3 Methodology

Box 3. Methodology Summary

This section describes the synthetic control methodology in more detail. The aim of the methodology is to estimate a counterfactual – in our case this would entail estimating single malt exports to the US if there was no tariff introduced. The difference between this counterfactual and observed exports (with the tariff) would be the estimated tariff impact.

The counterfactual is constructed using a combination of countries which (a) do not have a change in tariff and (b) are comparable to the US in other respects. Many researchers suggest constructing this counterfactual ('synthetic control') using a weighted average, where the weights are chosen in such a way that our synthetic control resembles the US as closely as possible in its characteristics prior to the introduction of the tariff.

3.1 Synthetic Control

This section is adapted from the forthcoming Abadie (2020) paper discussing the synthetic control methodology and feasibility[28], as well as the Firpo and Possebom (2018) paper discussing alternative inference tests for the synthetic control methodology.

3.1.1 The Setting

Assume that we have data for all for $J+1$ units: $j = 1,2,\dots,J+1$, where J is the amount of control units and $j = 1$ is the treated unit (the United States, in our case). The donor pool $j = 2,\dots,J+1$ is a collection of untreated units assumed to be not affected by the intervention.

$j = 1$	$j = 2,\dots,J+1$
Treated (one unit)	Untreated (J units, referred to as the donor pool)

Assume that our data spans T periods, and that the first T_0 periods occur before the intervention. For each unit j and time t we observe the outcome of interest, Y_{jt}. For each unit j we also observe a set of k predictors of the outcome, $X_{1j},\dots,X_{kj}$, which

[28] Abadie, 2021. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects". Journal of Economic Literature (forthcoming issue). Preview (2020) available at: https://economics.mit.edu/files/17847.

are themselves unaffected by the intervention. These predictors may include pre-intervention values of Y_{jt}.

$t = 1, \dots, T_0$	$t = T_0 + 1, \dots, T$
Pre-treatment period (T_0 periods)	Post-treatment period ($T - T_0$)

The $k \times 1$ vectors $\boldsymbol{X}_1, \dots, \boldsymbol{X}_{J+1}$ contain the values of the predictors for units $j = 1, \dots, J+1$. The $k \times J$ matrix, $\boldsymbol{X}_0 = \left[\boldsymbol{X}_2 \cdots \boldsymbol{X}_{J+1}\right]$ collects the values of predictors for all untreated units (a $k \times 26$ matrix, in our case). For each unit j and time period t we can define $\mathrm{Y}_{\mathrm{jt}}^{\mathrm{N}}$ to be the potential response without intervention. For the single treated unit, $\mathrm{j} = 1$, and a post-intervention period $\mathrm{t} > \mathrm{T}_0$ we can define $\mathrm{Y}_{1\mathrm{t}}^{\mathrm{I}}$ to be the potential response with the intervention.

The effect of the intervention of interest for the treated unit in period $t > T_0$ can be written as:

$$\delta_{1t} = Y_{1t}^{I} - Y_{1t}^{N}$$

The challenge here is to estimate Y_{1t}^{N} for $t > T_0$: any outcome Y_{1t} we observe for the treated unit after the introduction of the tariff is by definition Y_{1t}^{I}. Note that the effect of the intervention can change over time (the t subscript is retained). For example, we may hypothesise that the tariff takes a number of months (or quarters) to reach its full impact as US importers work to find suitable substitutes with a lower price point, or set up new trading relations.

The synthetic control method approximates the treated unit by creating a weighted average of units in the donor pool. The synthetic control can be represented by a $J \times 1$ vector of weights, $\boldsymbol{W} = \left(w_2, \dots, w_{J+1}\right)'$. Using this set of weights, the synthetic control estimators of Y_{1t}^{N} and δ_{1t}, respectively, are:

$$\hat{Y}_{1t}^{N} = \sum_{j=2}^{J+1} w_j Y_{jt}$$

and

$$\hat{\delta}_{1t} = Y_{1t} - \hat{Y}_{1t}^{N} \qquad (1)$$

To avoid extrapolation, the weights can be restricted to be non-negative and to sum to one:

$$\sum_{j=2}^{J+1} w_j = 1$$

If we were using nominal export values, this would be an issue – the Unites States is the top export market for Scotch whisky (single malt or otherwise), which means that any weighted average of countries in the donor pool would not be sufficient to approximate the actual values for $t \leq T_0$. Scaling the values, for example by using per-capita values or growth rates, could alleviate this issue.

