Bridging Qualitative and Quantitative Methods in Organizational Research: Applications of Synthetic Control Methodology in the U.S. Automobile Industry

Adam R. Fremeth[‡], Guy L. F. Holburn[*], Brian K. Richter[†]

November 17, 2015

Abstract
We assess the utility of synthetic control, a recently developed empirical methodology, for applications in organizational research. Synthetic control acts as a bridge between qualitative and quantitative research methods by enabling researchers to estimate treatment effects in contexts with small samples or few occurrences of a phenomenon or treatment event. The method constructs a counterfactual of a focal firm, or other observational unit, based on an objectively-weighted combination of a small number of comparable but untreated firms. By comparing the firm's actual performance to its counterfactual replica without treatment, synthetic control estimates, under certain assumptions, the magnitude and direction of treatment effects. We illustrate and critique the method in the context of the U.S. auto industry by estimating (a) the effect of government intervention in Chrysler's management from 2009-2011 on its sales volumes, and (b) the impact of Toyota's 2010 'acceleration crisis' on Camry sales.

Keywords: Synthetic Control, Case Study Research, Qualitative Methods, Automobile Industry, Government Intervention, Product Recalls

[‡] Adam R. Fremeth. Assistant Professor of Business, Economics, and Public Policy; Ivey School of Business, University of Western Ontario, London, ON N6G 0N1 CANADA. afremeth@ivey.uwo.ca
[*] Guy L. F. Holburn. Suncor Chair in Energy Policy and Associate Professor of Business, Economics, and Public Policy; Ivey School of Business, University of Western Ontario, London, ON N6G 0N1 CANADA. gholburn@ivey.uwo.ca;
[†] Brian K. Richter. Assistant Professor of Business, Government, and Society; McCombs School of Business; University of Texas at Austin; Austin, TX 78705 USA. brian.richter@mccombs.utexas.edu

‡

1 INTRODUCTION

Organization scholars often debate the relative merits of qualitative and quantitative empirical methodologies when assessing causal relationships among phenomena of interest (Eisenhardt 1989; Gibbert et al. 2008; Shah and Corley 2006). Given few opportunities in organizational research to apply randomized experimental research designs, which are the scientific ideal, social scientists are typically left with non-random observational data, the properties of which can influence choices of methodological approaches for testing causal claims. In settings with small data samples, populations, or numbers of focal events - which limit statistical approaches to inference - qualitative case studies are often the preferred approach (March et al. 1991). While qualitative methods permit in-depth exploration of relationships in ways that quantitative techniques cannot achieve, they are not easily adapted to estimation of the direction or magnitude of a phenomenon's effects.

In this paper, we assess the utility and limitations of a recently developed empirical methodology, synthetic control (Abadie and Gardeazabal 2003), which acts as a bridge between, and complement to, qualitative and quantitative social science research methods. The synthetic control technique, which has taken root in economics and political science although not as yet in organizational research, is applicable in contexts where researchers want to estimate the impact of a phenomenon or event, but have only a limited sample size or data on a few occurrences of the phenomenon of interest. For example, organizational researchers may wish to determine the effects, if any, of NGO media campaigns on targeted firm sales when only one firm in an industry is targeted; the performance impact of a new management incentive structure adopted by just one subsidiary of a parent firm that owns similar subsidiaries in other jurisdictions; or the effect of vertical integration on organizational profitability when integration in the industry is a rare event. In each of these instances, 'small numbers' limit the feasibility of regression-based statistical analyses and inference.

The synthetic control method generates a counterfactual or clone of the treated unit based on a weighted average of a small number of other comparable units that were not subject to the treatment phenomenon. If the synthetic closely mimics the focal unit in both its descriptive characteristics and pre-treatment performance, any subsequent divergence in outcomes in the post-treatment period between the actual and synthetic units may, under certain assumptions, be attributed to the impact of the treatment. In this way, researchers can make inferences and quantify the magnitude and direction of treatment effects. The synthetic also objectively identifies which combination of untreated units in the sample is most comparable to the focal unit, and hence which candidates may be appropriate for in-depth comparative case study investigation.

