Synthesizing Cash for Clunkers: Stabilizing the Car Market, Hurting the Environment?

Stefan Klößner and Gregor Pfeifer

Saarland University, University of Hohenheim

23 July 2018

Synthesizing Cash for Clunkers: Stabilizing the Car Market, Hurting the Environment?

Stefan Klößner and Gregor Pfeifer*

July 23, 2018

Abstract

We examine the impact of the €5 billion German Cash for Clunkers program on vehicle registrations and respective CO2 emissions. To construct proper counterfactuals, we develop the multivariate synthetic control method using time series of economic predictors (MSCMT) and show (asymptotic) unbiasedness of the corresponding effect estimator under quite general conditions. Using cross-validation for determining an optimal specification of predictors, we do not find significant effects for CO2 emissions, while the stimulus' impact on vehicle sales is strongly positive. Modeling different buyer subgroups, we disentangle this effect: 530,000 purchases were simply windfall gains; 550,000 were pulled forward; and 850,000 vehicles would not have been purchased in absence of the subsidy, worth €17 billion.

JEL Codes: C31; C32; D04; D12; H23; H24; L62; Q51
Keywords: Generalized Synthetic Controls; Cross-Validation; Cash-for-Clunkers; CO2 Emissions

---

*Pfeifer (corresponding author): University of Hohenheim, Department of Economics, 520B, D-70593, Stuttgart, Germany; +4971145922193; g.pfeifer@uni-hohenheim.de. Klößner: Saarland University, Department of Economics, Campus Bldg. C3 1, 66123 Saarbrücken, Germany; s.kloessner@mx.uni-saarland.de. The authors are grateful for valuable remarks from Sandra Baar, Martin Becker, Bernhard Boockmann, Bernd Fitzenberger, Ralph Friedmann, Paul Grieco, Krzysztof Karbownik, Chris Knittel, Eric Leeper, Thierry Magnac, Thierry Mayer, Claudio Michelacci, Aderonke Osikominu, Bettina Peters, Imran Rasul, Morten Ravn, John Van Reenen, and Jeremy West as well as participants of COMPIE, the Econometric Society European Winter Meeting, the Research Seminar of the Centre for European Economic Research (ZEW), Ce^2/WIEM, the Research Colloquium at Jena University, the German Statistical Week, the Annual Conferences of the German Economic Association, RES, EARIE, IAAE, EEA–ESEM, and the Berlin Seminar on Energy and Climate Policy, German Institute for Economic Research (DIW). Lastly, we are very grateful to Patrik Guggenberger for hospitality at Penn State, where parts of this research were conducted. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

1 Introduction

In response to the recent economic downturn, governments around the globe intervened with vehicle retirement schemes worth more than €15 billion in total. European countries contributed almost 60% of this amount, with Germany affording the most expensive program utilizing €5 billion to subsidize two million car purchases. As for most countries, the primary purpose of such programs was twofold: stimulating the declining demand for new vehicles, which threatened not only the local automotive market but the economy as a whole; and reducing road traffic related carbon dioxide (CO2) emissions.

In this paper, we seek to measure whether these goals were accomplished, examining the impact of the German scrappage scheme on purchases of new vehicles and greenhouse gas emissions. To be more precise, we are evaluating effects on new passenger car registrations and how these can be disentangled into windfall gains, pull-forward and delayed purchases, as well as on-top sales. In a next step, we estimate how corresponding CO2 emissions of new passenger cars behave in consequence of the policy intervention and relate those findings to our sales results.

The main challenge regarding an impact analysis of this kind arises from the difficulty to construct a reasonable counterfactual of new car registrations and CO2 emissions in absence of the subsidy. To identify our effects of interest, we use a rich data pool of 12 European countries, Germany with and eleven countries without a scrappage scheme, as well as several covariates to which we apply a generalized and extended version of synthetic control methods (SCM).[1] SCM rest upon the comparison of outcome variables between a unit representing the case of interest, i.e. Germany as affected by the policy intervention, and otherwise similar but unaffected units reproducing an accurate counterfactual. An algorithm-derived combination of precisely weighted comparison units is supposed to depict better the characteristics of Germany than either any single comparison unit alone or an equally weighted combination of all or several available control units. Using an appropriate set of covariates, also known as

[1]SCM, introduced in Abadie and Gardeazabal (2003) and Abadie et al. (2010), were applied, i.a., by Cavallo et al. (2013) (natural disasters), Jinjarak et al. (2013) (capital inflows), Kleven et al. (2013) (taxation of athletes), Acemoglu et al. (2016) (political connections), and Gardeazabal and Vega-Bayo (2017) (political/economic integration).

economic predictors, SCM select the counterfactual unit as the optimally weighted average of such comparison units that best resembles the pre-treatment characteristics of Germany.

The original SCM approach addresses one respective variable of interest at a time. The M dimension of our multivariate synthetic control method using time series (MSCMT), however, allows to consider multiple outcomes simultaneously. Therefore, it takes into account possible interdependencies between the outcomes of interest, new car registrations and CO2 emissions of newly registered cars. The standard SCM approach is also not capable of treating time-varying covariates as separate time series objects. Thus, if one wants to incorporate such economic predictors, one is restricted to either use only aggregate data like their mean, or to artificially treat every value of a time-varying covariate as a separate economic predictor. Both these alternatives have significant drawbacks. The latter is violating the principle of parsimony and comes with practical problems because the number of quantities that need to be estimated grows too large. The former comes at the cost of potentially biased estimates when the variable(s) of interest depend on lagged values of the economic predictor(s). The T dimension of our MSCMT approach overcomes these problems. It uses whole time series of the economic predictors and treats them as a single time series object each, thereby exploiting the predictors' variation over time in order to find the most precise counterfactual for Germany.

In accordance with these extensions, we formally show that under conditions more general than those considered in the original approach by Abadie et al. (2010), a potential bias of the effect estimator is a linear function of the pre-intervention approximation errors, and that in case of unobserved confounders, the effect estimator is still asymptotically unbiased. Eventually, the majority of SCM applications merely operate in-sample, making it difficult to assess the counterfactual's validity. To mitigate this problem and to determine optimal predictor weights throughout a specification search, we incorporate a generalized cross-validation approach: the pre-treatment timespan is divided into a training and a validation period, and predictor weights are selected by minimizing the out-of-sample error in the validation window. Lastly, we also use this technique to show how powerful it is to use the T-dimension of our covariates as compared to simply relying on their means.

Applying MSCMT, we find that the German scrappage program had an immensely positive effect on new car registrations with more than one million program-induced vehicles during its peak time. In order to disentangle program-affected purchases, we build on a model of different sales sub-groups and estimate corresponding shares in the overall effect. We find that 530,000 sales were windfall gains, i.e., subsidized purchases that would also have happened in absence of the policy intervention. In terms of shifting demand in time, the program triggered the pulling forward of 550,000 sales, because buyers wanted to profit from the scrappage premium. Contrarily, 260,000 buyers, which were not eligible for the subsidy, delayed their purchases until the German program had ended. Eventually, the program also induced 850,000 purchases of new cars that would not have happened at all in absence of the German subsidy. In monetary terms, these on-top sales translate into more than three times the €5 billion budget, notably boosting the German car market.

Regarding the second variable of interest, CO2 emissions, our MSCMT estimates yield non-significant effects throughout the program period. Disentangling group-specific impacts, we find that the non-significant result in 2009 is due to two different effects essentially balancing each other. On the one hand, average CO2 emissions of newly registered cars in 2009 were reduced since the program mainly pushed demand towards small, eco-friendly vehicles. On the other hand, the €2,500 subsidy induced many consumers to upgrade their purchases towards larger, more polluting ones. Due to these upgrades, increased emissions over the lifetime of the new vehicles might more than offset the reduction of CO2 emissions resulting from scrapping the clunkers, such that the environmental premium might have hurt the environment in the long-run.

By analyzing the consequences of scrappage schemes, we join a dynamic literature examining stimulus impacts of, i.a., tax rebates, income tax changes, or government spending on infrastructure and education (Parker et al., 2013; House and Shapiro, 2006; Feyrer and Sacerdote, 2011). On a more profound level, our paper is closely related to several studies of scrappage programs as a reaction to (recent) economic crises. Amongst others, Mian and Sufi (2012), Li et al. (2013), Copeland and Kahn (2013), and Hoekstra et al. (2017) use alternative identification strategies to show that the positive effect of the 2009 U.S. scrappage program 'CARS' on new vehicle sales during its two months duration came entirely at cost of a reversed effect the following months.[2] Studies looking at recent scrappage subsidies and their influences on the environment are, i.a., Knittel (2009) and, again, Li et al. (2013). Such papers look at the cost of reducing CO2 and find that, with about \$300-\$450 per ton, the U.S. Cash for Clunkers program was an expensive way to reduce greenhouse gases.[3]

Our paper contributes to the literature in several ways. First, we provide counterfactual evidence regarding new vehicle sales *and* greenhouse gas emissions caused by the German program. Second, we are, to the best of our knowledge, first to expand on such results by providing a detailed model to precisely estimate the share and impact of different sub-groups in the overall sales and emission effect: windfall gains, pull-forward and delayed effects, as well as on-top sales. We use those findings to understand better whether the intervention was effective in achieving its declared goals and how to potentially improve scrappage programs to come, issues that are of particular importance to policy makers. Third, since such analyses demand a proper method of identification allowing for multiple, interdependent outcomes of interest, we develop MSCMT. This generalized and extended version of SCM not only allows to jointly synthesize with respect to several dependent variables, but also to utilize entire time series of economic predictors. In contrast to Gobillon and Magnac (2016) and Xu (2017), who also use *time-varying covariates*, we do not use post-treatment values of these variables. This would only be justified if the covariates' actual and counterfactual post-treatment values coincide, an (implicit) assumption that we do not want to make for the case of the German scrappage program. Robbins et al. (2017), who also consider *multiple outcomes*, look at a setting with high-dimensional data at a granular level, where the treated area has several cases and a large number of untreated comparison cases exists. This is quite different from our aggregated, country-specific Cash-for-Clunkers scenario, which is a rather typical setting for synthetic control applications. We show that the combination of our M and T generalizations leads to (asymptotically) unbiased estimates under quite general conditions. Lastly, we demonstrate how to link this approach with a cross-validation technique determining optimal predictor

[2] Adda and Cooper (2000), Licandro and Sampayo (2006), Schiraldi (2011), and Grigolon et al. (2016) evaluate European scrappage schemes.

[3] Moreover, Sandler (2012) analyzes the cost-effectiveness of a long-running local scrapping program in California (1996-2010), finding the average cost-effectiveness to be much lower than expected. More environmental analyses of past programs are, e.g., those of Hahn (1995) or Szwarcfiter et al. (2005).

weights.

The remainder of the paper is structured as follows: Section 2 describes car scrappage programs during the crisis with a particular focus on Germany and presents our data set including first descriptive evidence of the two outcomes of interest. We develop MSCMT and present its properties in Section 3, while proofs are relegated to Appendix A. Section 4 provides overall MSCMT results for our two outcomes including cross-validation and standard placebo tests. Section 5 lays out our foundation for modeling different vehicle sub-groups and presents estimates on how the program helped to stabilize the car market at the cost of potentially hurting the environment in the long-run. Section 6 concludes.

2 Program, Data, and First Evidence

Cash for Clunkers

During 2008–2010, there were similar scrappage programs in 22 different countries around the globe. The biggest overall budget was provided by Germany with €5 billion, followed by Japan with about €2.9 billion. The U.S. spent about €2 billion and ranks third. In per capita figures, Germany invested about €61, almost three times as much as Japan, Italy, and Luxembourg respectively, who all spent a little more than €20. The U.S. ranks twelfth with about €6.50 per capita. Worldwide, about €15.3 billion were spent on vehicle retirement programs while the European Union contributed €8.8 billion or 58% of this sum. The €5 billion spent by Germany totaled 33% of the worldwide budget and 57% of what was spent by all European countries.

In Germany, the idea for a scrappage program was introduced by then vice-chancellor Frank-Walter Steinmeier in an interview on December 27, 2008. Only two weeks later, the federal government passed the Economic Stimulus Package II including a scrappage scheme—called “environmental premium”. This extremely narrow timing, i.e. fast implementation of the program, comprises a very nice feature of this policy intervention since it almost acted as a true exogenous shock to the demand side. The program officially started on January 14, 2009, financed by the Investment and Repayment Fund. First key points were published on January

16, 2009, by the responsible agency BAFA[4]. The subsidy of €2,500 could be requested by private individuals who scrapped an old passenger car which had to be at least nine years old and licensed to the applicant for at least 12 months. These eligibility criteria were, again, unforeseeable and actually arbitrary, introducing even more exogeneity. Eventually, the new car had to be a passenger car fulfilling at least the emission standard Euro 4[5]. Applications were possible until the end of 2009 or the exhaustion of the budget. The latter happened on September 2, 2009, when 2 million new cars had been subsidized.[6]

Data & Outcomes of Interest

For our upcoming analysis, we gathered data from EUROSTAT covering Germany and 11 European countries without a (recent) car scrappage program. The latter are Belgium, Czech Republic, Denmark, Estonia, Finland, Hungary, Latvia, Lithuania, Poland, Slovenia, and Sweden. Our two outcomes of interest are the monthly figures of new passenger car registrations and the annual average CO2 emissions of newly registered passenger cars, with time series ranging from 2004 until 2012.