3.1.2 Choosing weights

Weights in the $J \times 1$ vector $\boldsymbol{W} = \left(w_2, \ldots, w_{J+1}\right)'$ can be chosen by the researcher (e.g. equal weights for a simple average or population-weights). Abadie (2020) proposes to choose $w_2, \ldots, w_{J+1}$ so that the resulting weighted average (synthetic control) not only best resembles the pre-intervention outcome Y, but also the pre-intervention predictors for the treated unit.

Given the non-negative constants $v_1, \ldots, v_k$ (for k predictors), a set of weights $\boldsymbol{W}^* = \left(w_2{}^*, \ldots, w_{J+1}{}^*\right)'$ is chosen that minimises:

$$\|\boldsymbol{X}_1 - \boldsymbol{X}_0\boldsymbol{W}\| = \sqrt{\sum_{h=1}^{k} v_h\left(X_{h1} - w_2 X_{h2} - \cdots - w_{j+1}X_{hJ+1}\right)^2} \qquad (2)$$

such that weights $w_2, \ldots, w_{J+1}$ are non-negative and sum to one. This can be referred to as the 'inner optimisation'. Some predictors will be more 'important' than others in this minimisation exercise. The positive constants $v_1, \ldots, v_k$ therefore reflect the weight placed on each of the k predictors when reproducing the values of the treated unit's predictors using the donor pool's predictor values. The estimated treatment effect for time $t > T_0$ is then:

$$\hat{\delta}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j{}^* Y_{jt}$$

The set of weights $\boldsymbol{W}^*$ which minimises equation (2) needs a given set of constants $v_1, \ldots, v_k$. For each choice of $\boldsymbol{V} = (v_1, \ldots, v_k)$, a different synthetic control is generated:

$$\boldsymbol{W}(\boldsymbol{V}) = \begin{pmatrix} w_2(\boldsymbol{V}) \\ \vdots \\ w_{J+1}(\boldsymbol{V}) \end{pmatrix}$$

Much like choosing the set of weights $\boldsymbol{W}$, the choice of $\boldsymbol{V}$ can be left up to the researcher. For example, Abadie suggests choosing $\boldsymbol{V}$ such that the synthetic

control $\boldsymbol{W}(\boldsymbol{V})$ minimises the mean squared prediction error (for example, by dividing up the pre-treatment sample in training and testing periods like in many other time series applications). Researchers may also simply minimise the mean squared error for the entire pre-treatment period (i.e. minimise the distance between the observed and synthetic outcome values). The process of this optimisation is referred to as the 'outer optimisation'.

3.1.3 Implementation

The process of choosing weights $\boldsymbol{W}^*$ and constants $\boldsymbol{V}$ (outer and inner optimisation, respectively) is handled by Becker and Klößner's R package **`MSCMT`** (multivariate synthetic control method using time series).[29] This implementation is able to handle time-series data, both in the outcome and predictors, as well as multiple outcome variables (if needed). A full description of the various optimisation algorithms and methods used in this package is available in Becker and Klößner (2018).[30]

3.2 Inference

3.2.1 Placebo tests

Abadie et al (2010) propose a benchmark similar to Fisher's Exact Hypothesis Test where they estimate, for each control unit $j = 2, \dots, J+1$ and post-treatment time period $t = T_0 + 1, \dots, T$ an estimate $\hat{\delta}_{jt}$. The distribution of these estimates, $\widehat{\boldsymbol{\delta}}_j = \left(\hat{\delta}_{jT_0+1} \cdots \hat{\delta}_{jT}\right)'$, can then be compared to the vector of estimates for the treated unit, $\widehat{\boldsymbol{\delta}}_1 = \left(\hat{\delta}_{1T_0+1} \cdots \hat{\delta}_{1T}\right)'$.

