Outside the Box: Using Synthetic Control Methods as a Forecasting Technique

Stefan Klößner (Saarland University)
Gregor Pfeifer (University of Hohenheim)*

October 6, 2016

Abstract

We introduce synthetic control methods (SCM) as a forecasting technique. Using i) as economic predictors solely the outcome itself, i.e., lagged values of the dependent variable, and ii) lagged time series of the outcome to build the donor pool, we let SCM choose and weight appropriate values in order to come up with a sensible forecast of U.S. GDP growth. This procedure performs competitively viable compared with alternative forecasting methods.

Keywords: Synthetic Control Methods; Forecasting; GDP Growth
JEL-Codes: C53; C22

*Klößner: Saarland University, Statistics & Econometrics, Bldg. C3 1, D-66123 Saarbrücken, Germany; s.kloessner@mx.uni-saarland.de; Phone 0049 6831 302 3179. Pfeifer (corresponding author): University of Hohenheim, Department of Economics, 520 B, D-70593 Stuttgart, Germany; g.pfeifer@uni-hohenheim.de; Phone 0049 711 459 22193. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

1 Introduction

In their seminal papers, Abadie and Gardeazabal (2003) and Abadie et al. (2010) introduce synthetic control methods (SCM). In all applications so far, SCM rests upon the comparison of an outcome between a unit representing the case of interest, e.g., a unit affected by an intervention ("treated"), and otherwise similar but unaffected units ("non-treated") reproducing an accurate counterfactual of the unit of interest in absence of the intervention.[1] This is nothing different than *forecasting* the development of a counterfactual scenario for a variable of interest and then comparing this development to its actual counterpart in order to evaluate a certain treatment.

Now, why shouldn't we think about doing the exact analogon in cases where there is no treatment to evaluate, but where we are interested in plainly forecasting a certain economic measure? In such a case, we would not want to compare the forecast against any value, but to simply *create* this forecast properly. This paper explores how we can make use of SCM to compute such forecasts.

In the usual SCM approach, synthetic control unit and actual unit should resemble each other with respect to pre-treatment values of a) the outcome variable and b) predictors of the outcome. These predictors can consist of variables with predictive power for explaining the dependent variable ("covariates") as well as lagged values of the outcome.

As for general forecasting techniques, it is common to solely use lags of the outcome in order to forecast corresponding future values. This is exactly what we do, simply by applying SCM: we use a) deferred time series of the dependent variable as potential donor units and b) lagged values of the outcome, and no covariates, as potential economic predictors, and let the SCM algorithm endogenously choose optimally weighted lags to create a synthetic, one-period forecast of U.S. GDP growth.

[1] See, for instance, Cavallo et al. (2013) (natural disasters), Kleven et al. (2013) (taxation of athletes), Acemoglu et al. (2016) (political connections), or Gobillon and Magnac (2016) (enterprise zones).

2 Computing a Forecast with SCM

In the SCM approach,[2] the synthetic control unit is created out of a "donor pool" of J control units. The comparability to the treated unit is determined by a set of predictors from T pre-intervention periods: M linear combinations of Y and r (other) covariates with explanatory power for Y. All k predictors (with $k = r + M$) are combined in a $(k \times 1)$ vector X_1 for the treated unit and in a $(k \times J)$ matrix X_0 for all control units.

In one part of the optimization process (the inner optimization), one tries to find a linear combination of the columns of X_0 that represents X_1 best. The distance metric used to measure this difference is: $\|X_1 - X_0W\|_V = \sqrt{(X_1 - X_0W)' V (X_1 - X_0W)}$, where the weights used to construct the synthetic control unit are denoted by the vector W, and the weights of the predictors are given by the non-negative diagonal matrix V. The inner optimization is then, for given predictor weights V, the task of finding non-negative control unit weights W, summing up to unity, such that:

$$\sqrt{(X_1 - X_0W)' V (X_1 - X_0W)} \xrightarrow{W} \min \tag{1}$$

The solution to this problem is denoted by $W^*(V)$, which typically contains many vanishing components as these cannot become negative.

The second part of the optimization (the outer optimization) deals with finding optimal predictor weights. It usually follows a data-driven approach, where V is chosen among all positive definite and diagonal matrices such that the mean squared prediction error (MSPE) of the outcome variable Y is minimized over the pre-intervention periods. To this end, $Y_t^{(j)}$ denotes the value of Y for unit j at time $t = 1, \ldots, T$, where $j = 1$ denotes the treated unit, and $j = 2, \ldots, J + 1$ denote the control units.