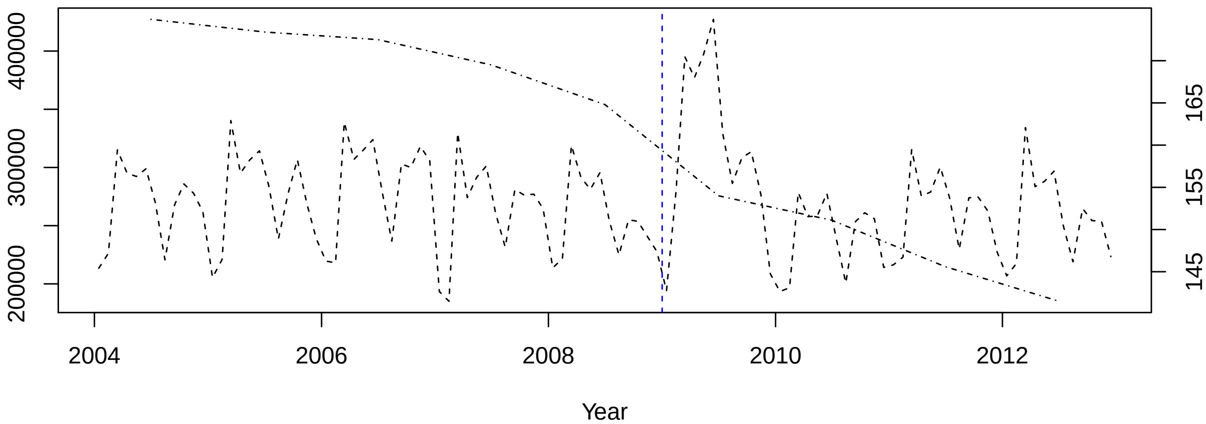
We have chosen these variables for several reasons. First, we use figures for registrations of new cars instead of, e.g., sales, because the latter almost immediately translate into the former and registration figures are, in contrast to sales figures, robust to cross-border consumption and trade. Additionally, data on new car registrations are collected by official agencies at a monthly frequency. Only with data of this frequency, it becomes possible to disentangle the program's effects, calculating how many of the subsidized vehicles would have been purchased anyway (windfall gains), how many had been shifted in time (pull-forward and delayed sales), and how many would not have happened in the absence of the intervention (on-top purchases).

Second, concerning environmental pollution, we work with average CO2 emissions of newly registered cars because these data allow to project overall CO2 emissions over the entire

[4] *Bundesamt für Wirtschaft und Ausfuhrkontrolle* (Federal Office of Economics and Export Control).

[5] European emission standards define the acceptable limits for exhaust emissions of new vehicles sold in EU member states. Actually, for the German case, this prerequisite was redundant since all new cars bought in 2009 had to be Euro 4 equipped anyway.

[6] More precisely, program administration had to be covered by the budget, which implied that the actual number of subsidized purchases was slightly lower than 2 million: 1,933,090.



Note: The dashed line exhibits monthly new passenger car registrations (absolute level, left scale), the dashed-dotted line shows annual CO2 emissions of newly registered passenger cars (g/km, right scale). The blue vertical line indicates the beginning of the German scrappage program.

Figure 1: Monthly Registrations and Annual CO2 Emissions of New Passenger Cars in Germany

lifetime of the newly registered vehicles, separately for the different groups of buyers. This would not be possible for alternative measures like average tons of CO2 emissions of *overall* passenger cars. Moreover, such alternatives mostly feature a shorter time horizon as well as an undesirable signal-to-noise ratio, as we are solely interested in differences induced by the scrappage program, which affected only a rather small subgroup of cars.

Figure 1 displays the development of our variables of interest from 2004 to 2012.[7] Overall, new car registrations were quite stable over the years 2004 till 2008, with about 270,000 registrations per month and very consistent seasonalities. Nevertheless, one can see a slightly positive development from the beginning of 2004 until the end of 2006 and, from then on, a crisis-induced decline until the end of 2008. In 2009, eventually, there was a striking jump up to more than 400,000 monthly new car registrations. One year later, this number dropped back to a little less than its regular 2004–2008 level (240,000 vs. 270,000), featuring a dip which, however, seems to be much smaller than the spike in 2009. Afterwards, once again, there is a positive development of new passenger car registrations until the end of 2012. Our second dependent variable, average CO2 emissions of newly registered passenger cars, shows

[7]In graphics, we use the following convention: when plotting monthly series, values are attributed to the 15th of the corresponding month and then connected by an interpolating straight line. Analogously, annual values are attributed to July 1st of the corresponding year.

a considerable decline of about 11 grams per kilometer during the intervention period. This is by far the biggest drop throughout our observation window with a yearly emission decline of about 2.5 g/km on average over the first four years. In the following post-treatment periods until the end of our observation frame, emissions further fall, but rather slowly.

While this descriptive depiction suggests that both of our outcomes react as intended due to the treatment, it is not sufficient to draw any firm conclusions from it. To be able to do so and, moreover, to disentangle potential subgroups of the overall effects, we apply a suitable estimation strategy, which delivers reliable counterfactual values: *Synthetic Control Methods* (SCM).

3 Multivariate SCM using Time Series

For calculating synthetic controls, we develop the multivariate synthetic control method using time series (MSCMT). This method generalizes and extends the SCM approach of Abadie and Gardeazabal (2003) and Abadie et al. (2010) in two respects simultaneously. First, similar to the approach of Robbins et al. (2017), MSCMT allows to consider more than one variable of interest at a time (M dimension added to SCM), i.e., it takes account of possible interdependencies between such outcomes. Second, like in Gobillon and Magnac (2016) and Xu (2017), MSCMT allows to use entire time series of all variables (T dimension added to SCM), while the original SCM approach is rather designed to using their means. By exploiting the economic predictors' entire variation over time, we efficiently use all data in order to find the best counterfactual for Germany. Note that the two generalizations are independent of each other: one might, depending on the application, synthesize with respect to several variables of interest, relying only on means of the covariates, i.e., employ an MSCM approach; alternatively, one might consider synthesizing with respect to only one variable of interest using time series data of the economic predictors, i.e., employ an SCMT approach. For the case at hand, where we consider both extensions, MSCMT provides a counterfactual Germany that mimics actual Germany with respect to both the number of new car registrations and the average CO2 emissions per km of newly registered cars, additionally exploiting the variation over time of data-chosen covariates.

Before we turn our attention to the specific application in Section 4, we formally introduce the multivariate synthetic control method using time series, discuss theory on the estimator's properties, and explain how we use cross-validation in combination with synthetic control methods.

General Model Setup

In general, synthetic control methods aim at producing an appropriate counterfactual for the treated unit, which describes how the variables of interest would have developed if there had been no treatment. In other words, the ultimate goal of MSCMT is to come up with approximations $\widehat{Y}_{k1t}$ for $\widetilde{Y}_{k1t}$, with the latter denoting the counterfactual values one would have observed for the treated unit if there had been no intervention, while Y_{k1t} denotes the actual value of the k-th variable of interest for the treated unit at some post-treatment time t. The true but unknown effect of the intervention on the k-th variable of interest at post-treatment times t, which is given by

$$\mathrm{Eff}_{k,t} := Y_{k1t} - \widetilde{Y}_{k1t}, \tag{1}$$

is then approximated by

$$\widehat{\mathrm{Eff}}_{k,t} := Y_{k1t} - \widehat{Y}_{k1t}. \tag{2}$$

The error in estimating the intervention's effect, $\widehat{\mathrm{Eff}}_{k,t} - \mathrm{Eff}_{k,t}$, coincides with $\widetilde{Y}_{k1t} - \widehat{Y}_{k1t}$. Thus, it is essential that $\widehat{Y}_{k1t}$ is a good approximation to $\widetilde{Y}_{k1t}$. For constructing $\widehat{Y}_{k1t}$, data from J so-called donor units are used, which are to some extent similar to the treated unit but have not been exposed to the intervention. These donor units are combined to form a so-called synthetic control unit, which serves as an approximation for the treated unit's counterfactual development. For determining the synthetic control unit, one uses non-negative weights $w_2, \ldots, w_{J+1}$, which sum up to unity and are collected in a J-dimensional vector $W = (w_2, \ldots, w_{J+1})'$. In our application, we have $J = 11$, and w_2 will be denoting the weight of Belgium and w_{12} that of Sweden.[8] Given such weights W, the counterfactual values of the dependent variables for the treated unit, $\widetilde{Y}_{k1t}$, are approximated by

$$\widehat{Y}_{k1t}(W) := \sum_{j=2}^{J+1} w_j Y_{kjt}. \tag{3}$$

As stated above, the approximation aims at $\widehat{Y}_{k1t}(W)$ being close to the counterfactual values $\widetilde{Y}_{k1t}$ for post-treatment times t. One therefore would like to choose the weights W such that the post-treatment approximation error $\widetilde{Y}_{k1t} - \widehat{Y}_{k1t}(W)$ is minimized. Unfortunately, this is not operational, as the counterfactual values $\widetilde{Y}_{k1t}$ are unknown. In their seminal papers, Abadie and Gardeazabal (2003) and Abadie et al. (2010) have introduced a principle which is operational and can be shown to produce reasonable estimators: instead of trying to minimize *post-treatment* differences between outcome variables of the synthetic control and unavailable counterfactual outcome values $\widetilde{Y}_{k1t}$, one aims at minimizing *pre-treatment* differences between actual outcome values Y_{k1t} and synthetic outcome values $\widehat{Y}_{k1t}(W)$ as well as *pre-treatment* differences between actual and synthetic values of covariates.[9] The covariates are variables which have predictive power for the outcomes of interest and which we allow to be time-varying (T dimension added to SCM). Thus, we will in the sequel denote by C_{l1t} the actual pre-treatment value of the l-th covariate for the treated unit at time t, by C_{ljt} the actual pre-treatment value of the l-th covariate for the j-th unit at time t, and by $\widehat{C}_{l1t}(W) := \sum_{j=2}^{J+1} w_j C_{ljt}$ the pre-treatment value of the l-th covariate for the synthetic control unit at time t.

In contrast to Gobillon and Magnac (2016) and Xu (2017), who also use time-varying covariates, we do not use *post-treatment* values of these variables. This would only be justified if the covariates' actual and counterfactual post-treatment values coincide, an (implicit) assumption that we do not want to make for the case of the German scrappage program. For instance, we use GDP as a covariate,[10] and as the rebate aimed at stabilizing the car market

[8] As is standard in the literature on SCM, we also use the terms "donors", "donor units", and "control units" for the countries that can potentially be used to synthesize Germany. Taken together, all those control units make up the so-called "donor pool". For standard assumptions concerning the donor pool, see Abadie et al. (2015). Analogously, we also call Germany the "treated unit".

[9] Note that this principle, which is standard in the SCM context, implicitly assumes that actual and counterfactual *pre-treatment* values coincide, both for the outcome variables as well as for the covariates.

[10] We will elaborate on our set of predictors in detail in Section 4.

by incentivizing owners of clunkers to buy new cars, the intervention might potentially have affected GDP post-treatment.

Overall, the multivariate synthetic control method using time series thus consists of matching the pre-treatment values of the outcome variables, $\widehat{Y}_{k1t}(W)$, to the corresponding values Y_{k1t}, and matching the pre-treatment values of the covariates, $\widehat{C}_{l1t}(W)$, to the corresponding values C_{l1t}. For estimating the effect of the intervention, we use $\widehat{\mathrm{Eff}}_{k,t}(W) := Y_{k1t} - \widehat{Y}_{k1t}(W)$ for post-treatment time points t. Combining equations (1), (2), and (3), the error in estimating the true effect $\mathrm{Eff}_{k,t}$ is then found to be

$$\widehat{\mathrm{Eff}}_{k,t}(W) - \mathrm{Eff}_{k,t} = \widetilde{Y}_{k1t} - \sum_{j=2}^{J+1} w_j Y_{kjt}. \tag{4}$$

Properties of the MSCMT estimator

For studying the properties of the MSCMT estimator under different data generating processes, we first introduce some notation: for every unit ($j = 1, \ldots, J+1$), we denote by

$$X_{j,t} := (Y_{1jt}, \ldots, Y_{Kjt}, C_{1jt}, \ldots, C_{Ljt})'$$

the $(M := K + L)$-dimensional stacked vector consisting of the outcomes of interest's actual values (first K elements of X) and the corresponding actual values of the covariates (last L elements of X) at *pre*-treatment time t. Similarly, for *post*-treatment times t, we denote by

$$X_{1,t} := \left(\widetilde{Y}_{11t}, \ldots, \widetilde{Y}_{K1t}, \widetilde{C}_{1t}, \ldots, \widetilde{C}_{L1t}\right)', X_{j,t} := (Y_{1jt}, \ldots, Y_{Kjt}, C_{1jt}, \ldots, C_{Ljt})' \, (j = 2, \ldots, J+1)$$

the corresponding vector of the treated unit's counterfactual and the donor units' actual values at time t. Therefore, $X_{j,t}$ describes how outcomes and covariates would have evolved in absence of the intervention.[11]

For given weights $W = (w_2, \ldots, w_{J+1})'$, we denote by $e_t(W) := X_{1,t} - \sum_{j=2}^{J+1} w_j X_{j,t}$ the M-

[11] Recall that it is implicitly assumed that the actual *pre*-treatment values of outcomes and covariates are identical to the corresponding counterfactual ones. Thus, it is essential that the intervention could not be anticipated. Additionally, one builds on the implicit assumption that, at any time, the donor units' values are not affected by the intervention.

dimensional approximation error at time t induced by the weights W. For *pre*-treatment times t, $e_t(W)$ consists of K components which give the errors when approximating the treated unit's actual outcome values, and L components which describe the corresponding approximation errors with respect to the covariates. For *post*-treatment times t, we find

$$e_t(W) = \left(\widetilde{Y}_{11t}, \ldots, \widetilde{Y}_{K1t}, \widetilde{C}_{1t}, \ldots, \widetilde{C}_{L1t}\right)' - \sum_{j=2}^{J+1} w_j \left(Y_{1jt}, \ldots, Y_{Kjt}, C_{1jt}, \ldots, C_{Ljt}\right)'. \tag{5}$$

Due to Equation (4), the first K components of $e_t(W)$ coincide with the estimation errors of the effect estimators.

