Panel data approach vs synthetic control method

WAN, Shui Ki; Xie, Yimeng; Hsiao, Cheng

Published in:
Economics Letters

DOI:
10.1016/j.econlet.2018.01.019

Published: 01/03/2018

Document Version:
Peer reviewed version

Panel Data Approach vs Synthetic Control Method

Shui-Ki, Wan[a], Yimeng Xie[b], Cheng Hsiao[b,c]

[a]*Hong Kong Baptist University*
[b]*University of Southern California*
[c]*Xiamen University*

Abstract

We explore the pros and cons of two counterfactual analysis methodologies – the panel data approach (PDA) and the synthetic control method (SCM) through a discussion of the differences in their underlying assumptions and a series of simulations.

Keywords: Counterfactual analysis; Panel data approach; Synthetic control

JEL Codes: C3, C5

1. Introduction

Let Y^1 and Y^0 denote the potential outcome with and without treatment. The treatment effect is then $(Y^1 - Y^0)$. Since Y^1 and Y^0 are not observed simultaneously, the major challenge in the treatment literature is to construct counterfactual for the missing Y^1 or Y^0 (e.g. Heckman and Vytlacil, 2007). SCM (Abadie *et al.*, 2010) and PDA (Hsiao *et al.*, 2012) are the two measurement-without-theory approaches that exploit the correlations among cross-sectional units to construct counterfactuals. Gardeazabel and Vega-Bayo (2016) (GV) compare the two approaches via "real data" cases and simulations. They conclude that: "(i) SCM generates a smaller post-treatment mean square error (MSE), mean absolute percentage error (MAPE), and mean error (ME) with a smaller interquartile range *whenever there is a good enough match*, (ii) SCM is better than PDA when more preintervention periods and covariates are available, (iii) SCM's estimation seems to be more robust to changes in the donor pool than the PDA[1]". However, all statistical methods are based on some maintained hypotheses. No method is universally applicable. Through contrasting the differences in their underlying assumptions and a series of simulations, we show that PDA appears to be more applicable in a wide array of situations.

[1] GV provide the support of (iii) via the empirical studies between Hong Kong and mainland China and conclude that contrary to Hsiao *et al.* (2012), there is a significant political integration effects based on the Abadie *et al.*'s (2010) placebo test using only a subset of Hsiao *et al.*'s original 24 control units. Had we conducted the test on all the 24 control units there is no significant political integration effect.

In Section 2 we highlight the differences in their underlying assumptions. Since there are ways to examine which method is more likely to yield measurements that are more stable and closer to the unknown "truth" only when the true data generating process (DGP) is known, we conduct simulation experiments in a wide array of situations and summarize our main findings in Section 3. Concluding remarks are in Section 4.

2. SCM and PDA

Suppose there are $J+1$ units. We assume an intervention occurs during the period from T_0+1 to T which affect the first unit but not the remaining J units. Thus, the measurement of the treatment effect for the first unit is essentially an issue of how best to predict $Y_{1,t}^0$ for $t = T_0+1, \dots, T$.

SCM assumes that in addition to $Y_{j,t}$, there could be r covariates $Z_{j,t}$ that either drive $Y_{j,t}$ or correlate with $Y_{j,t}$. Let X_1 and X_0 be respectively a $(T_0+r)\times 1$ vector and a $(T_0+r)\times J$ matrix collecting the preintervention data $(Y_1, \bar{Z}_1)$ and $\left(Y_j, \bar{Z}_j, j = 2, \dots, J+1\right)$ where $\bar{Z}_1$ and $\bar{Z}_j$ denote their time-mean.

The SCM presumably makes the assumption that the distribution of $Y_{1,t}^0$, $f\left(Y_{1,t}^0\right)$, and the distribution of $\bar{Z}_1$, $f(\bar{Z}_1)$, lie in the convex hull of the distributions $Y_{j,t}^0$ and $\bar{Z}_j$, respectively, thus, suggests predicting $Y_{1,t}^0$ by

$$\hat{Y}_{1,t}^{0} = \sum_{j=2}^{J+1} w_j Y_{j,t} \tag{1}$$

where the weights are obtained by minimizing:

$$\sqrt{(X_1 - X_0 w)' V (X_1 - X_0 w)} \tag{2}$$

subject to:

$$Y_{1,t} = \sum_{j=2}^{J+1} w_j Y_{j,t}\,, t = 1, \dots, T_0, \tag{3}$$

$$\bar{z}_{i,1} = \sum_{j=2}^{J+1} w_j \bar{z}_{i,j}\ , i = 1, \dots, r, \tag{4}$$

$$w_j \geq 0\ and \sum_{j=2}^{J+1} w_j = 1 \tag{5}$$

where V is a positive-definite matrix.