Using a discount factor $\beta \leq 1$ to put more weight on more recent observations, the outer

[2]See Abadie and Gardeazabal (2003) and Abadie et al. (2010).

optimization problem looks as follows:

$$\sum_{t=1}^{T} \beta^{T-t} \left(Y_t^{(1)} - \sum_{j=2}^{J+1} W^*(V)_j Y_t^{(j)} \right)^2 \xrightarrow{V} \min . \tag{2}$$

As explained above, in traditional applications of SCM, the treated unit is synthesized by units from a so-called donor pool. For instance, when considering U.S. GDP growth, the variable of interest is 'GDP growth', and for synthesizing the U.S., one would look for GDP data from, e.g., Canada, the U.K., and other countries. In the context of forecasting, however, the aim of this paper is to forecast U.S. GDP growth relying solely on U.S. data. The role of donors is thus taken by lags of the original time series, i.e., time-shifted versions of the series for different lags. More precisely, denoting the original time series' value at time t by Y_t, we denote for lags $1, 2, \ldots$ the correspondingly lagged time series by $Y_t^{(1)} := Y_{t-1}, Y_t^{(2)} := Y_{t-2}, \ldots$, respectively.

With regard to economic predictors, we do not use any covariates ($r = 0$) but consider the following linear combinations of lagged values of Y: the time average, defined as $\frac{1}{T} \sum_{t=1}^{T} Y_t^{(j)}$ for every j, and single values at different points in time—the first value in time, $Y_1^{(j)}$, the 'middle' value, $Y_{\lceil T/2 \rceil}^{(j)}$, and the last, most recent value, $Y_T^{(j)}$.

In our subsequent empirical application, we evaluate the following five SCM specifications: 'average', where $M = 1$ and only the time average is used, 'last', where again $M = 1$ and only the most recent value is used, 'average & last', where $M = 2$ and both the time average and the most recent value are used, 'first, middle, last', where $M = 3$ and the first, middle, and last point in time are used, and 'all', where $M = T$ and every single value in time, i.e., $Y_t^{(j)}$ for $t = 1, \ldots, T$ is used.

3 Empirical Application

Figure 1 depicts the quarterly time series of U.S. real and seasonally adjusted GDP growth from the second quarter of 1947 to the first quarter of 2015,[3] for which we will create forecasts applying SCM. We let discount factors take values in $\{0.8, 0.9, 0.95, 1\}$, while we use lags $1, \ldots, H$ as potential donors, where $H \in \{8, 12, 16, 20, 24, 28, 32, 36, 40\}$ is the maximal lag included. Theoretically, every lag $1, \ldots, H$ may be endogenously chosen to contribute to W^*. However, for reasons explained above, typically only a small number of lags will be attributed positive weight: SCM-based estimators therefore feature a built-in shrinkage behavior.

For given H and β, the five different specifications introduced above will be denoted by $\text{SCM}^{\beta}_{\text{A}}(H)$ for 'average', $\text{SCM}^{\beta}_{\text{L}}(H)$ for 'last', $\text{SCM}^{\beta}_{\text{A,L}}(H)$ for 'average & last', $\text{SCM}^{\beta}_{\text{F,M,L}}(H)$ for 'first, middle, last', and $\text{SCM}^{\beta}_{\text{All}}(H)$ for 'all'.

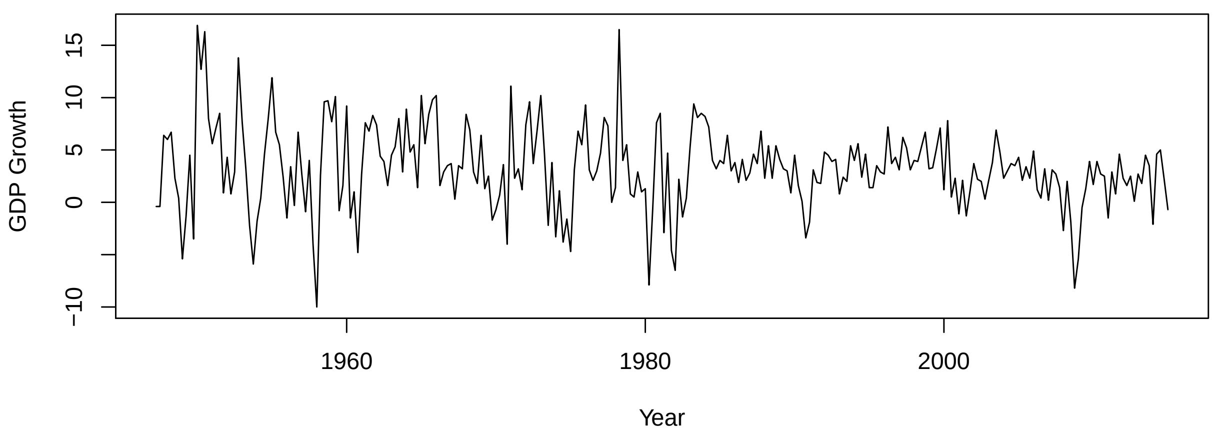

Figure 1: U.S. GDP Growth (%) between 1947 and 2015.