In order to study the properties of $e_t(W)$, we further denote by $\widetilde{A}_{0,t}, \ldots, \widetilde{A}_{p,t}$ possibly time-varying $(M \times M)$-dimensional coefficient matrices and by $\widetilde{\varepsilon}_{j,t}$ an M-dimensional error term, while $\widetilde{\delta}_t$ denotes an M-dimensional common trend. We assume that, for every unit j, $X_{j,t}$ is given by a vector autoregression of order p plus the common trend $\widetilde{\delta}_t$:

$$X_{j,t} = \sum_{h=0}^{p} \widetilde{A}_{h,t} X_{j,t-h} + \widetilde{\delta}_t + \widetilde{\varepsilon}_{j,t} = \widetilde{A}_{0,t} X_{j,t} + \ldots + \widetilde{A}_{p,t} X_{j,t-p} + \widetilde{\delta}_t + \widetilde{\varepsilon}_{j,t},$$

for which $I - \widetilde{A}_{0,t}$ is invertible for all t, with I denoting the M-dimensional identity matrix. The above equation can then be rewritten as

$$X_{j,t} = \left(I - \widetilde{A}_{0,t}\right)^{-1} \left(\sum_{h=1}^{p} \widetilde{A}_{h,t} X_{j,t-h} + \widetilde{\delta}_t + \widetilde{\varepsilon}_{j,t}\right) = \sum_{h=1}^{p} A_{h,t} X_{j,t-h} + \delta_t + \varepsilon_{j,t}, \tag{6}$$

with $A_{h,t} := \left(I - \widetilde{A}_{0,t}\right)^{-1} \widetilde{A}_{h,t}$, $\delta_t := \left(I - \widetilde{A}_{0,t}\right)^{-1} \widetilde{\delta}_t$, and $\varepsilon_{j,t} := \left(I - \widetilde{A}_{0,t}\right)^{-1} \widetilde{\varepsilon}_{j,t}$. For $e_t(W)$, the derivation

$$
\begin{aligned}
e_t(W) &= X_{1,t} - \sum_{j=2}^{J+1} w_j X_{j,t} \\
&= \sum_{h=1}^{p} A_{h,t} X_{1,t-h} + \delta_t + \varepsilon_{1,t} - \sum_{j=2}^{J+1} w_j \left(\sum_{h=1}^{p} A_{h,t} X_{j,t-h} + \delta_t + \varepsilon_{j,t} \right) \\
&= \sum_{h=1}^{p} A_{h,t} \left(X_{1,t-h} - \sum_{j=2}^{J+1} w_j X_{j,t-h} \right) + \delta_t \left(1 - \sum_{j=2}^{J+1} w_j \right) + \varepsilon_{1,t} - \sum_{j=2}^{J+1} w_j \varepsilon_{j,t} \\
&= \sum_{h=1}^{p} A_{h,t} e_{t-h}(W) + \varepsilon_{1,t} - \sum_{j=2}^{J+1} w_j \varepsilon_{j,t}.
\end{aligned}
\tag{7}
$$

shows that $e_t(W)$ is composed of two terms: an autoregression of order p and the noise-like term $\varepsilon_{1,t} - \sum_{j=2}^{J+1} w_j \varepsilon_{j,t}$. The latter term has expectation zero for *post*-treatment times t and is characterized by a positive variance which, due to the presence of $\varepsilon_{1,t}$, cannot be made arbitrarily small by cleverly choosing donor weights W—even for a large number of donor units. Therefore, $\sum_{h=1}^{p} A_{h,t} e_{t-h}(W)$, the autoregressive part of $e_t(W)$, is the decisive one with respect to choosing W: if the donor weights W are such that the approximation errors $e_t(W)$ are annihilated or at least close to 0 immediately before the treatment, the estimation errors of the effect estimators will essentially consist of the unavoidable part $\varepsilon_{1,t} - \sum_{j=2}^{J+1} w_j \varepsilon_{j,t}$. More formally, the following proposition can be stated, which generalizes the result (5) of Abadie et al. (2010, p. 496) from a univariate autoregressive process of order 1 to a multivariate autoregressive process of order p.

Proposition 1. *Denoting by T_0 the time of the intervention, we have:*

1. *the bias of the vector of effect estimators, $(Y_{11t} - \widehat{Y}_{11t}(W), \ldots, Y_{K1t} - \widehat{Y}_{K1t}(W))'$, conditional on $\mathcal{F}_{T_0}$, the information up to time T_0, is a linear function of the p approximation errors $e_{T_0}(W), \ldots, e_{T_0-(p-1)}(W)$,*

2. *if $T_0 \geq p$ and the weights $W^* = (w_2^*, \ldots, w_{J+1}^*)$ satisfy $X_{1,t} = \sum_{j=2}^{J+1} w_j^* X_{j,t}$ for $t = T_0, \ldots, T_0 - (p-1)$, then $E(e_t(W)|\mathcal{F}_{T_0}) = 0$ for all $t \geq T_0 + 1$, and the vector of effect estimators is unbiased.*

Proof. See Appendix A. □

From Proposition 1 we learn that a perfect fit of the last p observations of the outcomes and covariates prior to the treatment renders the vector of effect estimators unbiased, i.e., the effects on all outcomes are estimated without bias. Furthermore, the first part of the proposition shows that the better the approximations of outcomes and covariates are prior to the treatment, the smaller a potential bias of the effect estimators will be.

We will now consider the more general case, which allows for time-varying, unobserved confounders to be present, too, leading to an additional term

$$u_{j,t} := \Big(\lambda_{1,t}\mu_{1,j}, \ldots, \lambda_{M,t}\mu_{M,j}\Big)',$$

with $\lambda_{m,t} \in \mathbb{R}^{1\times F_m}$ describing the values of the F_m unobserved confounders affecting the m-th component of the stacked vector X, and $\mu_{m,j} \in \mathbb{R}^{F_m \times 1}$ denoting the corresponding factor loadings of unit j. All in all, we thus assume that the stacked vector of outcomes and covariates, X, is given by

$$X_{j,t} = \sum_{h=1}^{p} A_{h,t}X_{j,t-h} + \delta_t + u_{j,t} + \varepsilon_{j,t} = A_{1,t}X_{j,t-1} + \ldots + A_{p,t}X_{j,t-p} + \delta_t + u_{j,t} + \varepsilon_{j,t}.$$

In this more general setting, formula (7) generalizes to

$$e_t(W) = \sum_{h=1}^{p} A_{h,t}e_{t-h}(W) + u_{1,t} - \sum_{j=2}^{J+1} w_j u_{j,t} + \varepsilon_{1,t} - \sum_{j=2}^{J+1} w_j \varepsilon_{j,t}. \tag{8}$$

and, hence, $e_t(w)$ contains the additional term

$$u_{1,t} - \sum_{j=2}^{J+1} w_j u_{j,t} = \left(\lambda_{1,t}\Big(\mu_{1,1} - \sum_{j=2}^{J+1} w_j\mu_{1,j}\Big), \ldots, \lambda_{M,t}\Big(\mu_{M,1} - \sum_{j=2}^{J+1} w_j\mu_{M,j}\Big)\right)',$$

which might render the effect estimators biased, as there is no guarantee that the approximations $\sum_{j=2}^{J+1} w_j\mu_{m,j}$ are close to the unobservable factor loadings $\mu_{m,1}$. However, generalizing the result (1) of Abadie et al. (2010, p. 495), the following proposition shows that a potential bias of the effect estimators will shrink to zero when the pre-treatment time span grows large.

Proposition 2. *In the presence of unobserved confounders, if* W^* *is such that* $X_{1,t}$ $=$

$\sum_{j=2}^{J+1} w_j^* X_{j,t}$ *for* $t = 1, \ldots, T_0$*, the vector of effect estimators is asymptotically unbiased under some technical assumptions*[12]*, i.e., the bias vanishes when the pre-treatment time span grows to infinity.*

Proof. See Appendix A. □

Overall, for both data generating processes considered above, the MSCMT estimators provide good estimators for measuring the intervention's effects, at least as long as a sufficiently large number of pre-treatment observations of both outcomes and covariates are close to each other for the actually treated and the corresponding synthetic control unit.

MSCMT Predictor Weights by Cross-Validation

While it is clear from the theory outlined above that the donor weights W should ideally be chosen such that the synthetic control unit comes as close as possible to the actually treated unit in terms of pre-treatment values of outcome variables and covariates, it remains unclear how this should be operationalized. Typically, it will be impossible to obtain a perfect fit for all so-called economic predictors, i.e., for the pre-treatment values of all outcomes and covariates. Thus, a weighting scheme is needed, which determines how much weight is put on each predictor when matching them. When determining such predictor weights, it must be taken into account that some variables might have more predictive power for the outcomes than others, or that some of them might even have no predictive power at all, having been added to the model just to make sure that no potentially important predictors are omitted.[13] Therefore, so-called predictor weights $V = (v_1, \ldots, v_M)$ are introduced, where each v_m acts as a weight describing how much emphasis should be put on fitting the m-th predictor variable. For determining these weights from the data, we slightly generalize the cross-validation technique introduced by Abadie et al. (2015) and refined by Klößner et al. (2018) and Becker et al. (2018).

[12]For details on the technical assumptions, see Appendix A.

[13]For the models discussed above, predictive power hinges on the entries of the coefficient matrices $A_{h,t}$. For instance, if the (k, m)-entry of A_{1,T_0} is large, then Equations (7) and (8) entail that errors in approximating X_{m1T_0} immediately translate into large errors in estimating Eff_{k,T_0+1}.

The principal idea of the cross-validation technique is to split up the pre-treatment time span into two disjoint parts, the 'training period' and the 'validation period', with the former preceding the latter. The training period's data on the economic predictors is used to determine, for any given predictor weights V, donor weights $W^*_{\text{train}}(V)$, which produce the best fit of the predictor variables in the training period. The validation period's data on the outcome variables is then used to assess the fit that $W^*_{\text{train}}(V)$ implies for the outcome variables, and to search for those predictor weights V^* optimizing this fit.

More precisely, within the training part of the so-called cross-validation step, given any hypothetical predictor weights V, $W^*_{\text{train}}(V)$ is determined as the very non-negative vector $W = (w_2, \ldots, w_{J+1})$, with components summing to unity, for which $\sum_{m=1}^{M} v_m \, \text{MSE}^{\text{train}}_{X,m}(W)$ becomes minimal. Here, $\text{MSE}^{\text{train}}_{X,m}(W)$ denotes the appropriately scaled mean squared approximation error that results from approximating the m-th predictor variable's values in the training period, X_{m1t}, by its synthetic counterpart $\sum_{j=2}^{J+1} w_j X_{mjt}$:

$$\text{MSE}^{\text{train}}_{X,m}(W) := \frac{\frac{1}{N^{\text{train}}_m} \sum_{t \in \mathcal{T}_{\text{train}}} \left(X_{m1t} - \sum_{j=2}^{J+1} w_j X_{mjt} \right)^2}{\text{Var}(\{X_{mjt} : j = 1, \ldots, J+1, t \in \mathcal{T}_{\text{train}}\})}, \tag{9}$$

with N^{train}_m denoting the number of observations of the m-th economic predictor during the training period, $\mathcal{T}_{\text{train}}$. Note that the denominator appearing in the definition of $\text{MSE}^{\text{train}}_{X,m}$ ensures that the mean squared approximation errors are comparable across predictors, because mean squared differences between actual and synthetic values are divided by the amount by which the data on the corresponding predictor vary during the training period.[14] Table 1 provides an overview over the different time periods employed by the cross-validation technique.

In the validation part of the cross-validation step, one then searches for those predictor weights V^* that would have produced the smallest prediction error for the outcome variables in the validation period. To this end, the overall prediction error of the outcome variables, which is also called cross-validation criterion, is defined as the square root of the unweighted

[14] We follow the literature by using the data on *all* units in the denominator of Equation (9). The advantage of this approach is that the data can be rescaled prior to all calculations and do not have to be rescaled again when conducting placebo studies.

Table 1: Overview over Time Periods employed by Cross-Validation Technique

	cross-validation step	main step
for fitting predictors	$\mathcal{T}_{\text{train}}$ (2004–2006)	$\mathcal{T}_{\text{pred}}$ (2006–2008)
for estimating outcomes	$\mathcal{T}_{\text{valid}}$ (2007–2008)	$\mathcal{T}_{\text{est}}$ (2009–2010)

mean of all outcomes' mean squared (prediction) errors $\text{MSE}^{\text{valid}}_{Y,k}(W^*_{\text{train}}(V))$:[15]

$$\Delta_Y^{\text{C-V}}(V) := \sqrt{\frac{1}{K}\sum_{k=1}^{K} \text{MSE}^{\text{valid}}_{Y,k}(W^*_{\text{train}}(V))} \tag{10}$$

These outcome-specific mean squared prediction errors are in turn defined as the appropriately scaled mean squared errors resulting from approximating Y_{k1t}, the k-th outcome variable's values in the validation period $\mathcal{T}_{\text{valid}}$, by $\sum\limits_{j=2}^{J+1} w^*_{\text{train},j}(V)Y_{kjt}$, their synthetic counterpart:

$$\text{MSE}^{\text{valid}}_{Y,k}(W^*_{\text{train}}(V)) := \frac{\frac{1}{N_k^{\text{valid}}}\sum\limits_{t\in\mathcal{T}_{\text{valid}}}\left(Y_{k1t} - \sum\limits_{j=2}^{J+1} w^*_{\text{train},j}(V)Y_{kjt}\right)^2}{\text{Var}(\{Y_{kjt} : j = 1,\dots,J+1, t \in \mathcal{T}_{\text{valid}}\})}, \tag{11}$$

with N_k^{valid} denoting the number of observations of the k-th outcome variable in the validation period $\mathcal{T}_{\text{valid}}$.