In the next section we present an overview of the synthetic control method, its assumptions and limitations, and the situations in which it can be a useful tool for organizational research. We then illustrate its application in estimating the impact on organizational performance of two major events in the U.S. automobile industry: the impact of direct federal government intervention in Chrysler's management from 2009-2011, after the provision of TARP financing in late 2008, on Chrysler's sales; and the effect of Toyota's 2010 'acceleration crisis' on the Camry sub-brand subsequent sales.[1] The unique nature of these events and the small number of automobile firms active in the U.S. market limit or preclude difference-in-differences or matched sample models, but make synthetic control a methodological candidate, albeit subject to caveats, for quantifying their effects.

2 THE SYNTHETIC CONTROL METHOD

The synthetic control methodology has been utilized in economics, political science, and law to measure the effects of a variety of phenomena, including the impact of domestic terrorism on regional growth within the Basque region of Spain (Abadie and Gardeazabal 2003); economic liberalization on real GDP growth (Billmeier and Nannicini 2013); new tobacco tax policy on cigarette sales in California (Abadie et al. 2010); reunification of Germany on per-capita wealth in the former West Germany (Abadie et al. 2015); and gun control laws on crime (Donohue and Aneja 2012). Despite the breadth of topics studied, the questions researched using synthetic control are focused around a single event or application of a treatment phenomenon, and cannot easily be answered with case study analyses due to the absence of obvious stand-alone counterfactuals among the few untreated units.

2.1 Overview

The synthetic control technique creates a counterfactual observation unit, such as a firm, corporate division or employee, whose performance can be compared to a focal unit that has undergone an event or treatment. The counterfactual unit, or synthetic control, is constructed as a weighted average of mathematically-selected untreated comparison or control units. The technique maximizes the ability of the synthetic unit to perform (generate outcome data) as if it were the focal unit had it not been treated. It does so by using data from the pre-treatment period (i.e. the pre-event or pre-intervention period) to minimize the difference between (i) observable values of predictor attributes and the outcome variable of the focal unit and (ii) values of the same

[1] The authors will make publicly available software code and data to enable replication of all analyses in the paper.

variables in the synthetic unit. In essence, the underlying algorithm calculates positive or null weights, which sum to one, on all potential control units in the pre-treatment period to create a synthetic unit that replicates, as best it can, the outcome variable in the treated unit during the pre-treatment period.

The weights on the control units, determined using pre-treatment data, can be applied to generate post-treatment outcomes for the synthetic unit. Post-treatment outcomes may then be interpreted as if they were the counterfactual outcome values, assuming an acceptable fit can be created such that the synthetic and the focal unit track one another in the pre-treatment period. Divergence in outcome values between the synthetic and focal unit may happen in the post-treatment period if the intervention has a significant effect.

The pool of potential control units consists of similar units in the population that did not receive the treatment and for which panel data can be collected on *k* attributes, including potential predictor variables and the outcome variable in question. The method subjects the comparison units' predictor variables and outcome variable data in the pre-treatment period to an optimization process that minimizes

$$\sum_{m=1}^{k} v_m \left(X_{1m} - X_{0m}W\right)^2$$

by selecting the optimal values of both W and v_m—where X_{1m} is the value of the *m*-th attribute of the focal unit (including the outcome variable); X_{0m} is a 1 x *j* vector containing the values of the *m*-th attribute of each of the *j* potential comparison or control units; W is a vector of weights on control units in the donor pool; and v_m is a vector of weights on attributes of the control units such that they maximize the ability to predict the outcome variable of interest (Abadie et al. 2015).[2] This optimization process minimizes prediction error between the actual and the synthetic in the pre-treatment period. If the optimal W and v_m create a good synthetic match then the synthetic's outcome variable will closely track the actual outcome variable's values during the pre-treatment period. In this case, the synthetic may be deemed as a reasonable replica of the focal unit.