Our corresponding SCM forecasts are compared against the subsequent, well-known forecasting methods: Holt-Winters ('H-W'), Random Walk ('R-W'), and ARMA(p,q) models, where $p, q \in \{0, 1, 2, 3, 4\}$ are chosen by Akaike's and Bayes' Information Criterion ('AIC', 'BIC'), respectively.[4] While we exclusively focus on one-period-ahead forecasts, we study dif-

[3] Available here: `http://research.stlouisfed.org/fred2`; last accessed on August 18, 2016.

[4] All calculations were done using statistical software R (R Core Team (2016)) and package MSCMT (Becker

ferent window lengths T, stepping from three years (12 quarters) to ten years (40 quarters). For the smallest window length of three years, the times for which forecasts are computed range from the second quarter of 1960 to the first quarter of 2015. With increasing window length, the time points for the first forecast move to the right, finally resulting in forecast times for the largest window length of ten years ranging from the second quarter of 1967 to the first quarter of 2015.

Table 1 shows the winners of horse races according to three different criteria: the average rank of an estimator according to the corresponding forecasts' precision, the mean absolute prediction error (MAPE), and the root mean square prediction error (RMSPE). About half of the horse races are won by SCM based forecasts, where the criterion RMSPE favors SCM a little more compared with the other two. Throughout the SCM winners, it becomes obvious that higher β provide more precise estimates. Regarding specifications, 'all' and 'average & last' seem to perform better than the rest, especially better than 'first, middle, last'. This suggests that predictors based on old values seem less important than predictors based on more recent ones. With respect to donors, we observe that particularly the maximal lag number of 36 supports SCM winners, while we still have corresponding winners with, e.g., maximal lags of 8 and 12. Overall, when averaging over all window lengths, 'MAPE' and 'RMSPE' yield SCM winners with a β of 1, an 'average & last' and 'all' specification, respectively, and 36 lags as potential donors.

Table 2 displays values of the criteria, averaged over all window lengths, for all estimators that managed to win at least one horse race (as of Table 1), as well as for the remaining (non-winning) competing estimators—Holt-Winters and Random Walk. Over all criteria, these latter estimators perform much worse than any of the race winners. Among such winners, the 36 lags based SCM estimator with $\beta = 1$ of specification 'average & last', $\mathrm{SCM}^1_{\mathrm{A,L}}(36)$, performs very well with respect to all three criteria and might therefore be considered the overall winner. What is more, out of 36 potential lags, the SCM algorithm attaches weights to only five of these values (on average) to build the synthetic forecast.

and Klößner (2016)).

Table 1: Winners of Different Horse Races

	Ranks	MAPE	RMSPE
Window Length 8	$\mathrm{SCM}^{1}_{\mathrm{All}}(8)$	$\mathrm{SCM}^{0.95}_{\mathrm{All}}(28)$	$\mathrm{SCM}^{1}_{\mathrm{All}}(40)$
Window Length 12	$\mathrm{SCM}^{0.95}_{\mathrm{A,L}}(36)$	$\mathrm{SCM}^{0.95}_{\mathrm{A,L}}(36)$	$\mathrm{SCM}^{1}_{\mathrm{All}}(36)$
Window Length 16	$\mathrm{SCM}^{1}_{\mathrm{A,L}}(8)$	$\mathrm{SCM}^{1}_{\mathrm{A,L}}(12)$	$\mathrm{SCM}^{1}_{\mathrm{A,L}}(12)$
Window Length 20	$\mathrm{SCM}^{0.9}_{\mathrm{A,L}}(36)$	BIC	$\mathrm{SCM}^{1}_{\mathrm{A,L}}(36)$
Window Length 24	BIC	$\mathrm{SCM}^{1}_{\mathrm{A,L}}(36)$	BIC
Window Length 28	BIC	BIC	BIC
Window Length 32	AIC	AIC	AIC
Window Length 36	AIC	BIC	BIC
Window Length 40	BIC	BIC	$\mathrm{SCM}^{1}_{\mathrm{A}}(28)$
Overall	BIC	$\mathrm{SCM}^{1}_{\mathrm{A,L}}(36)$	$\mathrm{SCM}^{1}_{\mathrm{All}}(36)$

Note: Criteria for winning a horse race are: 'Ranks', calculated as the average rank of the estimator among all estimators; 'MAPE', the mean absolute prediction error; and 'RMSPE', the root mean square prediction error.'Overall' denotes the winner w.r.t. the average over all window lengths.