Eventually, the optimal predictor weights V^* found in the cross-validation step are used within the so-called 'main step' for determining the final donor weights $W^*_{\text{main}}(V^*)$, which are used to build the synthetic control unit. Using the economic predictors' pre-treatment data during the 'main period' immediately prior to the intervention, $W^*_{\text{main}}(V^*)$ is calculated as the minimizer of $\sum\limits_{m=1}^{M} v^*_m \,\text{MSE}^{\text{pred}}_{X,m}(W)$, where $\text{MSE}^{\text{pred}}_{X,m}(W)$ measures the fit with respect to the predictors during the main period:

$$\text{MSE}^{\text{pred}}_{X,m}(W) := \frac{\frac{1}{N_m^{\text{pred}}}\sum\limits_{t\in\mathcal{T}_{\text{pred}}}\left(X_{m1t} - \sum\limits_{j=2}^{J+1} w_j X_{mjt}\right)^2}{\text{Var}(\{X_{mjt} : j = 1,\dots,J+1, t \in \mathcal{T}_{\text{pred}}\})}, \tag{12}$$

with N_m^{pred} denoting the number of observations of the m-th economic predictor during the economic predictors' main period.

[15] One might introduce appropriate weighting schemes here in order to take account of potential differences in the outcomes' importance.

In our application, the period of interest after the treatment is given by years 2009 and 2010, as the scrappage program might have induced buyers to buy new cars in 2009 and to pull forward their purchases from 2010 to 2009 in order to profit from the subsidy. We therefore set the validation period to years 2007 and 2008, making the validation period mimic the estimation window we are mainly interested in. As our data start in 2004, we set the training period to years 2004–2006 and use the cross-validation technique as described above to determine optimal predictor weights V^*. In order to ensure that the lengths of the time spans used during the cross-validation step match those used during the main step, we set the economic predictors' main period to years 2006–2008, and employ V^* to compute the donor weights $W^*_{\text{main}}(V^*)$ using the predictors' data from 2006 to 2008.

Overall, cross-validating MSCMT—as we employ it here—thus consists of a 'test run' of MSCMT using the predictors' data in the training period from 2004 to 2006 to predict the outcomes in the validation period consisting of years 2007 and 2008. Comparing these predictions delivered by MSCMT to the corresponding actual values, the predictor weights V^* are chosen optimally, i.e., such that the prediction errors are as small as possible. After this test run, the periods are shifted two years ahead in time to $\mathcal{T}_{\text{pred}} = \mathcal{T}_{\text{train}}+2$ and $\mathcal{T}_{\text{est}} = \mathcal{T}_{\text{valid}}+2$ (all times measured in years), and the main run of MSCMT is carried out using the optimal predictor weights V^*.

4 Empirical Application of MSCMT

Model Specification

Prior to any actual calculations, we transform the outcome variables in order to make them comparable across countries. This comparability is needed to ensure that the values for the treated unit are within the convex hull of the donors' values, which is important for the synthetic control approach to work properly, see Gobillon and Magnac (2016).[16] To this end, we transform 'monthly new car registrations' figures into change rates of registrations, by calculating the percentage change rate between the new registrations in a given month and corre-

[16] An alternative approach to potentially bypass the convex hull condition is discussed in Powell (2018).

sponding registrations twelve months earlier. In so doing, we eliminate any country-specific seasonality and construct a new time series which fluctuates around zero for all countries. Similarly, the annual time series on 'average CO2 emissions of newly registered cars per kilometer driven' is made comparable by expressing all values as percentages relative to the emissions at the beginning of the observation period in 2004. By this construction, emission figures take values close to 100 for all countries.

With respect to covariates, we build on an earlier draft of this paper where we used ten variables in an MSCMT approach *without* cross-validation. Of the ten covariates, seven variables broadly reflect the respective country's economy, the vehicle market, and the environmental periphery: annual GDP per capita (quarterly data), the unemployment rate (annual data), two harmonized consumer price indices (Cars and Energy, monthly data), the share of passenger cars on overall transportation (annual data), CO2 emissions per inhabitant (annual data), and environmental tax revenues from the passenger transportation sector (percentage of GDP, annual data). Three more covariates deliver individual and household specific indicators: net earnings (PPS, annual data), annual consumption expenditures per capita (quarterly data), and pensions per capita (annual data).[17]

In this paper, however, we *actually use* the cross-validation technique to select an optimal subset of these covariates.[18] We run the first step of the cross-validation approach, taking as potential set of covariates each of the 1,024 subsets of the ten variables in turn, and calculate the corresponding cross-validation criterion defined in Equation (10).[19] Delivering the smallest cross-validation criterion, $\Delta_Y^{\text{C-V}}(V) = 0.0464957$, the specification with CO2 emissions per inhabitant, environmental tax revenues from the passenger transportation sector, and GDP as covariates emerges as the winner of our specification search.[20]

This specification in particular outperforms the specification without covariates, as the

[17]All variables are EUROSTAT data, running from 2004 until 2008.

[18]Specification searches without objective selection criteria and their sensitivity to "cherry picking" are discussed in Ferman et al. (2017).

[19]For all calculations, we use the statistical software R (R Core Team, 2018) in combination with package MSCMT (Becker and Klößner, 2018b), as other software on synthetic control methods has been found to be not always reliable, see Becker and Klößner (2017) and Becker and Klößner (2018a).

[20]Note that models with a small set of covariates are not nested in models with a larger set of covariates, as predictor weights must be restricted away from zero both for theoretical as well as for numerical reasons, see Becker and Klößner (2018a).

latter is characterized by $\Delta_Y^{\text{C-V}}(V) = 0.05771181$, an increase of about 24% in the root mean squared prediction error in the validation period as compared to the optimal specification.[21] Furthermore, the 1,024 MSCM model variants—where the predictors' averages are used instead of treating them as time series—also perform much worse than the optimal specification, with the best of these delivering $\Delta_Y^{\text{C-V}}(V) = 0.05150044$. Thus, they perform substantially worse than the models using the entire time series variation.[22] All subsequent analyses will therefore be carried out using MSCMT and the winner specification consisting of the three covariates mentioned above.

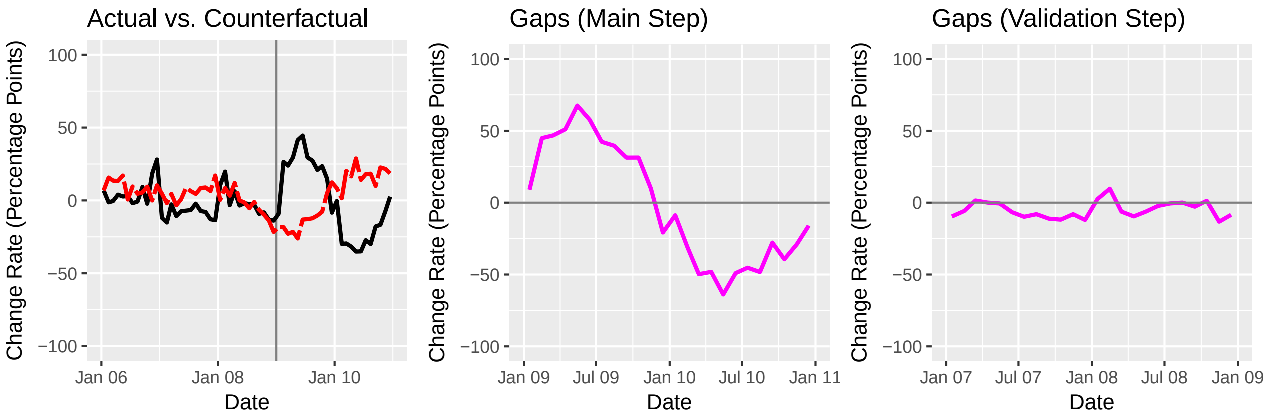
Estimation Results

The cross-validation technique finds the following predictor weighting scheme to be optimal with respect to the cross-validation criterion: 0.0000693 for the registration change rate, 0.69297 for CO2 emissions of newly registered cars, 0.29010 for CO2 emissions per inhabitant, 0.015653 for GDP, and 0.0012148 for environmental tax revenues from the passenger transportation sector. Employing these predictor weights, the donor weights for synthesizing Germany in the main step are as follows: Belgium 68.02%, Czech Republic 3.72%, Estonia 6.39%, Finland 0.095%, Poland 5.33%, and Sweden 16.46%; the remainder of the donor pool receives weights of zero, respectively. Using these donor weights to build the synthetic Germany, we can calculate counterfactual timelines for our outcome variables.

The left panel of Figure 2 displays the timelines of the new car registrations' change rates of actual Germany and its synthetic counterpart. The figure shows that in 2008, immediately prior to the intervention, actual and counterfactual change rates of registrations are quite close to each other, while as compared to their synthetic counterpart, actual change rates are considerably larger during 2009 and smaller during 2010. These effects can be explained as follows: while registration figures were stable during 2008, the scrappage premium induced an increase in new car registrations in 2009, leading to largely positive change rates in 2009.

[21] SCM cases without covariates, also called "constrained regressions", are discussed and applied in, e.g., Botosaru and Ferman (2017), Chernozhukov et al. (2017), Doudchenko and Imbens (2016), Kaul et al. (2015), Powell (2017), and Powell (2018).

[22] When using averages, Formulas (9) and (12) change with respect to the order of taking means and squares. Precisely, means of variables are calculated first and only then differences are squared.

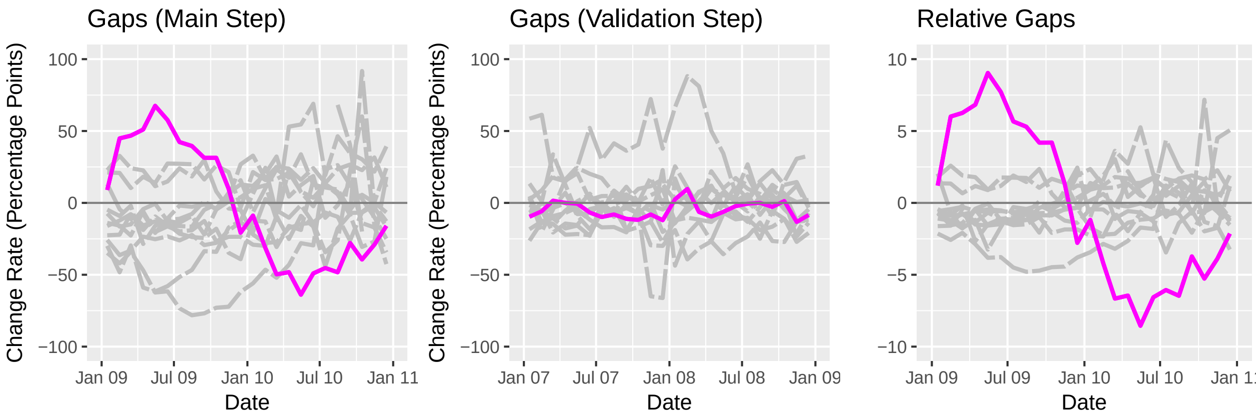


Note: Left panel: comparison of actual (black, solid) and synthetic (red, dashed) values of new registrations' change rate for Germany. Middle & right panel: differences between actual and synthetic values of new registrations' change rate for Germany during estimation window (middle) and evaluation window (right).

Figure 2: New Passenger Car Registration Results for Germany: Gap Plots using Cross-Validation

With respect to 2010, two effects were at work which both led to considerably negative values for the change rate of registrations. On the one hand, the 2009 registration figures were larger due to the intervention, which itself leads to negative change rates in 2010—even if the 2010 figures were 'normal', i.e., not influenced by the Cash-for-Clunkers program. On the other hand, registration figures in 2010 might have additionally been reduced by the intervention, because purchases were pulled forward from 2010 to 2009 in order to profit from the scrappage premium. Thus, one cannot easily infer whether the negative actual change rates in 2010 are driven only by additional purchases that happened in 2009, only by pull-forward purchases, or by both effects. We will tackle the problem of disentangling such potential mechanisms in Section 5 below.

The middle panel of Figure 2 plots the difference between actual and counterfactual change rates of registrations in the main, post-intervention estimation window from 2009 to 2010. This so-called gap plot shows that, during 2009, the effect of the intervention on the new registrations' change rate was largely positive, partially taking values well above 50 percentage points. This indicates that, for some months, absolute actual registration figures towered over counterfactual ones by more than half of the corresponding amount in 2008. Similarly, we find considerably large negative gaps during 2010. Although these gaps in change rates are quite impressive, one might wonder whether these differences might still be explained by random variation. In the right panel of Figure 2, we therefore contrast these gaps with the analogous



Note: Left & middle panel: differences between actual and synthetic values of new registrations' change rate for Germany (magenta, solid) and placebos (gray, dashed) during estimation window (left) and validation window (middle). Right panel: gaps in estimation window relative to average gap size in validation window.

Figure 3: New Passenger Car Registration Results for Germany: Gap-Plots using Cross-Validation incl. Placebos

gaps obtained for the validation period. As the latter take rather small absolute values close to zero, we have first evidence that the intervention caused the change rates to take larger values in 2009 and smaller values in 2010.

In order to further explore the significance of the gaps in the main estimation window, Figure 3 repeats parts of Figure 2, additionally showing so-called placebos results for inference. In the context of SCM, a growing number of rather informal inferential techniques have emerged, which are all rooted in the framework of permutation tests (see, e.g., Abadie et al., 2010, 2015; Ferman and Pinto, 2017; Firpo and Possebom, 2017). Using a placebo study, the treatment is reassigned to a comparison unit, meaning that the actually treated unit, Germany, moves into the donor pool while one of the control units is synthesized instead.[23] Applying this idea to each country from the original donor pool allows to compare the estimated effect of the German program to the distribution of placebo effects obtained for other countries. One can consider the effect of the program to be significant if the estimated impact for Germany is (absolutely) large relative to the outcome distribution of adequately synthesized placebo units. After the treatment-cutoff, Germany's impact on new registrations' change rates clearly stands out of the bulk of control units, featuring an evidently significant treatment effect (left panel of Figure 3). In contrast, Germany's pre-treatment prediction error in the validation window (middle panel of Figure 3) is completely covered by a band consisting of the placebo

[23]However, our subsequent placebo results do not depend on Germany being included in the donor pool.