If the vector of estimated effects for the United States is substantially different in value than the distribution of effects for all control units, Abadie et al reject the null hypothesis of *no effect*.

In some cases, certain time periods $t \in \{T_0 + 1, \dots, T\}$ may show a large effect while others do not. In these cases, it may be unclear whether to reject the null or not. To that end, Abadie et al (2010) propose two potential test statistics: one based on the post-treatment (root) mean squared prediction errors (MSPEs), and one based on the ratio of the (root) MSPEs pre- and post-treatment.

3.2.1.1 Inference using the post-treatment fit

Using the post-treatment RMSPE for country $j = 1, \dots, J+1$:

[29] See Becker and Klößner (2017): https://cran.r-project.org/package=MSCMT

[30] Martin Becker and Stefan Klößner, 2018. "Fast and reliable computation of generalized synthetic controls". Econometrics and Statistics, vol 5, pages 1-19. Preliminary version (2017) available at: http://www.oekonometrie.uni-saarland.de/papers/FastReliable.pdf

$$RMSPE_j^{post} = \sqrt{\frac{\sum_{T_0+1}^{T}\left(Y_{j,t} - \hat{Y}_{j,t}^{N}\right)^2}{T - T_0}}$$

the p-value proposed by Abadie et al (2010) is given by:

$$p = \frac{\sum_{j=1}^{J+1} \mathbb{I}\left[RMSPE_j^{post} \geq RMSPE_1^{post}\right]}{J+1} \quad (3)$$

where the indicator function $\mathbb{I}\left[RMSPE_J \geq RMSPE_1\right]$ takes a value of 1 when $RMSPE_J \geq RMSPE_1$, and 0 otherwise. Alternatively, the $MSPE$ can be used.

Intuitively, this compares the post-treatment fit of the treated unit to the post-treatment fit of the placebo tests, and if it is unusually large compared to the fits obtained by the control units' synthetic controls, the p-value is small.

This, however, requires us to limit the control units used by comparing their pre-treatment fit: one of the control units will ultimately have the largest export value per capita, and its synthetic control will fit poorly. Therefore, its post-treatment fit will also be poor, and its RMSPE will be high. Abadie et al propose imposing a restriction on the ratio between the treated unit's pre-treatment fit and the control units' pre-treatment fits to remedy this.

3.2.1.2 Inference using the ratio of post- to pre-treatment fits

One way to avoid limiting the size of the donor pool is by using the ratio of post- to pre-treatment fits instead of only the post-treatment fit. This is also alluded to in Abadie et al (2010) and is the test statistic of choice in Abadie et al (2015).

Using a ratio of root mean squared prediction errors (RMSPE) given by:

$$RMSPE_j^{ratio} = \frac{RMSPE_j^{post}}{RMSPE_j^{pre}} = \sqrt{\frac{\sum_{T_0+1}^{T}\left(Y_{j,t} - \hat{Y}_{j,t}^{N}\right)^2/(T - T_0)}{\sum_{t=1}^{T_0}\left(Y_{j,t} - \hat{Y}_{j,t}^{N}\right)^2/T_0}} \quad (4)$$

the p-value can be defined as:

$$p = \frac{\sum_{j=1}^{J+1} \mathbb{I}\left[RMSPE_j^{ratio} \geq RMSPE_1^{ratio}\right]}{J+1} \quad (5)$$

where the indicator function $\mathbb{I}\left[RMSPE_j^{ratio} \geq RMSPE_1^{ratio}\right]$ takes a value of 1 when $RMSPE_j^{ratio} \geq RMSPE_1^{ratio}$, and 0 otherwise.[31]

In short, the $RMSPE_j^{ratio}$ test statistic is a ratio of root mean squared prediction errors for country j before and after the intervention. When a large enough proportion of control units have a ratio of pre- to post-treatment RMPSEs larger than the treated unit $j = 1$, the p-value is large (and we fail to reject the null hypothesis of *no effect* at some significance level α).