Table 2: Average Values of Different Criteria for Horse Race Winners & Competing Non-Winners

	Ranks	MAPE	RMSPE
$\mathrm{SCM}^{0.95}_{\mathrm{All}}(28)$	92.43	2.823	3.754
$\mathrm{SCM}^{1}_{\mathrm{All}}(8)$	90.27	2.873	3.812
$\mathrm{SCM}^{1}_{\mathrm{All}}(36)$	90.17	2.746	3.646
$\mathrm{SCM}^{1}_{\mathrm{All}}(40)$	90.62	2.758	3.656
$\mathrm{SCM}^{1}_{\mathrm{A}}(28)$	92.45	2.875	3.808
$\mathrm{SCM}^{0.9}_{\mathrm{A,L}}(36)$	88.03	2.787	3.751
$\mathrm{SCM}^{0.95}_{\mathrm{A,L}}(36)$	87.77	2.766	3.721
$\mathrm{SCM}^{1}_{\mathrm{A,L}}(8)$	90.46	2.824	3.855
$\mathrm{SCM}^{1}_{\mathrm{A,L}}(12)$	91.27	2.840	3.786
$\mathrm{SCM}^{1}_{\mathrm{A,L}}(36)$	87.03	2.731	3.693
AIC	87.48	2.825	3.806
BIC	86.19	2.756	3.710
H-W	101.25	3.082	4.086
RW	98.91	2.986	3.969

Note: Values of criteria, averaged over all window lengths, for i) above horizontal line: all estimators winning at least one horse race; ii) below horizontal line: competing, never winning estimators.

Recall that such behavior resembles features of so-called shrinkage estimators, which is why we finally compute corresponding forecasting results with a Lasso-type estimator for fitting autoregressive time series models as discussed in Nardi and Rinaldo (2011). This estimator performs worse than our overall winner, $\text{SCM}^{1}_{\text{A,L}}(36)$, according to all three criteria: the version without cross-validation yields values 87.77 (Ranks), 2.782 (MAPE), and 3.765 (RMSPE), while corresponding values for the version with cross-validation are 89.69, 2.797, and 3.726, respectively. Eventually, over all window lengths and across all criteria, none of the forecasts based on Nardi and Rinaldo (2011) would emerge as a winner in Table 1.

4 Discussion

This paper initiates thinking about SCM as a forecasting technique. We consider outcome lags as economic predictors and create a donor pool out of deferred time series of the dependent variable. Eventually, we let the SCM procedure choose appropriately weighted lags to compute a one-period, synthetic forecast of U.S. quarterly GDP growth. We find that SCM performs competitively well compared with alternative forecasting techniques. Over several window lengths and precision criteria, a certain SCM specification even emerges as the overall horse race winner.

Certainly, one can think about creating the SCM donor pool differently. For instance, we could include properly constructed down- or upward trends. Moreover, one could want to not simply rely on transformations of the dependent variable, but also on covariates with predictive power for the outcome. Eventually, we can also think about re-designing our SCM model with, e.g., a donor pool consisting of other countries' GDP forecasts. In this case, SCM would perform something similar to an error-smoothing over such different predictions.

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *The American Economic Review*, 93(1):113–132.

Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., and Mitton, T. (2016). The value of connections in turbulent times: Evidence from the united states. *Journal of Financial Economics*, 121(2):368 – 391.

Becker, M. and Klößner, S. (2016). *MSCMT: Multivariate Synthetic Control Method Using Time Series*. R package version 1.0.0.

Cavallo, E., Galiani, S., Noy, I., and Pantano, J. (2013). Catastrophic Natural Disasters and Economic Growth. *The Review of Economics and Statistics*, 95(5):1549–1561.

Gobillon, L. and Magnac, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *The Review of Economics and Statistics*, 98(3):535–551.

Kleven, H. J., Landais, C., and Saez, E. (2013). Taxation and international migration of superstars: Evidence from the european football market. *The American Economic Review*, 103(5):1892–1924.

Nardi, Y. and Rinaldo, A. (2011). Autoregressive process modeling via the Lasso procedure. *Journal of Multivariate Analysis*, 102(3):528–549.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.