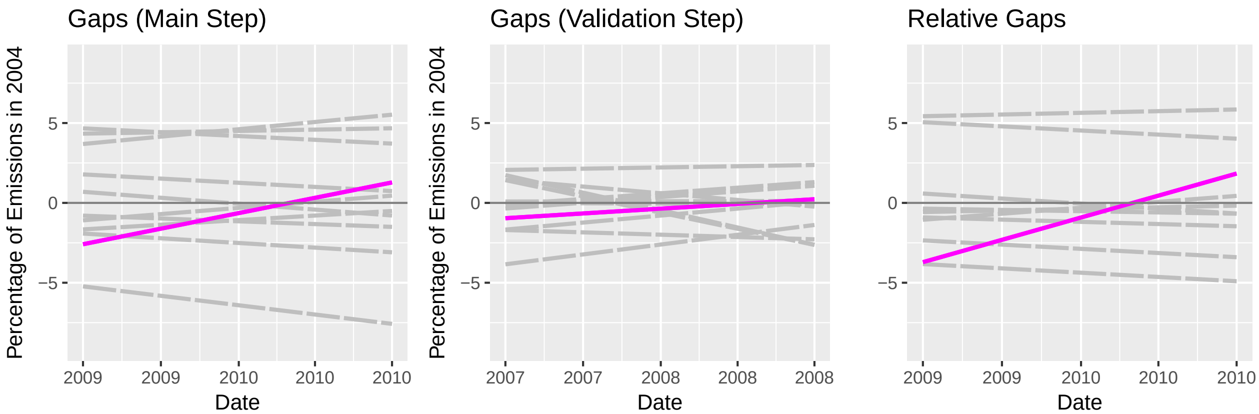
countries' results.

A last test is presented in the right panel of Figure 3. Here, we divide the country-specific monthly values of the left panel of Figure 3 by the average cross-validation RMSPE of the corresponding country as plotted in the middle panel of Figure 3.[24] More precisely, we scale the (monthly) result of the post-intervention period by the (average) precision of the out-of-sample-prediction in the validation period. This is a sensible procedure since a large (placebo) treatment effect is meaningless if the prediction error in the validation period between the respective unit and its synthetic control is large, too. Hence, we deflate the treatment effect of those countries that show a rather imprecise out-of-sample result in the validation window, while we inflate the treatment effect in case of a good prediction for the validation period. Again, Germany stands out of the mass of placebos, clearly featuring a significant impact. In particular, the relative gaps are close to 10 in absolute value, which shows that the estimated treatment effect (for some months) is almost ten times as large as the pre-intervention out-of-sample prediction error in the validation period. All in all, our confidence that Germany's sizable synthetic control estimate actually reflects the effect of the scrappage stimulus is strengthened since no similar or larger estimates arose when the treatment was artificially re-assigned to units not exposed to the intervention.

With respect to CO2 pollution, Figure 4 reveals that the point estimates for the treatment effect show a decline in emissions of the average newly registered car in 2009 by about 2.5% (of 2004 emissions), and an increase of about 1.25% in 2010. However, these effects are rather small when compared to the corresponding placebo results, and they also do not appear to be large in absolute value when compared to the corresponding out-of-sample prediction errors in the validation period. The insignificance of the CO2 results also remains true if we consider the relative gaps, i.e., when we scale the result of the post-intervention period by the average precision of the predictions for the validation period.

In a final step throughout the next section, we formally disentangle our measured effects, focusing on new passenger car registrations and mostly abstracting from corresponding CO2

[24]This approach is similar in spirit to the ratios of post- over pre-treatment RMSPE employed by Abadie et al. (2015).



Note: Left & middle panel: differences between actual and synthetic values of CO2 emissions in terms of 2004 emissions for Germany (magenta, solid) and placebos (gray, dashed) during estimation window (left) and validation window (middle). Right panel: gaps in estimation window relative to average gap size in validation window.

Figure 4: CO2 Emission Results for Germany: Gap-Plots using Cross-Validation incl. Placebos

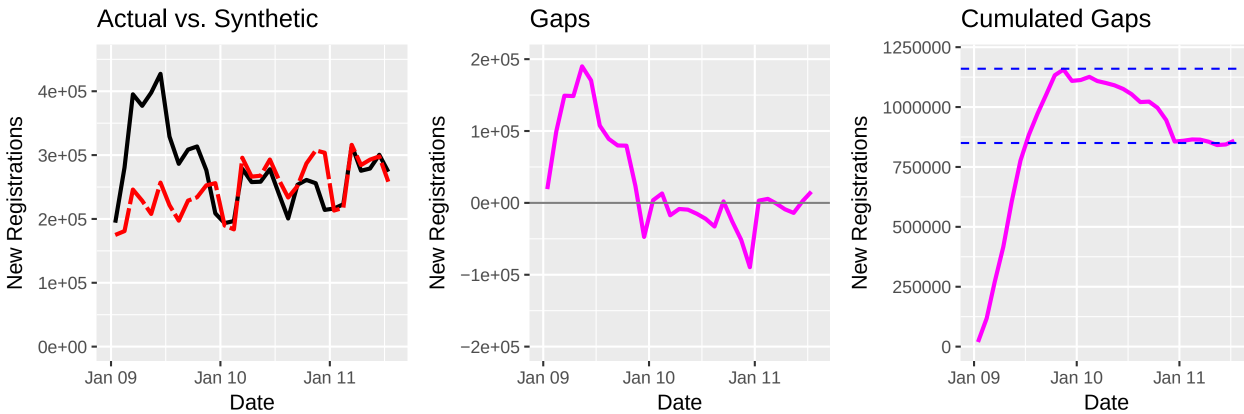
emissions, due to the (in)significance of our former findings.

5 Disentangling the MSCMT Effects

The Long-Run Impact and Time-Shifting of Sales

Figure 5 shows how the results on the registrations' change rates translate into absolute numbers of registrations.[25] The figure's left panel, which displays the actual and counterfactual timelines of new car registrations, reveals that actual registration figures in 2009 were much larger than their synthetic counterpart, showing that the intervention led to a huge increase in registrations in the first year. Contrarily, actual registration figures in 2010 are somewhat below their synthetic counterpart, indicating that some buyers did not purchase new cars in 2010, although they would have done so in the absence of the intervention. These buyers pulled their purchases forward to 2009 in order to pocket the scrappage premium. The same effect can also be seen from the middle and right panel of Figure 5, which yield the 'gaps' between actual and synthetic registration figures, i.e., the difference between actual and counterfactual numbers of registrations, and their cumulation over time, $\widehat{\text{Eff}_{\text{cum}}}(t) := \sum_{\tau \in [2009-01,t]} \widehat{\text{Eff}}_{1,\tau}$ which

[25]As this section mainly focuses on registrations, one might estimate an SCMT model for registrations only. We refrain from doing so, as the corresponding result only marginally deviates from the MSCMT result when the set of predictors is left unchanged.



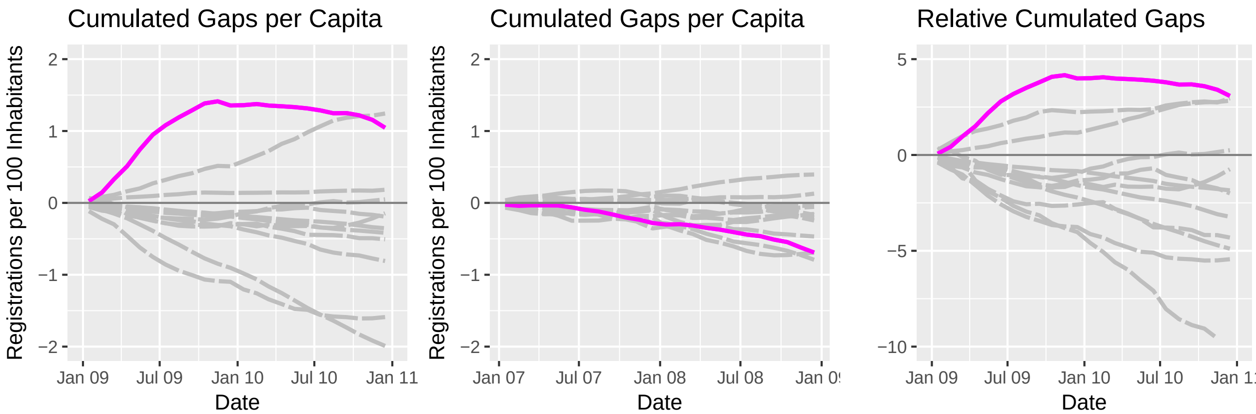
Note: Left panel: comparison of actual (black, solid) and synthetic (red, dashed) registration figures for Germany. Middle panel: difference between actual and synthetic registration figures for Germany (gaps). Right panel: cumulated gaps over time.

Figure 5: New Passenger Car Registrations Results for Germany: Transition from MSCMT Results to Cumulated Gap Results

is an estimate of the cumulated effect on registrations, $\mathrm{Eff}_{\mathrm{cum}}(t) := \sum_{\tau \in [2009-01,t]} \mathrm{Eff}_{1,\tau}$.

One can see from the figure that the gaps are positive during 2009, while they become negative during 2010, and that the positive difference in 2009 is much larger than the negative one in the next period. Correspondingly, the cumulated gaps drastically increase during 2009, while they decrease much less in 2010, until they stabilize at a level of approximately 850,000 at the end of that year. Therefore, it is evident that the long-run effect of the intervention consists of approximately 850,000 sales, which would not have been realized in the absence of a Cash-for-Clunkers program (on-top sales). Note that this is quite different from the U.S. case, for which the literature finds that 100% of program-induced purchases happened on cost of future periods (see, e.g., Mian and Sufi, 2012), resulting in a vanishing long-run impact. In our case, however, the positive overall long-term effect of the scrappage premium is actually significant. Figure 6 shows that Germany's cumulated gaps per capita are significantly above those of the placebo countries, especially when relating them to the root mean squared prediction error in the validation period.

At its peak in late 2009, the cumulated effect was even close to 1.16 million cars, implying that there was also a strong pull-forward effect, which can be roughly read off as the peak minus the long run value. Visual inspection suggests that the scrappage program led to a decrease of approximately 310,000 purchases between late 2009 and December 2010. Yet, we have to take into consideration that the intervention did not only incentivize people to pull forward



Note: Left & middle panel: cumulated gaps per capita for Germany (magenta, solid) and placebos (gray, dashed) during estimation window (left) and validation window (middle). Right panel: cumulated gaps in estimation window relative to average cumulated gap size in validation window.

Figure 6: Cumulated Sales Effect of German Scrappage Program

purchases, but also that it potentially shifted demand for non-subsidized (or "regular") cars from the program period to later times. In fact, car dealers report that, during the peak of the program period, demand was so strong that people sometimes would have had to accept long waiting times with lines ranging until outside the dealerships. Additionally, empirical evidence shows that throughout the program period non-subsidized buyers were price discriminated against when purchasing large, i.e. expensive, cars compared with subsidized people (see Kaul et al., 2016). Therefore, potential car buyers, not being eligible for the subsidy but aware of such facts, might have decided to delay their purchases until the policy intervention had ended. All in all, the program therefore implied two different effects with respect to shifting demand in time, which partially offset each other. On the one hand, people pulling forward their purchases in order to profit from the scrappage premium. On the other hand, people delaying their regular purchase until the program had run off. The above mentioned difference of ca. 310,000 purchases therefore equals the difference between pull-forward and delayed purchases. Without disentangling these two effects, we can only conclude that at least 310,000 purchases were pulled forward and that at most 770,000 (≈ 1,933,000-850,000-310,000) people profited from windfall gains, since they had intended to buy a new car anyway.

A Non-Linear Model

In order to estimate the amounts of pull-forward and delayed purchases, we build a non-linear model to describe the timeline of the program's cumulated effect as shown in Figure 5's right panel. First of all, we measure the time elapsed in months since 2009-01-01, and consider time points $t \in [0, 32]$, corresponding to the time span from January 2009 through August 2011.[26] By introducing the parameter $0 \leq t_0$, which describes the start of the intervention's effect on registrations, we allow for some delay between the official start of the program and its taking effect on registrations. Similarly, we employ t_1 ($\geq t_0$) to denote the point in time when the effect of on-top sales on registration figures ends.

We model by $\mathrm{Eff}^{(\mathrm{o})}(t) := \kappa_0 \dfrac{t_1 - t}{t_1 - t_0} + \kappa_1 \dfrac{t - t_0}{t_1 - t_0}$ for $t_0 \leq t \leq t_1$ the rate of on-top (additional) purchases during the program's lifetime. Here, κ_0 and κ_1 denote the rates at which on-top registrations occur at the program's start and end, respectively. The timeline of $\mathrm{Eff}^{(\mathrm{o})}(t)$ is given in the left panel of Figure 7.[27] The cumulated effect of on-top purchases, $\mathrm{Eff}^{(\mathrm{o})}_{\mathrm{cum}} := \int_0^t \mathrm{Eff}^{(\mathrm{o})}(\tau) d\tau$, can be written as

$$\mathrm{Eff}^{(o)}_{\mathrm{cum}} = \beta^{(\mathrm{o})} \widetilde{\mathrm{lin}}_{t_0,t_1} + (\kappa_0 - \kappa_1) \widetilde{\mathrm{quad}}_{t_0,t_1},$$

with $\beta^{(\mathrm{o})} := \frac{t_1 - t_0}{2} (\kappa_0 + \kappa_1)$ denoting the overall amount of additional purchases[28]. The corresponding timeline is given by the blue, dotted line in the right panel of Figure 7.