Applying the weights, *W*, estimated in the pre-treatment period to the control units' post-treatment outcome values, Y_o, creates the synthetic's post-treatment outcome data, i.e. Y_oW. This represents the untreated counterfactual of the focal unit in the post-treatment period. Y_1 denotes the observed outcome data for the focal treated unit in the post-treatment period. Whether the gap between the synthetic counterfactual and actual

[2] More precisely, $v_m = argmin_{v\in V}\left((Z_1 - Z_0W^*(V))'(Z_1 - Z_0W^*(V))\right)$ where Z_0 is a *m* x *j* vector containing values of predictor attributes for the *J* potential control firms, where W^* is the value of W that minimizes the equation above.

outcome variable in the post-treatment period (i.e. the value of $Y_1 - Y_o W$) remains the same size, becomes increasingly positive, or increasingly negative allows researchers to make inferences about the direction and the magnitude of the treatment effect. In other words, if the outcome variable of the synthetic control diverges significantly from the actual outcome behavior in the post-treatment period, the performance gap may be attributed to the effect of the treatment. The strength of the inference can be quantified by constructing a randomization-based p-value from the results of placebo tests and further validated through robustness checks aimed at falsifying underlying assumptions.

Abadie et al. (2010, 2015) provide additional technical details and proofs supporting the underlying synthetic control methodology. Abadie et al. (2015) also formally prove that the underlying mathematics of the synthetic control methodology collapse down to those in a traditional regression with an additional restriction that the linear combination of weights in synthetic control must sum to one, whereas equivalent weights derived from a regression model need not be restricted to doing so. Software is available in Stata, R and Matlab to implement synthetic control methods.

Inference in Synthetic Control

Due to the small sample sizes used in synthetic control, assessing the validity of inference requires different approaches than calculating frequentist p-values used in regression settings. Assuming that a good match is achieved between the actual unit and its synthetic in the pre-treatment period, placebo tests are one approach for testing whether estimated results in the post-treatment period are spurious (Bertrand et al. 2004). The central premise of placebo tests is that replicating the methodological analysis using a hypothetical intervention on untreated units, or at different points in time, should not generate a meaningful divergence.

'Across-unit' placebo tests examine whether synthetic control analysis produces large estimated effects when hypothetical treatments are applied to units that were not subject to the intervention on the actual treatment date. Causal inference can be supported when they do not produce gaps between observed and synthetic outcomes that are as large as those for the focal unit at the actual treatment date. Creating a synthetic for each untreated control unit in the population enables researchers to ascertain whether the estimated effect for the focal unit is of unique magnitude and direction relative to all other untreated units. These synthetics constructed for untreated control units also provide the basis for calculating randomization-based p-values. The ratio of the post-treatment prediction errors to the pre-treatment prediction errors can be used to calculate a scale-free measure of the extremity of the impact of the hypothetical treatment on each untreated control unit. The empirical distribution of these extremity measures allows researchers to compute p-values based on permutation inference in the population (Rosenbaum 2002a, 2002b) rather than frequentist inference, which relies on assumptions about the functional form of underlying distributions. These randomization-based p-values can be interpreted, nevertheless, similarly to frequentist p-values in regression settings to assess the strength of the statistical inference.

'In time' placebo tests are analogous to across-unit placebo tests but are conducted by applying hypothetical treatments to the focal unit at different points in time from the actual treatment date. If the effect of the treatment on the focal unit is causal, then replicating the synthetic control analysis when the treatment is falsely applied at other dates should not generate a divergence between the new synthetic and the actual outcomes.

