'\theta\right)\right] \\
&+ \left[\sum_{j=2}^{J+1} w_j\left(\nu_{1T_0+1} - \nu_{jT_0+1}\right) - \frac{1}{T_0}\sum_{t=1}^{T_0}\left(\nu_{1t}^{pre} - \sum_{j=2}^{J+1} w_j \nu_{jt}^{pre}\right)\right] \\
&+ \left[\sum_{j=2}^{J+1} w_j\left(\epsilon_{1T_0+1} - \epsilon_{jT_0+1}\right) - \frac{1}{T_0}\sum_{t=1}^{T_0}\left(\epsilon_{1t}^{pre} - \sum_{j=2}^{J+1} w_j \epsilon_{jt}^{pre}\right)\right]
\end{aligned} \tag{9}
$$

Now suppose we find the synthetic control method weights $W^* = (w_2^*, \dots, w_{j+1}^*)$ such that for each $t \in (1, \dots, T_o)$ we get $\sum_{j=2}^{J+1} w_j^* y_{jt} = y_{1t}$ and $\sum_{j=2}^{J+1} w_j^* Z_{jt} = Z_{1t}$. Then $\frac{1}{T_0}\sum_{t=1}^{T_0}(y_{1t}^{pre} - \sum_{j=2}^{J+1} w_j^* y_{jt}^{pre})$, which is the average of total deviation between observed values of outcome for the first unit ($i = 1$) and the synthetic control unit, during the pre-intervention period, is equal to zero, and so is $\frac{1}{T_0}\sum_{t=1}^{T_0}(y_{1t_{(-1)}}^{pre} - \sum_{j=2}^{J+1} w_j^* y_{jt_{(-1)}}^{pre})$, and so is $\frac{1}{T_0}\sum_{t=1}^{T_0}(Z_{1t_{(-1)}}^{pre} - \sum_{j=2}^{J+1} w_j^* Z_{jt_{(-1)}}^{pre})$. For the last period before the intervention (T_0), we also have $y_{1T_0} - \sum_{j=2}^{J+1} w_j^* y_{jT_0} = 0$ and also $Z_{1T_0} - \sum_{j=2}^{J+1} w_j^* Z_{jT_0} = 0$. The rest of the terms are idiosyncratic error terms with mean zero. Therefore, eventually we can show for $t = T_0 + 1$:

$$
E\left[y_{1T_0+1}^{N} - \sum_{j=2}^{J+1} w_j^* y_{jT_0+1}^{N}\right] = 0 \tag{10}
$$

Period $T_0 + 2$:

Let us restate the dynamic panel data model below:

$$
y_{it}^{N} = \rho y_{it-1}^{N} + Z_{it}'\theta + X_i'\gamma + \mu_i + \delta_t + \epsilon_{it} \tag{11}
$$

$$
Z_{it} = A_1 Z_{it-1} + A_2 y_{it-1}^{N} + \mu_{zi} + \delta_{zt} + \nu_{it} \tag{12}
$$

and let us go one period back:

$$y_{it-1}^{N} = \rho y_{it-2}^{N} + Z_{it-1}'\theta + X_i'\gamma + \mu_i + \delta_{t-1} + \epsilon_{it-1} \tag{13}$$

$$Z_{it-1} = A_1 Z_{it-2} + A_2 y_{it-2}^{N} + \mu_{zi} + \delta_{zt-1} + \nu_{it-1} \tag{14}$$

So, by plugging in $Z_{it}$from Equation 12 we can write Equation 11 as:

$$y_{it}^{N} = \rho y_{it-1}^{N} + [A_1 Z_{it-1} + A_2 y_{it-1}^{N} + \mu_{zi} + \delta_{zt} + \nu_{it}]'\theta + X_i'\gamma + \mu_i + \delta_t + \epsilon_{it} \tag{15}$$

Now by inserting Equation 13 for y_{it-1} and Equation 14 for Z_{it-1} and rearranging the terms we have:

$$\begin{aligned} y_{it}^{N} = {} & G y_{it-2}^{N} + Z_{it-2}'H + \mu_{zi}'I + \delta_{zt-1}'B + \nu_{it-1}'B + X_i'J + K\mu_i + K\delta_{t-1} + K\epsilon_{it-1} + \delta_{zt}'\theta \\ & + \nu_{it}'\theta + \delta_t + \epsilon_{it} \end{aligned} \tag{16}$$

where $A = [(\rho + A_2'\theta)\rho]$ and $G = A + A_2'B$ are scalars. $H = A_1'B$ is a $(K \times 1)$ vector, μ_{zi}' is a $(1 \times K)$ row vector, $\mathrm{I} = B + \theta$ is a $(K \times 1)$ vector, and $B = [[\rho + A_2'\theta](\theta) + A_1'\theta]$ is a $(K \times 1)$ vector. $\mathrm{J} = C + \gamma$ where γ is a $(K_2 \times 1)$ vector of coefficients on the X_i and $C = \gamma[\rho + A_2'\theta]$ is a $(K_2 \times 1)$ vector as well, $K = [\rho + A_2'\theta]$ is a scalar and θ is a $(K \times 1)$ vector.