With respect to pull-forward sales, we denote by t_4 the furthest point in time from which purchases were pulled forward, and by $\mathrm{sh}^{(\mathrm{p})}(s, t)$ the rate of pulling forward from time s to time t. More precisely, for small time spans Δ_s and Δ_t, we assume that the amount of purchases pulled forward from $[s, s + \Delta_s]$ to $[t, t + \Delta_t]$ is approximately given by $\mathrm{sh}^{(\mathrm{p})}(s, t) \cdot \Delta_s \cdot \Delta_t$. Pulling forward takes place only for $t_0 \leq t \leq t_1 \leq s \leq t_4$, and we model it by $\mathrm{sh}^{(\mathrm{p})}(s, t) :=$

[26] We discard later observations because the effects of the program can safely be expected to have worn off in September 2011 due to institutional reasons discussed below (Subsection 'Effects on Sales'). However, the results of our analysis do not change when using all 48 months.

[27] All graphics relate to the parameters as estimated by our model.

[28] For $q < r$, we denote by $\widetilde{\mathrm{lin}}_{q,r}$ the piecewise linear function which takes the value 0 to the left of q and the value 1 to the right of r. Similarly, $\widetilde{\mathrm{quad}}_{q,r}$ is defined as the piecewise quadratic function vanishing outside of $[q, r]$ with local maximum $\frac{r-q}{8}$ at $\frac{q+r}{2}$. Formally, $\widetilde{\mathrm{lin}}_{q,r}(t) = \frac{t-q}{r-q}$ and $\widetilde{\mathrm{quad}}_{q,r}(t) = \frac{1}{2} \frac{(t-q)(r-t)}{r-q}$ for $q \leq t \leq r$.

$\mu^{(\mathrm{p})} + \lambda^{(\mathrm{p})} (t_4 - t_0) - \lambda^{(\mathrm{p})} (s - t)$. Thereby, $\mu^{(\mathrm{p})} \geq 0$ gives the rate of pull-forward sales for the maximal time lag between s and t ($s - t = t_4 - t_0$), while $\lambda^{(\mathrm{p})} \geq 0$ describes by how much the amount of pulling forward decreases with increasing time lag $s - t$. The term $-\lambda^{(\mathrm{p})} (s - t)$ is included to capture a potential decrease in the rate of pulling forward due to increased costs of this shifting in time with increasing time span of pulling forward. The rate of pull-forward sales, $\mathrm{sh}^{(\mathrm{p})}$, is visualized as the green, dashed line in the middle panel of Figure 7, for the special case of $\lambda^{(\mathrm{p})} = 0$.[29] Denoting by $\mathrm{Eff}^{(\mathrm{p})}(t) := \int_0^{32} \left(\mathrm{sh}^{(\mathrm{p})}(s,t) - \mathrm{sh}^{(\mathrm{p})}(t,s)\right) ds$ the net effect of pull-forward sales at time t, we find for the cumulated effect of pull-forward purchases, $\mathrm{Eff}^{(\mathrm{p})}_{\mathrm{cum}}(t) := \int_0^{t} \mathrm{Eff}^{(\mathrm{p})}(\tau) d\tau$:[30]

$$\mathrm{Eff}^{(\mathrm{p})}_{\mathrm{cum}} = \beta^{(\mathrm{p})} \, \mathrm{lin}_{t_0,t_1,t_4} - \lambda^{(\mathrm{p})} \, \mathrm{quad}_{t_0,t_1,t_4},$$

with $\beta^{(\mathrm{p})} := (t_4 - t_1)(t_1 - t_0)\left(\mu^{(\mathrm{p})} + \lambda^{(\mathrm{p})} \frac{t_1 - t_0}{2}\right)$ denoting the overall amount of pull-forward purchases. The corresponding timeline is given as the green, dashed line in the right panel of Figure 7.

For the group of people who delayed their purchases due to the scrappage program, we assume that they initially wanted to buy in $[t_0, t_2]$, and that their delayed purchases were rectified during $[t_2, t_3]$. We expect that $t_1 \leq t_2$, since these buyers postponed their purchases until the hype about Cash-for-Clunkers had worn off. We further expect that $t_3 \leq t_4$, since the scrappage premium was a rather strong monetary incentive to pull forward purchases that would have happened much later without such a subsidy, while people delaying their purchases due to the intervention had no more incentive for postponing further after the program had ended.

To model the delaying of purchases, we denote by $\mathrm{sh}^{(\mathrm{d})}(s,t)$ the rate of shifting from time s to time t. Delaying only takes place for $t_0 \leq s \leq t_2$, and delayed sales happen at

[29] Footnote 27 also applies here.

[30] For $q < r < s$, $\mathrm{lin}_{q,r,s}$ denotes the piecewise linear function which takes the value 1 at r and vanishes outside of $[q,s]$: for $q \leq t \leq r$, it takes the value $\frac{t-q}{r-q}$, while it equals $\frac{s-t}{s-r}$ for $r \leq t \leq s$. Similarly, $\mathrm{quad}_{q,r,s}$ is defined as the piecewise quadratic function vanishing outside of $[q,s]$ with local minima of $-\frac{(r-q)^2(s-r)}{8}$ and $-\frac{(q-r)(r-s)^2}{8}$ at $\frac{q+r}{2}$ and $\frac{r+s}{2}$, respectively. Formally, $\mathrm{quad}_{q,r,s}(t)$ is defined as $-\frac{1}{2}(s-r)(t-q)(r-t)$ for $q \leq t \leq r$, and as $-\frac{1}{2}(r-q)(t-r)(s-t)$ for $r \leq t \leq s$.

$t_2 \leq t \leq t_3$. For such s, t, we assume that the rate of delaying is given by $\mathrm{sh}^{(\mathrm{d})}(s,t) := \mu^{(\mathrm{d})} + \lambda^{(\mathrm{d})}(t_3 - t_0) - \lambda^{(\mathrm{d})}(t - s)$, where $\mu^{(\mathrm{d})} \geq 0$ gives the rate of delaying for the maximal time lag between s and t, $t_3 - t_0$, while $\lambda^{(\mathrm{d})} \geq 0$ describes by how much the rate of delaying decreases when the time lag $t - s$ increases. Intuitively, we subtract the term $\lambda^{(\mathrm{d})}(t-s)$ since it is costly to postpone one's purchase: the larger the time difference, the less likely it becomes to delay a purchase to time t, which, without the scrappage program, would have happened at time s. The rate of delaying, $\mathrm{sh}^{(\mathrm{d})}$, is visualized as the grey, dashed-dotted line in the middle panel of Figure 7. The net effect of delayed purchases at time t, given by $\mathrm{Eff}^{(\mathrm{d})}(t) := \int_0^{32} \left(\mathrm{sh}^{(\mathrm{d})}(s,t) - \mathrm{sh}^{(\mathrm{d})}(t,s)\right) ds$, can be used to calculate the cumulated effect of delayed purchases at time t: $\mathrm{Eff}^{(\mathrm{d})}_{\mathrm{cum}}(t) := \int_0^{t} \mathrm{Eff}^{(\mathrm{d})}(\tau) d\tau$. Straightforward calculations show that $\mathrm{Eff}^{(\mathrm{d})}_{\mathrm{cum}}(t)$ can be written as follows:

$$\mathrm{Eff}^{(\mathrm{d})}_{\mathrm{cum}} = -\beta^{(\mathrm{d})}\, \mathrm{lin}_{t_0,t_2,t_3} + \lambda^{(\mathrm{d})}\, \mathrm{quad}_{t_0,t_2,t_3},$$

with $\beta^{(\mathrm{d})} := (t_3 - t_2)(t_2 - t_0)\left(\mu^{(\mathrm{d})} + \lambda^{(\mathrm{d})} \frac{t_3 - t_0}{2}\right)$ denoting the overall amount of delayed purchases. The corresponding timeline is given by the grey, dashed-dotted line in the right panel of Figure 7.

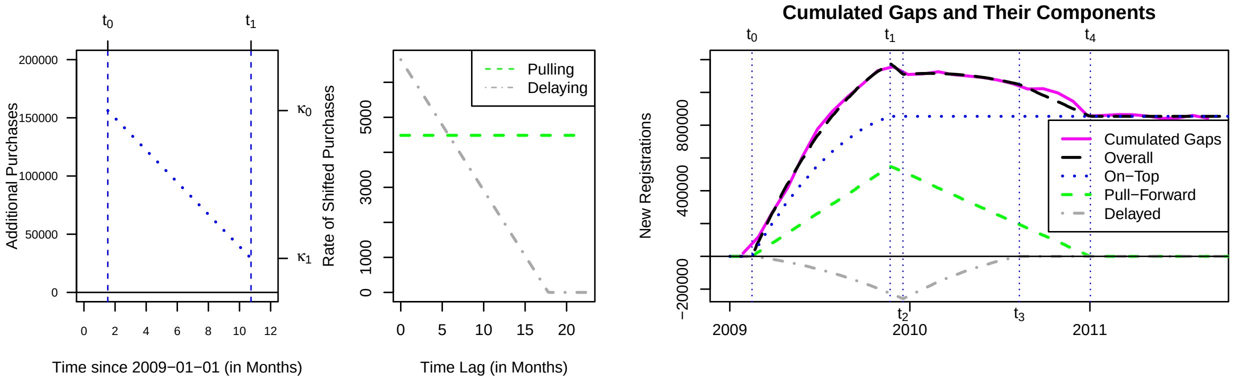
All in all, we arrive at the following model for the scrappage program's cumulated effect on registrations, $\mathrm{Eff}_{\mathrm{cum}} := \mathrm{Eff}^{(\mathrm{o})}_{\mathrm{cum}} + \mathrm{Eff}^{(\mathrm{p})}_{\mathrm{cum}} + \mathrm{Eff}^{(\mathrm{d})}_{\mathrm{cum}}$:[31]

$$\begin{aligned}\mathrm{Eff}_{\mathrm{cum}} = \beta^{(\mathrm{o})} \widetilde{\mathrm{lin}}_{t_0,t_1} + \beta^{(\mathrm{p})}\, \mathrm{lin}_{t_0,t_1,t_4} - \beta^{(\mathrm{d})}\, \mathrm{lin}_{t_0,t_2,t_3} \\ + (\kappa_0 - \kappa_1) \widetilde{\mathrm{quad}}_{t_0,t_1} - \lambda^{(\mathrm{p})}\, \mathrm{quad}_{t_0,t_1,t_4} + \lambda^{(\mathrm{d})}\, \mathrm{quad}_{t_0,t_2,t_3}.\end{aligned} \tag{13}$$

Effects on Sales

Equation (13), a function of the parameters $t_0, t_1, t_2, t_3, t_4, \kappa_0, \kappa_1, \mu^{(\mathrm{p})}, \lambda^{(\mathrm{p})}, \mu^{(\mathrm{d})}, \lambda^{(\mathrm{d})}$, is non-linear in t_0, t_1, t_2, t_3, t_4, and linear in the others. These quantities are restricted by $0 \leq t_0 \leq t_1 \leq t_2 \leq t_3 \leq t_4$, as well as by $\kappa_0, \kappa_1, \mu^{(\mathrm{p})}, \lambda^{(\mathrm{p})}, \mu^{(\mathrm{d})}, \lambda^{(\mathrm{d})} \geq 0$. For estimating these parameters,

[31]Note that there is no timeline for the effect of windfall gains, as these happen under both the actual and counterfactual scenario, implying a net effect of zero.



Note: Left panel: arrival rate of additional purchases as a function of time. Middle panel: rates of time-shifting for pull-forward and delayed sales as a function of the time amount by which the corresponding purchase is shifted. Right panel: cumulated gaps (magenta solid line) and their components. The timeline of on-top sales is represented by a blue dotted line, the green dashed-dotted line shows pull-forward sales, the gray long-dashed line highlights delayed sales, while the black long-dashed line depicts the sum thereof and shows the fitted values of our non-linear model.

Figure 7: Cumulated On-Top, Pull-Forward, Delayed, and Overall Program-Induced Sales

we approximate $\mathrm{Eff}_{\mathrm{cum}}$ by $\widehat{\mathrm{Eff}}_{\mathrm{cum}}$ and apply non-linear least squares, with $\widehat{\mathrm{Eff}}_{\mathrm{cum}}(t)$ denoting the cumulated differences between actual and synthetic Germany as defined above and plotted in the right panel of Figure 5. The corresponding results are presented in the right panel of Figure 7 and Table 2.

Denoting the residuals from the non-linear estimation by $\widehat{\widehat{\mathrm{Eff}}}_{\mathrm{cum}}$, the estimation yields a root mean square error between $\widehat{\mathrm{Eff}}_{\mathrm{cum}}$ and $\widehat{\widehat{\mathrm{Eff}}}_{\mathrm{cum}}$ of roughly 12,000 cars. Thus, the fit provided by the non-linear model is quite good. This can also be seen from the right panel of Figure 7, where the cumulated effect derived from the MSCMT results (magenta, solid line) is precisely approximated by the fitted values of the non-linear model (black, long-dashed line).