PDA assumes (i) $Y_{j,t}, t = 1, \dots, T$ should not be influenced by the treatment; and (ii) includes an intercept in the counterfactual estimation to take account the difference in individual-specific fixed effects between the treated and the control units; and (iii) no restrictions for the weights w_j. Assumptions (i) and (ii) are critical in generating unbiased prediction of counterfactuals (Hsiao *et al.*, 2012; Ferman, 2017) while (iii) could be important for efficient prediction. In a regression framework, PDA takes the form[2]

$$Y_{1,t}^{0} = \beta_0 + \tilde{X}_0 \boldsymbol{\beta} + u_{1t}, t = 1, \dots, T_0. \tag{6}$$

where $\tilde{X}_0$ includes any factors that satisfy assumptions (i) and (ii). To balance the

[2] Although Hsiao *et al.* (2012) only uses Y_j in their empirical applications, it does not preclude the use of Z_1 or Z_j.

in-sample fit and post-sample prediction accuracy, Hsiao *et al.* (2012) suggest using a model selection criterion, say corrected AIC, to choose the subset $\tilde{X}_0$[3]. Based on the estimation results of the selected model, PDA predicts the counterfactual by

$$\hat{Y}_{1,t}^0 = \hat{\beta}_0 + \tilde{X}_0\hat{\boldsymbol{\beta}}, t = T_0 + 1, \dots, T. \quad (7)$$

When the PDA assumptions (i) and (ii) hold for both approaches, the difference between PDA and SCM is that the former is an unconstrained regression while the latter is a constrained regression. When the constraints are valid, SCM is a more efficient method. When the constraints are invalid, SCM could lead to biased prediction of counterfactual. For instance, $f(Y_1^0)$ that lies within the convex hull of $f\left(Y_j^0\right)$ does not imply (3) and (4). Neither is (5) necessary. Consider a simple case

$$Y_{j,t}^0 = a_j + b_j f_t + \varepsilon_{j,t}, \quad (8)$$

where f_t is a common factor driving all $Y_{j,t}$ and $\varepsilon_{j,t}$ is the $j\text{-}th$ unit's idiosyncratic error with mean 0. If a_j and b_j are fixed constants, the unbiased prediction of $Y_{1,t}^0$ based on (1) and (5) requires $a_1 = \sum_{j=2}^{J+1} w_j a_j$ and $b_1 = \sum_{j=2}^{J+1} w_j b_j$. Even there could exist w_j satisfying the condition for b_1, SCM could still be biased as long as $a_1 \neq \sum_{j=2}^{J+1} w_j a_j$. Furthermore, there is also an issue of predictive efficiency under (5).

On the other hand, if $Y_{j,t}^0$ is identically distributed across j and strictly stationary over t, or a_j and b_j are random draws from a common distribution such that

[3] In principle, other selection criteria can be used, for instance, Li and Bell (2017) suggest using LASSO.

$E(a_j) = E(a)$ and $E(b_j) = E(b)$, then the constraint (5) may be satisfied and SCM could be better provided the treatment has no effects on the control units[4].

3. Simulation designs and findings

Since it is obvious that failing to satisfy assumptions (i) and (ii) would yield biased predictions for counterfactual, our simulations focus on how (5) impacts predictions. Apart from GV's designs (Design 1), we consider GV's settings with different parameter distributions (Design 2); with known and/ or irrelevant covariates Z (Design 3); some $Y_{j,t}$ are negatively correlated with $Y_{1,t}$ (Design 4); $Y_{j,t}$ are nonstationary but cointegrated (Design 5), and $Y_{1,t}$ and $Y_{j,t}$ follow a simple factor structure in (8) where (5) is satisfied (Design 6) for all combinations[5] where $(J, T_0) \in (5,10,20,40)$ and evaluate with $T - T_0 = 10$ or 20 observations. We compare their performance by MSE, ME and their variability with 1000 replications. GV suggest using $MAE_{0,SCM} < 0.2|\bar{y}_1^0|$ (MAE-rule) to select the "good matches", where $\bar{y}_1^0$ is the pretreatment mean of $Y_{1,t}^0$. This rule, when applied to SCM, often yields a much smaller number of replications than to PDA[6]. In our view, a fair comparison should be based on equal number of experimental outcomes. That is, we should use either all experimental outcomes (1000-rule), or equal number of "good matches". Therefore, if the MAE-rule

[4] We owe this example to a referee.
[5] We present 5 combinations only. Other cases are in the Supplementary Note.
[6] The GV's "good match" rules often drop out a great number of replications for SCM. For instance, in Design 5, less than 10% of SCM replications satisfy the rule.

selects K "good matches" for SCM, then the same number of good PDA matches are selected by adjusting the critical value c in $MAE_{0,PDA} < c|\bar{y}_1^0|$.