3.2.1.3 Null hypothesis

The (two-sided) null hypothesis for both of these tests is given by:

$$H_0 : \delta_{jt} = 0 \text{ for each region } j \in \{1, \dots, J+1\} \text{ and time period } t \in \{1, \dots, T\}$$

This is a fairly restrictive inference assumption, as recognised by Ferman, Pinto, and Possebom (2018). In the absence of random assignment (of the treatment), this p-value can be interpreted as the probability of obtaining an estimated value for the test statistic at least as large as the value obtained using the treated case, as if the treatment were randomly assigned among the data (i.e., our control units).

3.2.2 Confidence sets

Ferman, Pinto, and Possebom (2018) extend the inference procedure for the synthetic control method to allow for any sharp hypothesis, where the null hypothesis for a constant treatment effect is given by:

$$H_0^c : Y_{j,t}^I = Y_{j,t}^N + c \times \mathbb{I}[t \geq T_0 + 1]$$

in each region $j \in \{1, \dots, J+1\}$ and time period $t \in \{1, \dots, T\}$, and $c \in \mathbb{R}$. This can be rephrased as:

$$H_0^c : \delta_{jt} = c \times \mathbb{I}[t \geq T_0 + 1]$$

Ferman, Pinto, and Possebom also note that more general treatment effect functions can also be used – e.g., where the treatment effect is not constant over time, or varies by region as well as time.

[31] A ratio of mean squared error predictions can also be used. This would alter the test statistic for each country and simply shift the scale. This is used, for example, in Firpo and Possebom, 2018.

The test statistic seen in equation (4) can be modified to allow for this intervention effect:

$$RMSPE_j^c = \sqrt{\frac{\sum_{T_0+1}^{T}\left(Y_{j,t} - \hat{Y}_{j,t}^N - c \times \mathbb{I}[t \geq T_0]\right)^2/(T - T_0)}{\sum_{t=1}^{T_0}\left(Y_{j,t} - \hat{Y}_{j,t}^N - c \times \mathbb{I}[t \geq T_0]\right)^2/T_0}} \quad (6)$$

while the p-value in equation (5) becomes:

$$p^c = \frac{\sum_{j=1}^{J+1} \mathbb{I}\left[RMSPE_j^c \geq RMSPE_1^c\right]}{J+1} \quad (7)$$

Note that Ferman, Pinto, and Possebom also allow for country-weights in this ratio to vary using some sensitivity parameter $\phi \in \mathbb{R}_+$ and a vector $\boldsymbol{\nu} = \left(\nu_1, \ldots, \nu_{J+1}\right)$. Here, we focus on the case where $\phi = 0$ and $\boldsymbol{\nu} = (1, \ldots, 1)$, extending the equal-weight inference procedure in Abadie et al (2010) to test for any sharp hypothesis.[32]

Inverting the test statistic allows us to estimate confidence sets, where a general $(1 - \alpha)$ confidence set can be constructed as follows:

$$CS_{(1-\alpha)} = \left\{f \in \mathbb{R}^{\{1,\ldots,T\}} : f(t) = c \; and \; p^c > \alpha\right\}$$

This set contains all constant-in-time intervention effects whose associated null hypotheses are not rejected by the inference procedure. In some cases, a one-sided test may be desirable (for example, isolating only negative effects post-treatment).[33]

[32] Sensitivity analysis could be performed by varying ϕ and $\boldsymbol{\nu}$.

[33] A one-sided null hypothesis may be given by $H_0^c : \delta_{j,t} < c$ where a mean prediction error test statistic could be used, $MPE_j^c = \frac{\sum_{T_0+1}^{T}\left(Y_{j,t} - \hat{Y}_{j,t}^N - c \times \mathbb{I}[t \geq T_0]\right)/(T - T_0)}{\sum_{t=1}^{T_0}\left(Y_{j,t} - \hat{Y}_{j,t}^N - c \times \mathbb{I}[t \geq T_0]\right)/T_0}$, where the p-value p^c can be calculated as $p^c = \frac{\sum_{j=1}^{J+1} \mathbb{I}\left[MPE_j^c < MPE_1^c\right]}{J+1}$.