With respect to the time points, we find that program-induced registrations started mid-February, 2009 ($t_0 = 1.53$), while the effect ended in late November, 2009 ($t_1 = 10.74$). The delayed (non-subsidized) purchases started to be rectified shortly after, between mid-December, 2009, and early August, 2010 ($t_2 = 11.58, t_3 = 19.33$). Furthermore, the scrappage premium resulted in pull-forward purchases which otherwise would have been realized up to the very end of 2010 ($t_4 = 24.04$). Thus, purchases have been pulled forward over a time span of almost exactly two years ($t_4 - t_0 = 22.51$ months), with purchases that were originally planned for as late as December 2010, but realized already in early 2009. The finding of this time span being *two* years is very plausible, since it precisely reflects specific requirements

Table 2: Results of Non-Linear Model

Quantity	Estimate	Quantity	Estimate
t_0	1.53	$\mu^{(\mathrm{p})}$	4,485.68
t_1	10.74	$\lambda^{(\mathrm{p})}$	0.00
t_2	11.58	$\mu^{(\mathrm{d})}$	0.00
t_3	19.33	$\lambda^{(\mathrm{d})}$	373.47
t_4	24.04	$\beta^{(\mathrm{o})}$	853,864
κ_0	156,213.38	$\beta^{(\mathrm{p})}$	549,579
κ_1	29,220.00	$\beta^{(\mathrm{d})}$	258,975
$\kappa_0 - \kappa_1$	126,993.38	$\beta^{(\mathrm{w})}$	529,647

Note: Superscripts p, d, o, and w stand for the sales groups of pull-forward, delayed, on-top, and windfall, respectively. Time points t_0, t_1, t_2, t_3, t_4 are measured in months elapsed since 2009-01-01 and denote the start and end time of on-top sales, the start and end times of delaying, and the end time of pulling forward, respectively. κ_0 and κ_1 denote the arrival rate of on-top sales at the program's start and end, respectively. μ gives the rate of delaying and pulling forward at the maximal time lags, $t_3 - t_0$ and $t_4 - t_0$, respectively. λ denotes by how much the rates of delaying and pulling forward decrease with increasing time lag, respectively, while β denotes the overall amounts of on-top, pull-forward, delayed, and windfall sales, respectively.

that car holders must meet in Germany: after an initial period of three years following the initial vehicle purchase, passenger cars must be presented for inspection at officially approved facilities every *two* years (similar to the MOT test in the UK). Without passing such an inspection, cars must not be used on public roads. Especially when running an old vehicle, these inspections are typically quite costly, as they usually involve replacing or repairing some parts of the car. Thus, people who planned to continue driving their old car for more than two years either owned a vehicle for which they anticipated no or only small costs of replacing and repairing in 2009 or 2010, or they owned a vehicle for which investing into it was economically sensible. In both cases, the old car's worth probably exceeded €2,500, so that it did not make sense to scrap it off.

For the group of on-top sales, i.e., purchases only happening due to the intervention, we find that they were realized at a rate of roughly 155,000 per month at the start of the program (κ_0). At the end of the program, however, this rate had dropped to an estimate of roughly 29,000 per month (κ_1). This translates into an estimate of approximately 850,000 sales induced by the scrappage program ($\beta^{(o)}$), completely in line with the results obtained by visual inspection of Figure 5's right panel and discussed above. Extrapolating those program induced on-top purchases by using the mean price for German cars bought under the scrappage program (€20,000 according to BAFA, 2010) implies a notable multiplier effect with respect to the original €5 billion budget—it more than triples.

With respect to the group of pulled forward purchases, we find $\lambda^{(\mathrm{p})}$ to be estimated as zero. Thus, the rate of pull-forward sales is constant over time and given by $\mu^{(\mathrm{p})}$ =4,485.68. Therefore, from every month between December, 2009, through December, 2010, ca. 4,500 purchases were pulled forward to every month during the program period from February, 2009, through November, 2009. Overall, this translates into approximately 550,000 pull-forward purchases ($\beta^{(\mathrm{p})}$).

The rate of postponing by only a marginally amount of time is well above 6,000 (see the mid panel of Figure 7), and it shrinks to zero when the maximal time span of delaying, $t_3 - t_0 = 17.8$, is reached. Overall, this translates into almost 260,000 purchases ($\beta^{(\mathrm{d})}$) of non-subsidized cars that were postponed in order to avoid the time span during which the Cash-for-Clunkers program was in force.

Although no part of Model (13) is concerned with windfall gains, we can estimate the amount of this group indirectly by subtracting the amounts of on-top and pull-forward sales from the overall number of subsidized purchases. We thus find an estimate of almost 530,000 ($\beta^{(\mathrm{w})} :=$ 1,933,090 $-(\beta^{(\mathrm{o})} + \beta^{(\mathrm{p})})$), showing that more than 25% of the program's budget were wasted to subsidize people who had wanted to buy a new car anyway. Note that if we had ignored the possibility of delayed, non-subsidized purchases, we would have wrongly concluded that only 310,000 sales were pulled forward in time, and, as a consequence, we would have overestimated the number of windfall gains by 240,000.

Effects on the Environment: Back-of-the-Envelope Calculations

For the years 2009 and 2010, the time during which the Cash-for-Clunkers program was effective, Table 3 displays actual and counterfactual numbers of new registrations as well as corresponding CO2 emissions, and combines them into CO2 emissions per kilometer driven of all newly registered cars in 2009 and 2010. Again, we find that the actual, overall amount of registrations (6,681,693) exceeds the corresponding counterfactual amount (5,825,610) by the number of on-top purchases (856,083), i.e., the number of purchases that happened only

Table 3: Results for Registrations and Emissions

	Actual (CC)			Counterfactual (Non-CC)		
	2009	**2010**	**2009–2010**	**2009**	**2010**	**2009–2010**
Registrations	3,794,418	2,887,275	6,681,693	2,685,181	3,140,429	5,825,610
∅ CO2	154	151.1		158.53	148.86	
CO2/km	584,340,372	436,267,253	1,020,607,625	425,680,401	467,481,749	893,162,150

Note: Actual and counterfactual numbers of new registrations, CO2 emissions per average newly registered car (g/km), and CO2 emissions of all newly registered cars (g/km), for 2009, 2010, and its difference where applicable.

because of the subsidy.[32] With respect to CO2 emissions per kilometer driven, the actual and counterfactual pool of newly registered cars in 2009 and 2010 differ by 1,020,607,625 $-$ 893,162,150 = 127,445,475 g/km, i.e., by roughly 127 tons per kilometer. Thus, if we were to assume that non-subsidized buyers as well as new car owners who shifted their purchase in time were not influenced by the scrappage program with respect to the type of car they bought, then the difference in emissions could only stem from on-top sales. We would then deduce that these vehicles emit 127,445,475 / 853,864 = 148.87 gram of CO2 per kilometer driven. However, this number does not seem plausible: average CO2 emissions per kilometer driven of the two million subsidized cars were only 142 g/km, as reported by official institutions (IFEU, 2009). Moreover, if on-top sales emitted 148.87 g/km, then the groups of pull-forward purchases and windfall gains would belong to cars which emit only (1,933,090 $\times$ 142 - 853,864 $\times$ 148.87) / (1,033,090 - 853,864) = 136.56 g/km. Hence, the rather implausible conclusion would be that cars which had been purchased only because they were subsidized and consequently can be assumed to be rather small vehicles characterized by low emission figures, emitted more CO2 than cars that would have been bought anyway and thus can be assumed to be rather larger, more polluting cars.

If we adopt a more plausible, but still conservative assumption, namely that the cars bought on-top are not different from the cars which would have been purchased anyway, then the cars of all these groups emit 142 g/km on average. As a consequence, the difference between actual and counterfactual pollution, 127,445,475 g/km, is only partially explained by the emissions of on-top purchases, as these then only account for 853,864 $\times$ 142 = 121,248,688

[32] Recall that from the disentangling above, we obtained 853,864 for the number of on-top sales, re-assuring the validity of our model.

g/km. The difference of 127,445,475 − 121,248,688 = 6,196,787 g/km must consequently stem from 'upgrading' by buyers who wanted to buy a new car in 2009–2010 anyway, i.e., from using the scrappage premium to buy larger, more polluting cars than initially planned. More precisely, the average car of the pull-forward and windfall groups would in the absence of the intervention have emitted 6,196,787 / (1,933,090 − 853,864) = 5.74 g/km less than it actually does. Eventually, we consider an even more plausible assumption, namely that on-top sales on average emit 135 g/km, slightly less than the overall average of all subsidized vehicles of 142 g/km. Under this scenario, we arrive at an average upgrading by the groups of pull-forward and windfall of (127,445,475 − 853,864 × 135) / (1,933,090 − 853,864) = 11.28 g/km.

With respect to pull-forward sales only, there is an immediate reduction of emissions due to scrapping the old car and driving a new, typically more eco-friendly vehicle. According to official institutions (IFEU, 2009), scrapped cars on average emitted 200 g/km, considerably more than the subsidized new cars' average value of 142 g/km. As purchases on average were pulled forward by roughly one year, the reduction will equal 58 g per km driven for 550,000 cars over one year. However, this is a very short-lived effect that reduces overall emissions only during the one-year time span over which the purchase was pulled forward. For the group of windfall gains, there is even no such effect at all, as they would have bought a new car in 2009 anyway. However, there is considerable evidence that people who wanted to buy a new car in 2009 or 2010 (pull-forward and windfall groups) used the premium of €2,500 to acquire larger, more polluting cars. Over these new cars' lifetime, 15 years on average in Germany, this upgrading will lead to an increase of emissions. Even under the conservative scenario discussed above, this would amount to additional emissions of 5.74 g per km driven for roughly 1.1 million cars over 15 years. This is almost three times the amount of emissions saved by the group of pull-forward sales by replacing their clunkers early with new, eco-friendly cars. Thus, it might well be the case that the Cash-for-Clunkers program was not benefiting the environment, but even hurting it.[33]

[33]Note that it is essentially impossible to tell what effect the group of on-top purchases caused on the environment. There is no information available whether and when these car owners in absence of the intervention would have bought a used car, or whether they would have retired their old car completely, falling back to public transport, motorbikes, or some other means of transport.

6 Conclusion

Car scrappage subsidies are supposed to smooth negative impacts on one of the most important markets within the economy. Since vehicle retirement plans typically involve goals beyond hard sales figures—above all, environmental aspects—and demand a substantial budget, a sound decision on a particular policy intervention becomes not only more difficult but also potentially more expensive. In order to being able to examine the impact of scrapping schemes, in our case on car purchases and respective CO2 emissions in Germany, it is key to compute reliable counterfactuals. To this end, we develop MSCMT, the multivariate synthetic control method using time series of economic predictors, and combine this approach with a cross-validation to select optimal predictor weights. We further show that the MSCMT effect estimator is (asymptotically) unbiased under quite general conditions regarding the data generating process.

We find that the policy intervention indeed provided substantial short-run and long-run stimulus to the new car market. Being the number one European country in terms of employees within the automotive market and regarding new passenger cars production, the German scrappage program distinctly stabilized one of its most important markets and its overall economy. It did so by not only borrowing car purchases from the future but also by generating a large amount of additional sales on top of regular ones within the program period. Technically disentangling our estimates reveals that out of the 2 million subsidized vehicles about 530,000 would have been purchased anyway (windfall gains); without the intervention, roughly 550,000 cars would have been bought at a later point in time (pull-forward effects), while about 260,000 non-subsidized sales have been delayed due to the intervention; finally, slightly more than 850,000 vehicle sales would not have been realized at all without the introduction of the scrappage subsidy (on-top sales), amounting to a value of roughly €17 billion.

With respect to greenhouse gas emissions, however, analyzing the effects of the scrappage program yields a more complicated picture. Because of the pulling forward of about half a million purchases, the actual average CO2 emissions of newly registered cars in 2010 were larger than they would have been without the intervention, as purchases of these small, eco-friendly vehicles now took place during the program in 2009, but not in 2010. This reasoning also implies that actual CO2 emissions of newly registered cars in 2009 should have been much lower than their counterfactual equivalent. However, we do not find a statistically significant decrease of average CO2 emissions of new vehicles in 2009. This is probably because consumers profiting from windfall gains and consumers pulling forward their purchases were lured by the €2,500 subsidy into buying larger, less eco-friendly cars, and the corresponding increase in average CO2 emissions at least partially offsets the above mentioned decrease. Therefore, the scrappage premium had at least three effects, working in different directions. First, a reduction of emissions due to earlier replacing of old, heavily polluting cars. Second, a corresponding increase of emissions because many non-subsidized buyers postponed their purchases, thereby slightly prolonging the lifetime of their old cars. And third, an increase of emissions due to several subsidized buyers using parts of the scrappage premium to buy larger, more polluting cars, which probably was the largest of these effects.

To sum up, we find that the German scrappage program considerably backed its domestic economy by boosting vehicle sales, albeit at the potential cost of hurting the environment by increasing greenhouse gas emissions. Because of these economic and ecological effects working in very opposite directions, it is not trivial to give clear advice for car retirement schemes to come. We have seen that particularly the category of windfall gains led to undesirable consequences which are twofold. First, it wastes quite a share (one fourth) of the subsidy funds, since those cars would have also been purchased in absence of a scrappage program. Second, it hurts the environment by producing a substantial amount of additional CO2 emissions due to a subsidy-induced vehicle upgrade. In practice, such effects probably cannot be completely avoided, because it is virtually impossible to dis-incentivize people profiting from windfall gains. However, in times of crisis, policy makers should keep in mind that people who postpone their purchases today constitute tomorrow's group of windfall gains. Thus, if one opts for a scrappage program, it should be introduced as early as possible. Then, the replacement of old, heavily polluting cars will not be delayed during the crisis' early stage. This is not only directly beneficial for the environment, but also indirectly by preventing the piling up of old cars, i.e., the growth of the future group of windfall gains.

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., Diamond, A., and Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2):495–510.

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.

Adda, J. and Cooper, R. (2000). Balladurette and Juppette: A Discrete Analysis of Scrapping Subsidies. *Journal of Political Economy*, 108(4):778–806.

BAFA (2010). Abschlussbericht – Umweltprämie. `http://www.bafa.de/bafa/de/wirtschaftsfoerderung/umweltpraemie/publikationen/ump_abschlussbericht.pdf` (last accessed on 2016-03-07).

Becker, M. and Klößner, S. (2017). Estimating the economic costs of organized crime by synthetic control methods. *Journal of Applied Econometrics*, 32(7):1367–1369.

Becker, M. and Klößner, S. (2018a). Fast and reliable computation of generalized synthetic controls. *Econometrics and Statistics/The Annals of Computational and Financial Econometrics*, 5:1–19.

Becker, M. and Klößner, S. (2018b). *MSCMT: Multivariate Synthetic Control Method Using Time Series*. R package version 1.3.3.

Becker, M., Klößner, S., and Pfeifer, G. (2018). Cross-validating synthetic controls. *Economics Bulletin*, 38(1):603–609.

Botosaru, I. and Ferman, B. (2017). On the role of covariates in the synthetic control method. Working Paper.

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.