The DGPs are in Table 1. Table 2 presents the MSE under both rules for the 6 designs. In general, we find that PDA significantly dominates SCM except in a few cases where $J = T_0 = 5$ or J/T_0 is large, or for Design 6a where the best predictor is the simple average of the control units. The performance of PDA improves when T_0 increases, but it does not necessarily hold for SCM. The MAE-rule could also remove substantial number of replications for comparison.

4. Concluding Remarks

The construction of the counterfactuals has to be based on how the observed data are generated and on whether one takes a measurement-with-theory or measurement-without-theory approach. The PDA and SCM are essentially the latter. PDA rules out control units that could be affected by the treatment and places no restriction on the predictors while SCM does not place restrictions on the control units but constrains the weights to be non-negative and must add up to one. Our expanded simulation designs show that PDA actually dominates SCM in a majority of cases.

For PDA and SCM to yield reasonable estimates of counterfactuals, the control units must not be affected by the intervention. It could be hard to find a control group

that is invariant to such disruptions. For instance, it is not that easy to find control groups to measure the impact of the Iranian revolution on the Iranian economy[7]. Our simulation results also suggest that "a good match rule" or the ratio of J/T_0 could significantly impact the performance of PDA and SCM. Formal theories that address these issues could be of considerable interest to empirical researchers.

Experimental data and observational data are fundamentally different. Statistical analysis using observational data is based on hypothetical DGP. It is not a proof of the true DGP. We must be humble in reporting our numerical findings. We have to think of as many consequences of the hypothesis as possible, and to the extent possible, verify the consequences they follow. We are still only in the process of groping toward the truth, not discovering the truth.

Acknowledgement:

We would like to thank a referee and M.H. Pesaran and R. Smith for helpful comments. We also wish to acknowledge the financial support of China National Science Foundation grant #71631004.

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*

[7] We owe this example to M.H. Pesaran.

Statistical Association **105**: 493–505.

Ferman B. 2017. Revisiting the synthetic control estimator? Working paper.

Gardeazabal J, Vega-Bayo A. 2016. An empirical comparison between the synthetic control method and Hsiao *et al.*'s panel data approach to program evaluation. *Journal of Applied Econometrics*, 32(5), 983-1002.

Heckman J.J. and Vytlacil E. (2007). Econometric evaluation of social programs. Part I: Causal Models. In: Handbook of Econometrics Vol. 6B edited by Heckman JJ, Leamer E. Amsterdam: North-Holland.

Hsiao C, Ching S, Wan SK. 2012. A panel data approach for program evaluation: Measuring the benefits of political and economic integration of Hong Kong with mainland China. *Journal of Applied Econometrics* **27**(5): 705–740.

Li KT, Bell DR. 2017. Estimation of average treatment effects with panel data: Asymptotic theory and implementation. *Journal of Econometrics* **197**(1): 65–75.