A second approach to testing the robustness is to selectively modify the underlying structural relationship embedded in weightings of control firms in the baseline synthetic. 'Leave-one-out' tests replicate the synthetic analysis but exclude, on a one-by-one basis, each control unit that was originally included in the synthetic from the pool of those eligible to comprise the new synthetics. The test is useful in situations where there may be concerns about indirect effects of the focal unit's treatment on other units, which may arise, for instance, due to competitive interactions among dominant firms within an industry. If the performance of each of the 'leave-one-out' synthetics mirrors the baseline synthetic then it is less likely that the results are biased by the inclusion of any single control unit. 'Out-of-sample' tests examine the sensitivity of the core synthetic counterfactual to alternative specifications, in this case using a different time frame in the pre-treatment period to estimate structural weights on control units that comprise the synthetic. Out-of-sample synthetics that are similar in composition and performance to the baseline can help alleviate any concerns that the structural relationship between the focal unit and synthetic's control units significantly shifted or changed during the pre-treatment period.

2.2 Comparisons with Other Methods

Synthetic control has similarities to other empirical methods that estimate treatment effects, such as matched sampling (Rosenbaum and Rubin 1983), difference-in-differences (D-i-D) models (Card and Krueger 1994), and event studies (McWilliams and Siegel 1997), but also some distinctive differences (see Table 1), which make each of these methods more applicable in certain contexts but not in others.

<Insert Table 1 here>

Matched sampling and D-i-D methods are similar to synthetic control in that they construct counterfactuals based on untreated comparison groups. Matched sampling uses cross-sectional (or pooled) data to selectively 'match' each treated unit with an untreated unit that is comparable on measurable attributes that are predicted to affect the outcome variable. Propensity score matching is one of a number of alternative matching techniques that select untreated units with a similar ex ante probability of receiving treatment based on their characteristics (Dehejia and Wahba 2002); treated and untreated pairs of units are matched on their estimated probability of receiving treatment. The average treatment effect is estimated by comparing the means of the outcome variable across the treated and matched comparison groups, and testing for statistical difference. A statistically significant coefficient for the treatment variable may be interpreted as evidence of causality under the (potentially strong) assumption that there is no selection bias due to unobserved characteristics.

D-i-D models use pre- and post-treatment period panel data to estimate the average impact of a treatment, controlling for unobserved, time-invariant characteristics, as well as time-varying measurable factors that may influence the outcome variable. After also controlling for time period effects, any statistically significant divergence in outcomes between treated and untreated units may be interpreted as evidence of a causal effect of the treatment. In the standard D-i-D model, the identifying assumptions for inference are that there are no omitted time-varying differences or trends between treated and treated groups that have a causal effect on outcomes, and that treatment itself is exogenous.

One distinction between synthetic control and D-i-D models is the ability to deal with endogeneity biases arising from omitted variables: while D-i-D approaches correct for unobserved time-invariant confounders, synthetic control can account for *time-varying* unobserved confounders, assuming a good pre-treatment fit is achieved (which we discuss below). Abadie et al (2010) demonstrate that if a synthetic unit's outcome data tracks the focal unit's well over a sufficiently long pre-treatment period, omission of unobserved variables need not lead to biased estimates: only control units that are similar to the focal unit in terms of observed and unobserved attributes should produce similar outcomes and trajectories over extended durations. Hence, once a synthetic is established that closely mimics the actual in pre-treatment outcome behavior, any divergence in outcomes that arises in the post-treatment period may be attributed to the treatment itself (assuming there are no post-treatment idiosyncratic shocks either to the focal unit or to included control units).

A second distinction for synthetic control is the size of the data sample required for implementation. Matching and D-i-D models require sufficiently large samples of treated and untreated units to construct the comparison groups, and to conduct statistical tests. Recent empirical studies that use matching methods have used samples with more than several hundred observational units (e.g. De Figueiredo et al. 2013; Dahlander and O'Mahony 2011), while D-i-D studies typically require even more (e.g. Fabrizio 2012; Rawley and Simcoe 2013; Toh and Kim, 2013). By contrast, synthetic control can operate with fewer than 20 units in the sample. Based on our review of published papers that use synthetic control, the majority use data sets with fewer than 50 units.