Chernozhukov, V., Wuthrich, K., and Zhu, Y. (2017). An exact and robust conformal inference method for counterfactual and synthetic controls. Working Paper.

Copeland, A. and Kahn, J. (2013). The Production Impact of 'Cash-for-Clunkers': Implications for Stabilization Policy. *Economic Inquiry*, 51(1):288–303.

Doudchenko, N. and Imbens, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. NBER Working Paper No. 22791.

Ferman, B. and Pinto, C. (2017). Placebo tests for synthetic controls. Working Paper.

Ferman, B., Pinto, C., and Possebom, V. (2017). Cherry Picking with Synthetic Controls. MPRA Paper 78213, University Library of Munich, Germany.

Feyrer, J. D. and Sacerdote, B. (2011). Did the Stimulus Stimulate? Real Time Estimates of the Effects of the American Recovery and Reinvestment Act. NBER Working Papers 16759, National Bureau of Economic Research, Inc.

Firpo, S. and Possebom, V. (2017). Synthetic control method: Inference, sensitivity analysis and confidence sets. Working Paper.

Gardeazabal, J. and Vega-Bayo, A. (2017). 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.

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.

Grigolon, L., Leheyda, N., and Verboven, F. (2016). Scrapping subsidies during the financial crisis — Evidence from Europe. *International Journal of Industrial Organization*, 44:41–59.

Hahn, R. W. (1995). An Economic Analysis of Scrappage. *The RAND Journal of Economics*, 26(2):222–242.

Hoekstra, M., Puller, S. L., and West, J. (2017). Cash for corollas: When stimulus reduces spending. *American Economic Journal: Applied Economics*, 9(3):1–35.

House, C. L. and Shapiro, M. D. (2006). Phased-In Tax Cuts and Economic Activity. *The American Economic Review*, 96(5):1835–1849.

IFEU (2009). Abwrackprämie und Umwelt - eine erste Bilanz. Technical report, IFEU (Institut für Energie- und Umweltforschung Heidelberg GmbH).

Jinjarak, Y., Noy, I., and Zheng, H. (2013). Capital Controls in Brazil–Stemming a Tide with a Signal? *Journal of Banking and Finance*, 37(8):2938–2952.

Kaul, A., Klößner, S., Pfeifer, G., and Schieler, M. (2015). Synthetic Control Methods: Never Use All Pre-Intervention Outcomes Together With Covariates. Working paper.

Kaul, A., Pfeifer, G., and Witte, S. (2016). The Incidence of Cash for Clunkers: An Analysis of the 2009 Car Scrappage Scheme in Germany. *International Tax and Public Finance*, 23(6):1093–1125.

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.

Klößner, S., Kaul, A., Pfeifer, G., and Schieler, M. (2018). Comparative Politics and the Synthetic Control Method Revisited: A Note on Abadie et al. (2015). *Swiss Journal of Economics and Statistics*, 154(11).

Knittel, C. R. (2009). The Implied Cost of Carbon Dioxide Under the Cash for Clunkers Program. Working Paper, Sloan School of Management, Massachusetts Institute of Technology.

Li, S., Linn, J., and Spiller, E. (2013). Evaluating 'Cash-for-Clunkers': Program Effects on Auto Sales and the Environment. *Journal of Environmental Economics and Management*, 65(2):175–193.

Licandro, O. and Sampayo, A. R. (2006). The Effects of Replacement Schemes on Car Sales: The Spanish Case. *Investigaciones Economicas*, 30(2):239–282.

Mian, A. and Sufi, A. (2012). The Effects of Fiscal Stimulus: Evidence from the 2009 Cash for Clunkers Program. *The Quarterly Journal of Economics*, 127(3):1107–1142.

Parker, J. A., Souleles, N. S., Johnson, D. S., and McClelland, R. (2013). Consumer Spending and the Economic Stimulus Payments of 2008. *The American Economic Review*, 103(6):2530–2553.

Powell, D. (2017). Synthetic control estimation beyond case studies: Does the minimum wage reduce employment? Working Paper.

Powell, D. (2018). Imperfect synthetic controls: Did the massachusetts health care reform save lives? Working Paper.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robbins, M. W., Saunders, J., and Kilmer, B. (2017). A Framework for Synthetic Control Methods With High-Dimensional, Micro-Level Data: Evaluating a Neighborhood-Specific Crime Intervention. *Journal of the American Statistical Association*, 112(517):109–126.

Sandler, R. (2012). Clunkers or Junkers? Adverse Selection in a Vehicle Retirement Program. *American Economic Journal: Economic Policy*, 4(4):253–281.

Schiraldi, P. (2011). Automobile Replacement: A Dynamic Structural Approach. *The RAND Journal of Economics*, 42(2):266–291.

Szwarcfiter, L., Mendes, F. E., and La Rovere, E. L. (2005). Enhancing the Effects of the Brazilian Program to Reduce Atmospheric Pollutant Emissions from Vehicles. *Transportation Research: Part D: Transport and Environment*, 10(2):153–160.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed efects models. *Political Analysis*, 25(1):57–76.

A Proofs

Proposition 3 (Restatement of Proposition 1)**.** *Denoting by T_0 the time of the intervention, we have:*

1. *the bias of the vector of effect estimators, $(Y_{11t} - \widehat{Y}_{11t}(W), \ldots, Y_{K1t} - \widehat{Y}_{K1t}(W))'$, conditional on $\mathcal{F}_{T_0}$, the information up to time T_0, is a linear function of the p approximation errors $e_{T_0}(W), \ldots, e_{T_0-(p-1)}(W)$,*
2. *if $T_0 \geq p$ and the weights $W^* = (w_2^*, \ldots, w_{J+1}^*)$ satisfy $X_{1,t} = \sum_{j=2}^{J+1} w_j^* X_{j,t}$ for $t = T_0, \ldots, T_0 - (p-1)$, then $E(e_t(W)|\mathcal{F}_{T_0}) = 0$ for all $t \geq T_0 + 1$, and the vector of effect estimators is unbiased.*

Proof. Equations (4) and (5) entail that, in order to prove the first assertion, it is enough to show that $E(e_t(W)|\mathcal{F}_{T_0}) = 0$ for $t > T_0$. We start by considering the case $t = T_0 + 1$, for which

$$e_{T_0+1}(W) = \sum_{h=1}^{p} A_{h,T_0+1} e_{T_0+1-h}(W) + \sum_{j=2}^{J+1} w_j \left(\varepsilon_{1,T_0+1} - \varepsilon_{j,T_0+1}\right).$$

The expectation of $\sum_{j=2}^{J+1} w_j \left(\varepsilon_{1,T_0+1} - \varepsilon_{j,T_0+1}\right)$ conditional on $\mathcal{F}_{T_0}$, is zero. Thus, the conditional expectation of $e_{T_0+1}(W)$ is given by

$$E\left(e_{T_0+1}(W)|\mathcal{F}_{T_0}\right) = E\left(A_{1,T_0+1}|\mathcal{F}_{T_0}\right) e_{T_0}(W) + \ldots + E\left(A_{p,T_0+1}|\mathcal{F}_{T_0}\right) e_{T_0-(p-1)}(W),$$

which completes the proof of the first assertion for the case $t = T_0 + 1$. For $t > T_0$, we can proceed by using induction, as $e_t(W)$ is given by the recursion $e_t(W) = \sum_{h=1}^{p} A_{h,t} e_{t-h}(W) + \sum_{j=2}^{J+1} w_j \left(\varepsilon_{1,t} - \varepsilon_{j,t}\right)$, where the conditional expectations of $\sum_{j=2}^{J+1} w_j \left(\varepsilon_{1,t} - \varepsilon_{j,t}\right)$ given $\mathcal{F}_{T_0}$ vanish for all $t > T_0$.

The second assertion follows immediately from the first one, as $X_{1,t} = \sum_{j=2}^{J+1} w_j^* X_{j,t}$ for $t = T_0, \ldots, T_0 - (p-1)$ is tantamount to $e_t(W^*) = 0$ for $t = T_0, \ldots, T_0 - (p-1)$. □

Proposition 4 (Restatement of Proposition 2)**.** *In the presence of unobserved confounders, if W^* is such that $X_{1,t} = \sum_{j=2}^{J+1} w_j^* X_{j,t}$ for $t = 1, \ldots, T_0$, the vector of effect estimators is asymptotically unbiased under some technical assumptions, i.e., the bias vanishes when the pre-treatment time span grows to infinity.*

Proof. For $t = 1, \ldots, T_0$, we have $e_t(W^*) = 0$ due to $X_{1,t} = \sum_{j=2}^{j+1} w_j^* X_{j,t}$. This implies $0 = \sum_{j=2}^{J+1} w_j^* \left(u_{1,t} - u_{j,t} + \varepsilon_{1,t} - \varepsilon_{j,t}\right)$ for $t = p+1, \ldots, T_0$ as well as $e_{T_0+1}(W^*) =$

$\sum_{j=2}^{J+1} w_j^* \left(u_{1,T_0+1} - u_{j,T_0+1} + \varepsilon_{1,T_0+1} - \varepsilon_{j,T_0+1}\right)$. For every $k \in \{1, \ldots, K\}$, considering the k-th component, we thus have $0 = \sum_{j=2}^{J+1} w_j^* \left(u_{1,t}^{(k)} - u_{j,t}^{(k)} + \varepsilon_{1,t}^{(k)} - \varepsilon_{j,t}^{(k)}\right)$ for $t = p+1, \ldots, T_0$ and $e_{T_0+1}(W^*)^{(k)} = \sum_{j=2}^{J+1} w_j^* \left(u_{1,T_0+1}^{(k)} - u_{j,T_0+1}^{(k)} + \varepsilon_{1,T_0+1}^{(k)} - \varepsilon_{j,T_0+1}^{(k)}\right)$. Inserting the definition of $u_{j,t}$, we get $u_{j,t}^{(k)} = \lambda_{k,t}\mu_{k,j}$, implying $0 = \sum_{j=2}^{J+1} w_j^* \left(\lambda_{k,t}\left(\mu_{k,1} - \mu_{k,j}\right) + \varepsilon_{1,t}^{(k)} - \varepsilon_{j,t}^{(k)}\right)$ for $t = p+1, \ldots, T_0$ as well as $e_{T_0+1}^{(k)}(W^*) = \sum_{j=2}^{J+1} w_j^* \left(\lambda_{k,T_0+1}\left(\mu_{k,1} - \mu_{k,j}\right) + \varepsilon_{1,T_0+1}^{(k)} - \varepsilon_{j,T_0+1}^{(k)}\right)$. Stacking terms for pre-treatment times $t = p+1, \ldots, T_0$ and denoting corresponding vectors and matrices by the subscript 'P', we get:

$$0 = \sum_{j=2}^{J+1} w_j^* \left(\lambda_{k,P}\left(\mu_{k,1} - \mu_{k,j}\right) + \varepsilon_{1,P}^{(k)} - \varepsilon_{j,P}^{(k)}\right) = \lambda_{k,P} \sum_{j=2}^{J+1} w_j^* \left(\mu_{k,1} - \mu_{k,j}\right) + \sum_{j=2}^{J+1} w_j^* \left(\varepsilon_{1,P}^{(k)} - \varepsilon_{j,P}^{(k)}\right),$$

with $\lambda_{k,P} \in \mathbb{R}^{(T_0-p)\times F_k}$ and $\varepsilon_{1,P}^{(k)}, \varepsilon_{j,P}^{(k)} \in \mathbb{R}^{T_0-p}$. From this, we find

$$\sum_{j=2}^{J+1} w_j^* \left(\mu_{k,1} - \mu_{k,j}\right) = -\left(\lambda_{k,P}'\lambda_{k,P}\right)^{-1} \lambda_{k,P}' \sum_{j=2}^{J+1} w_j^* \left(\varepsilon_{1,P}^{(k)} - \varepsilon_{j,P}^{(k)}\right),$$

which, inserted into the formula for $e_{T_0+1}^{(k)}(W^*)$, delivers

$$e_{T_0+1}^{(k)}(W^*) = \sum_{j=2}^{J+1} w_j^* \left(\varepsilon_{1,T_0+1}^{(k)} - \varepsilon_{j,T_0+1}^{(k)}\right) - \lambda_{k,T_0+1}\left(\lambda_{k,P}'\lambda_{k,P}\right)^{-1} \lambda_{k,P}' \sum_{j=2}^{J+1} w_j^* \left(\varepsilon_{1,P}^{(k)} - \varepsilon_{j,P}^{(k)}\right).$$

We decompose $e_{T_0+1}^{(k)}(W^*)$ as $e_{T_0+1}^{(k)}(W^*) = R_{1,T_0+1} + R_{2,T_0+1} + R_{3,T_0+1}$, with

$$\begin{aligned}
R_{1,T_0+1} &:= \lambda_{k,T_0+1}\left(\lambda_{k,P}'\lambda_{k,P}\right)^{-1} \lambda_{k,P}' \sum_{j=2}^{J+1} w_j^* \varepsilon_{j,P}^{(k)} \\
R_{2,T_0+1} &:= -\lambda_{k,T_0+1}\left(\lambda_{k,P}'\lambda_{k,P}\right)^{-1} \lambda_{k,P}' \varepsilon_{1,P}^{(k)} \\
R_{3,T_0+1} &:= \sum_{j=2}^{J+1} w_j^* \left(\varepsilon_{1,T_0+1}^{(k)} - \varepsilon_{j,T_0+1}^{(k)}\right).
\end{aligned}$$

Under some standard, technical conditions, stated throughout their proof, Abadie et al. (2010, p. 504) show that the expected values of R_{2,T_0+1} and R_{3,T_0+1} vanish, while the mean of R_{1,T_0+1} can be bounded by a function that goes to zero as the number of pre-treatment observations, T_0, increases. This concludes the proof for $t = T_0 + 1$, while for $t > T_0 + 1$, the result follows by induction, as $e_t(W^*)$ is a linear function of $e_{t-h}(W^*)$, $h = 1, \ldots, p$. □