$$\begin{aligned} \sum_{j=2}^{J+1} w_j y_{jt}^{N} = {} & G \sum_{j=2}^{J+1} w_j y_{jt-2}^{N} + \sum_{j=2}^{J+1} w_j Z_{jt-2}' H \\ & + \sum_{j=2}^{J+1} w_j X_j' J \\ & + \sum_{j=2}^{J+1} w_j \mu_{zj}' I \\ & + K \sum_{j=2}^{J+1} w_j \mu_j + \delta_{zt-1}'B + \delta_{zt}'\theta + K\delta_{t-1} + \delta_t \\ & + \sum_{j=2}^{J+1} w_j \nu_{jt-1}' B + \sum_{j=2}^{J+1} w_j \nu_{jt}'\theta + K \sum_{j=2}^{J+1} w_j \epsilon_{jt-1} + \sum_{j=2}^{J+1} w_j \epsilon_{jt} \end{aligned} \tag{17}$$

Now we can write:

$$
\begin{aligned}
y_{1t}^{N} - \sum_{j=2}^{J+1} w_j\, y_{jt}^{N}
&= G\left(y_{1t-2}^{N} - \sum_{j=2}^{J+1} w_j\, y_{jt-2}^{N}\right) + \left(Z_{1t-2} - \sum_{j=2}^{J+1} w_j\, Z_{jt-2}\right)' H \\
&+ \left(X_1 - \sum_{j=2}^{J+1} w_j\, X_j\right)' J + \left(\mu_{z1} - \sum_{j=2}^{J+1} w_j\, \mu_{zj}\right)' I + K\left(\mu_1 - \sum_{j=2}^{J+1} w_j\, \mu_j\right) \\
&+ \sum_{j=2}^{J+1} w_j\, (\nu_{1t-1} - \nu_{jt-1})' B + \sum_{j=2}^{J+1} w_j\, (\nu_{1t} - \nu_{jt})' \theta \\
&+ K\left(\sum_{j=2}^{J+1} w_j\, (\epsilon_{1t-1} - \epsilon_{jt-1})\right) + \sum_{j=2}^{J+1} w_j\, (\epsilon_{1t} - \epsilon_{jt})
\end{aligned}
\tag{18}
$$

So, if we stack the pre-intervention matrices we have:

$$
\begin{aligned}
Y_1^{pre} - \sum_{j=2}^{J+1} w_j\, Y_j^{pre}
&= G\left[Y_{1_{(-2)}}^{pre} - \sum_{j=2}^{J+1} w_j\, Y_{j_{(-2)}}^{pre}\right] + \left[\left(Z_{1_{(-2)}}^{pre} - \sum_{j=2}^{J+1} w_j\, Z_{j_{(-2)}}^{pre}\right)\right]' H \\
&+ \iota_{T_0}\left(X_1 - \sum_{j=2}^{J+1} w_j\, X_j\right)' J + \iota_{T_0}\left(\mu_{z1} - \sum_{j=2}^{J+1} w_j\, \mu_{zj}\right)' I + \iota_{T_0} K\left(\mu_1 - \sum_{j=2}^{J+1} w_j\, \mu_j\right) \\
&+ \left[\left(\nu_{1_{(-1)}}^{pre} - \sum_{j=2}^{J+1} w_j\, \nu_{j_{(-1)}}^{pre}\right)\right]' B + \left[\left(\nu_1^{pre} - \sum_{j=2}^{J+1} w_j\, \nu_j^{pre}\right)\right]' \theta \\
&+ K\left(\epsilon_{1_{(-1)}}^{pre} - \sum_{j=2}^{J+1} w_j\, \epsilon_{j_{(-1)}}^{pre}\right) + \left(\epsilon_1^{pre} - \sum_{j=2}^{J+1} w_j\, \epsilon_j^{pre}\right)
\end{aligned}
\tag{19}
$$