Synthetic control can also be applied in situations where treatments are rare events, or affect just one unit, which often precludes causal inference based on matched sampling or D-i-D methods. It thus permits researchers to estimate the specific impact of a single treatment event on a single unit rather than the *average* effect across all units of a larger number of related events within the same class. The ability to quantify the effect of a specific event is valuable when there is substantial heterogeneity within the population of an event type that is not readily observed or measured by the researcher. For instance, while most product or automobile recalls are relatively minor in magnitude and receive minimal media attention, a small number of recalls within an industry are 'blockbuster' events that are expected to have deeper consequences for organizational performance and operations. Synthetic control thus enables researchers to focus attention on singleton events that have unique characteristics and to quantitatively estimate their effects.

Normally, researchers would be restricted to qualitative case study methods when just one unit in a population is subject to treatment, though identifying a comparable untreated unit as the counterfactual may not always be feasible or straightforward. An advantage of synthetic control is that it can help guide researchers in choosing counterfactual(s) for comparative case study analysis through its optimization methodology – which objectively constructs the synthetic comparator. In this sense, synthetic control is a complementary method for case study analysis by providing a more objective approach to the selection of comparable case study organizations (Eisenhardt 1989; March et al. 1991; Runde and de Rond 2010).

Another potential method for assessing the effect of a single event on a single organization are event studies of stock price returns, but this restricts enquiry to publicly quoted organizations and to the stock price as the sole possible dependent variable. Organization scholars are often interested in broader measures of performance than abnormal stock returns, and synthetic control (like regression analysis) can be applied to any outcome dependent variable for which data is available. Event studies also focus on changes in stock prices over a short time window in assessing treatment effects, typically several days or less, whereas synthetic control has the flexibility to adopt time horizons of months or years, depending on the nature of the outcome variable.

2.3 Assumptions and Limitations of Synthetic Control

Despite some beneficial attributes, caution is needed in implementing synthetic control, and also in interpreting estimated results, due to assumptions and limitations of the method.

The utility of synthetic control depends substantially on achieving a good match in the pre-treatment period between the actual and synthetic's outcome variable, which indicates that a plausible counterfactual can be constructed for the purpose of causal inference. However, it is not always possible to achieve a close fit, as when the mean values of the synthetic and actual outcomes in the pre-treatment period are far apart, and the time series of each frequently move in opposite directions. Several conditions or assumptions should be satisfied for a good match to be created. First, there should be a donor pool of untreated control units that are sufficiently similar in their attributes to the focal treated unit from which to construct the synthetic. If there is significant heterogeneity in the attributes of the control pool units, or if there are few untreated units comparable to the focal, a poor synthetic match is more likely to occur. Accordingly, it is not technically feasible to create a good match for units that have extreme values of the outcome variable in the population since it is not possible to create a weighted linear combination of other units (where the weights are non-negative and sum to one) that produce either the maximum or minimum value of the outcome variable.

A second assumption is that a relatively stable structural weighting relationship exists between the focal unit and the synthetic's included control units during the pre-treatment period. Major shocks or events that occur in the pre-treatment period affecting control units would make it more difficult to construct a synthetic that performs similarly to the focal unit throughout the duration of the pre-treatment time frame (though such units may be dropped from the donor pool when constructing the synthetic). Out-of-sample tests can be used to test the stability of structural relationships in the period before the treatment event occurs.

Third, interpretation of the synthetic's outcome performance is affected by the assumption that there are no significant shocks in the post-treatment period (other than the treatment event) either to the focal unit or to the synthetic's included control units. (This is akin to the "common trends" assumption in D-i-D models.) If control units' outcomes were affected by idiosyncratic events that occurred after the treatment, it may not be possible to attribute any divergence in outcomes between the synthetic and actual unit as being primarily due to the impact of the focal treatment event. Again, however, control units subject to post-treatment shocks may be omitted from the pool of potential donor units used to construct the synthetic.