Table 1 Simulation Designs

1	$\begin{cases} Y_{jt}^0 = \sum_{k=1}^{4} \mu_{kj}\lambda_{kt} + \xi_{jt} \\ \lambda_{it} = \eta_i + \lambda_{it-1} + \varepsilon_{it}, i = 1,2 \\ \lambda_{it} = 0.8\lambda_{it-1} + \varepsilon_{it}, i = 3,4 \end{cases}$	$\begin{cases} \mu_{i,j}, \eta_i \sim \chi^2(1), i = 1,2 \\ \mu_{ij} \sim N(1,1), i = 3,4 \\ \lambda_{1,0}, \lambda_{2,0} \sim lrgdp \\ \varepsilon_{kt} \sim N(0,0.5) \\ \xi_{jt} \sim N(0,0.25) \end{cases}$
2	$\begin{gathered} Y_{jt}^0 = \gamma_j + \theta_t z_j + \mu_j \lambda_t + \xi_{jt} \\ \theta_t = \{-2, \dots, 2\} \\ \lambda_t = 0.5\lambda_{t-1} + \varepsilon_t \end{gathered}$	$\begin{cases} \gamma_j \sim N(-1,1) \\ z_j \sim lrgdp \\ \mu_j \sim N(1,1) \\ \varepsilon_t \sim N(0,0.5) \\ \xi_{jt} \sim N(0,0.25) \end{cases}$
3	$\begin{cases} Y_{jt}^0 = \gamma_j + \theta_t z_{1j} + \mu_j \lambda_t + \xi_{jt} \\ \theta_t = \{0.3, \dots, 1.8\} \\ \lambda_t = 0.5\lambda_{t-1} + \varepsilon_t \end{cases}$	$\begin{cases} \gamma_j, z_{1j} \sim lrgdp \\ z_{2j} \sim U[-5,-2] \\ z_{3j} \sim N(0,2) \\ \mu_j \sim N(1,2) \\ \varepsilon_t \sim N(0,0.5) \\ \xi_{jt} \sim N(0,0.25) \end{cases}$
4	$\begin{cases} Y_{jt}^0 = \gamma_j + \theta_t z_j + \mu_j \lambda_t + \xi_{jt} \\ \theta_t = \{0.3, \dots, 1.8\} \\ \lambda_t = \eta + \lambda_{t-1} + \varepsilon_t \end{cases}$	$\begin{cases} \gamma_j, z_j \sim lrgdp \\ \mu_{1j} \sim N(1,1), \mu_{1,1} = -\mu_{1,J+1} \\ \eta \sim \chi^2(1) \\ \lambda_0 \sim lrgdp \\ \varepsilon_t \sim N(0,0.5) \\ \xi_{jt} \sim N(0,0.25) \end{cases}$
5	$\begin{cases} Y_{1t}^0 = \sum_{j=2}^{J+1} Y_{jt}^0 \beta_j + \xi_{1t} \\ Y_{jt}^0 = Y_{j,t-1}^0 + u_{jt}, \ j = 2, \dots, J+1 \\ \beta_j = \{1, \dots, 2\} \end{cases}$	$\begin{cases} u_t \sim N(0,0.5) \\ \xi_{1t} \sim N(0,0.5) \end{cases}$
6	$\begin{cases} Y_{j,t}^0 = \lambda_t + \xi_{j,t} \\ \lambda_t = \eta + \lambda_{t-1} + \varepsilon_t \end{cases}$	$\begin{cases} \xi_{jt} \sim N(0,0.25) \\ \eta \sim \chi^2(1) \\ \varepsilon_t \sim N(0,0.5) \end{cases}$

Note: "$x \sim lrgdp$" means that x is randomly drawn from a cross-sectional pool $\ln(rgdp_{j,1980}),\ j = 1, \dots, 143$ obtained from the Penn World Table.

Table 2 MSE of SCM and PDA

		1000-Rule						***MAE*-Rule**					
(J, T_0)		**1a**	**2a**	**3a**	**4a**	**5a**	**6a**	**1a**	**2a**	**3a**	**4a**	**5a**[###]	**6a**
(5,5)	PDA	308	2.63	4.22	3.31	23.17	0.91	2.36	2.47	2.41	2.8	12.72	0.77
	SCM	22269.42	2.94	2.78	124.44	41.85	0.1	2.67	2.86	2.04	10.38	10.33	0.1
(5,20)	PDA	2.44	0.22	0.22	0.23	1.17	0.14	0.38	0.19	0.18	0.22	0.21	0.14
	SCM	1776.59	2.78	2.47	579.39	58.75	0.09	3.33	2.05	1.51	12.22	5.18	0.09
(10,20)	PDA	1.11	0.19	0.18	0.21	14.63	0.15	0.34	0.18	0.18	0.2	-	0.15
	SCM	373.3	1.28	1.29	129.78[#]	151.54[#]	0.08	3.16	1.15	0.97	22.81	-	0.08
(20,20)	PDA	0.82	0.26	0.27	0.3	83.06	0.23	0.45	0.26	0.28	0.31	-	0.24
	SCM	2056.63	0.74[#]	0.75[#]	68.25[#]	338.55[#]	0.08	0.67	0.67	0.57	4.28	-	0.08
(5,40)[##]	PDA	3.00	0.15	0.17	0.16	0.52	0.11	0.5	0.12	0.13	0.14	-	0.11
	SCM	5809.77	2.84	2.62	4563.83	103.85	0.08	7.67	2.18	1.45	38.49	-	0.08

[#] The MSEs are computed based on 992 to 999 replications only as SCM does not necessarily converge.
[##] $T - T_0 = 20$ observations are used.
[###] "-" means that 0 replication satisfies the MAE-rule.