We multiply both sides of Equation 19 by $(\iota'_{T_0}\iota_{T_0})^{-1}\iota'_{T_0}$, and we add and subtract the resulting expression from Equation 18. So, we can write:

$$
\begin{aligned}
y^N_{1T_0+2} - \sum_{j=2}^{J+1} w_j\, y^N_{jT_0+2}
&= \left[(\iota'_{T_0}\iota_{T_0})^{-1}\iota'_{T_0}\left(Y_1^{pre} - \sum_{j=2}^{J+1} w_j\, Y_j^{pre}\right)\right] \\
&+ \left[\left(y_{1T_0} - \sum_{j=2}^{J+1} w_j\, y_{jT_0}\right)' - (\iota'_{T_0}\iota_{T_0})^{-1}\iota'_{T_0}\left(y^{pre}_{1_{(-1)}} - \sum_{j=2}^{J+1} w_j\, y^{pre}_{j_{(-1)}}\right)'\right] G \\
&+ \left[\left(Z_{1T_0} - \sum_{j=2}^{J+1} w_j\, Z_{jT_0}\right)' - (\iota'_{T_0}\iota_{T_0})^{-1}\iota'_{T_0}\left(\left(Z^{pre}_{1_{(-1)}} - \sum_{j=2}^{J+1} w_j\, Z^{pre}_{j_{(-1)}}\right)'\right)\right] H \\
&+ \left[\left(X_1 - \sum_{j=2}^{J+1} w_j\, X_j\right)' - (\iota'_{T_0}\iota_{T_0})^{-1}\iota'_{T_0}\iota_{T_0}\left(X_1 - \sum_{j=2}^{J+1} w_j\, X_j\right)'\right] J \\
&+ \left[\left(\mu_{z1} - \sum_{j=2}^{J+1} w_j\, \mu_{zj}\right)' - (\iota'_{T_0}\iota_{T_0})^{-1}\iota'_{T_0}\iota_{T_0}\left(\mu_{z1} - \sum_{j=2}^{J+1} w_j\, \mu_{zj}\right)'\right] I \\
&+ \left[\left(\mu_1 - \sum_{j=2}^{J+1} w_j\, \mu_j\right) - (\iota'_{T_0}\iota_{T_0})^{-1}\iota'_{T_0}\iota_{T_0}\left(\mu_1 - \sum_{j=2}^{J+1} w_j\, \mu_j\right)\right] K \\
&+ \left[\sum_{j=2}^{J+1} w_j\left(\nu_{1T_0+1} - \nu_{jT_0+1}\right)' - (\iota'_{T_0}\iota_{T_0})^{-1}\iota'_{T_0}\left(\nu^{pre}_{1_{(-1)}} - \sum_{j=2}^{J+1} w_j\, \nu^{pre}_{j_{(-1)}}\right)'\right] B \\
&+ \left[\sum_{j=2}^{J+1} w_j\left(\nu_{1T_0+2} - \nu_{jT_0+2}\right) - (\iota'_{T_0}\iota_{T_0})^{-1}\iota'_{T_0}\left(\nu^{pre}_1 - \sum_{j=2}^{J+1} w_j\, \nu^{pre}_j\right)\right]' \theta \\
&+ \left[\sum_{j=2}^{J+1} w_j\left(\epsilon_{1T_0+1} - \epsilon_{jT_0+1}\right) - (\iota'_{T_0}\iota_{T_0})^{-1}\iota'_{T_0}\left(\epsilon^{pre}_{1_{(-1)}} - \sum_{j=2}^{J+1} w_j\, \epsilon^{pre}_{j_{(-1)}}\right)\right] K \\
&+ \left[\sum_{j=2}^{J+1} w_j\left(\epsilon_{1T_0+2} - \epsilon_{jT_0+2}\right) - (\iota'_{T_0}\iota_{T_0})^{-1}\iota'_{T_0}\left(\epsilon^{pre}_1 - \sum_{j=2}^{J+1} w_j\, \epsilon^{pre}_j\right)\right]
\end{aligned}
\tag{20}
$$

Finally, we can write:

$$
\begin{aligned}
y^N_{1T_0+2} - \sum_{j=2}^{J+1} w_j\, y^N_{jT_0+2} \\
&= \frac{1}{T_0}\sum_{t=1}^{T_0}\left(y^{pre}_{1t} - \sum_{j=2}^{J+1} w_j\, y^{pre}_{jt}\right) \\
&+ \left[\left(y_{1T_0} - \sum_{j=2}^{J+1} w_j y_{jT_0}\right)' G - \frac{1}{T_0}\sum_{t=1}^{T_0}\left(y^{pre}_{1t_{(-2)}} - \sum_{j=2}^{J+1} w_j\, y^{pre}_{jt_{(-2)}}\right)' G\right] \\
&+ \left[\left(Z_{1T_0} - \sum_{j=2}^{J+1} w_j Z_{jT_0}\right)' H - \frac{1}{T_0}\sum_{t=1}^{T_0}\left(\left(Z^{pre}_{1t_{(-2)}} - \sum_{j=2}^{J+1} w_j\, Z^{pre}_{jt_{(-2)}}\right)' H\right)\right] \\
&+ \left[\sum_{j=2}^{J+1} w_j \left(\nu_{1T_0+1} - \nu_{jT_0+1}\right)' - \frac{1}{T_0}\sum_{t=1}^{T_0}\left(\nu^{pre}_{1_{(-1)}} - \sum_{j=2}^{J+1} w_j\, \nu^{pre}_{j_{(-1)}}\right)'\right] B \\
&+ \left[\sum_{j=2}^{J+1} w_j \left(\nu_{1T_0+2} - \nu_{jT_0+2}\right) - \frac{1}{T_0}\sum_{t=1}^{T_0}\left(\nu^{pre}_{1} - \sum_{j=2}^{J+1} w_j\, \nu^{pre}_{j}\right)\right]' \theta \\
&+ \left[\sum_{j=2}^{J+1} w_j \left(\epsilon_{1T_0+1} - \epsilon_{jT_0+1}\right) - \frac{1}{T_0}\sum_{t=1}^{T_0}\left(\epsilon^{pre}_{1_{(-1)}} - \sum_{j=2}^{J+1} w_j\, \epsilon^{pre}_{j_{(-1)}}\right)\right] K \\
&+ \left[\sum_{j=2}^{J+1} w_j \left(\epsilon_{1T_0+2} - \epsilon_{jT_0+2}\right) - \frac{1}{T_0}\sum_{t=1}^{T_0}\left(\epsilon^{pre}_{1} - \sum_{j=2}^{J+1} w_j\, \epsilon^{pre}_{j}\right)\right]
\end{aligned}
\tag{21}
$$

Now again consider the synthetic control method weights $W^* = (w_2^*, \dots, w_{j+1}^*)$ such that for each $t \in (1, \dots, T_0)$ we get $\sum_{j=2}^{J+1} w_j^* y_{jt} = y_{1t}$ and $\sum_{j=2}^{J+1} w_j^* Z_{jt} = Z_{1t}$. Then, $\frac{1}{T_0}\sum_{t=1}^{T_0}(y^{pre}_{1t} - \sum_{j=2}^{J+1} w_j^* y^{pre}_{jt})$ which is the average of total deviation between observed values of outcome for the first unit ($i = 1$) and the synthetic control unit, during the pre-intervention period, is equal to zero, and so is $\frac{1}{T_0}\sum_{t=1}^{T_0}(y^{pre}_{1t_{(-2)}} - \sum_{j=2}^{J+1} w_j^* y^{pre}_{jt_{(-2)}})$, and so is $\frac{1}{T_0}\sum_{t=1}^{T_0}(Z^{pre}_{1t_{(-2)}} - \sum_{j=2}^{J+1} w_j^* Z^{pre}_{jt_{(-2)}})$.

For the last period before the intervention (T_0) we also have $y_{1T_0} - \sum_{j=2}^{J+1} w_j^* y_{jT_0} = 0$ and also $Z_{1T_0} - \sum_{j=2}^{J+1} w_j^* Z_{jT_0} = 0$. The rest of the terms are idiosyncratic error terms with mean zero. Therefore, eventually we can show for $t = T_0 + 2$:

$$
E\left[y^N_{1T_0+2} - \sum_{j=2}^{J+1} w_j^* y^N_{jT_0+2}\right] = 0
$$

This can be done with all the post-intervention periods. Therefore, this shows the bias of the synthetic control estimator, given the assumptions mentioned above, is zero in a DPD data generating process such as Equation 1.

3. Conclusion

In comparative case studies we need similarities in the covarites between the treated unit and the controls. The synthetic control method creates a control unit which resembles the values of outcome variable as well as predictors, almost perfectly, while a traditional regression creates predicted unit using the data and the defined model. The underlying model or data generating process of the synthetic control estimator is assumed to be a factor model with time variant coefficients and time invariant variables and we find this model to be restrictive. In this paper, in order to expand the future studies regarding the synthetic control method and to compare this method with other estimation methods, we look at the underlying model of the synthetic control method and we show if the data generating process of the outcome variable of interest is assumed to be a dynamic panel data model in which the variables can be time variant, the synthetic control estimator would provide an unbiased estimator. We believe this may be a more appropriate model to work with in aggregate comparative studies.

Acknowledgment
I gratefully thank Professor Wim Vijverberg for providing his expert advice and immensely valuable comments and suggestions.

References

Abadie , A., Diamond, A. & Hainmueller, J., 2010. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *American Statistical Association*, pp. 493-505.