The fourth assumption, which also shapes interpretation of post-treatment synthetic outcomes, is that treatment of the focal unit does not have meaningful indirect spillover effects on the outcomes of control units included in the synthetic. Spillovers may introduce upward or downward bias into estimates of post-treatment performance, either magnifying or diminishing the estimated size of the treatment effect on the focal unit. Recognizing the direction of any bias introduced by spillovers will thus lead to modified assessment of core synthetic estimates. Leave-one-out robustness tests enable researchers to test the sensitivity of results to the inclusion of specific control units in the synthetic if there is concern about such indirect effects. Creating placebo synthetics for all donor units is also a method to test whether any single unit experienced indirect spillover effects.

Beyond these assumptions, a limitation of the synthetic control method is that measuring the goodness of fit between the synthetic and actual unit during the pre-treatment time period - to judge whether a good match has been achieved - relies partially on a visual assessment of graphical results depicting how closely the synthetic tracks the actual outcome variable. Although the method produces a root mean squared prediction error (RMSPE) measure of model fit, its scale is specific to the outcome variable and hence cannot be used to compare across models with different outcome variables. While there may be common agreement among researchers about visually clear good or bad matches, there may be less agreement about mediocre matches and whether they warrant using synthetic control. Discretion is thus required in selecting synthetic control as an appropriate technique for quantifying treatment effects.

3 APPLICATION 1: GOVERNMENT INTERVENTION IN CHRYSLER

We turn now to our first application of synthetic control in an organizational context, an analysis of the effect of government ownership of and control over Chrysler following the 2008 financial crisis on Chrysler's sales performance. The estimates we generate contribute to a broader academic debate in management literature on the role of government in private industry (Mahoney et al. 2009; Kivleniece and Quelin 2012), in which predictions about the impact of government intervention on firm performance range from positive to negative.

The small number of major domestic and foreign auto firms selling in the U.S. (19), and the even smaller number of firms that received government intervention (2), challenges the data requirements of difference-in-differences and matched sampling methods.[3] Event studies using stock market data are precluded since the focal firm, Chrysler, was privately-owned. Synthetic control is nevertheless potentially suited to this context given the limited population of firms in the auto industry, and the availability of balanced panel data on firm attributes and performance measures over multiple years.

3.1 Background

During 2008 the U.S. automobile sector experienced a significant demand shock as the domestic economy entered recession, consumer credit markets contracted following Lehman's bankruptcy, and as oil prices peaked at $145 per barrel. Consequently, industry-wide sales of new vehicles declined by 18% in 2008 as compared to the previous year, the worst annual fall since the early 1970s. Although sales rebounded in late 2009, the demand drop created significant short-term financial challenges for all auto manufacturers, including Toyota which experienced its first financial loss since 1950, and led the three major Detroit-based auto manufacturers to request federal government assistance. In December 2008, President Bush authorized the Treasury Department to extend an initial $13.4 billion of TARP funds to Chrysler and General Motors, explicitly spelling out some conditions for assistance while leaving others to the discretion of the Treasury. Under President Obama, a dedicated Presidential Task Force on the Auto Industry oversaw the firms' usage of TARP funds beginning in early 2009: both companies were required to obtain government approval for major restructuring plans and to undergo government-facilitated bankruptcy processes, which led to the government assuming equity ownership positions. Chrysler emerged from those proceedings in June 2009, partly owned by the U.S. government with a 9.8% interest. The U.S. government maintained its oversight over and shareholder position in Chrysler for a period of 29 months from January 2009 until May 2011, when Fiat bought the government's equity stake and paid down its final debt, six years earlier than originally planned.

While the Treasury stated that it would seek to "maximize overall investment returns" on the government's ownership stake in Chrysler, its guiding principles also stated that it would "promote stability for and prevent disruption of…the economy" and "protect taxpayer investments", goals that need not have aligned with private shareholder interests (Black 2010). Accordingly, the government conditioned its financing on various

[3] It is not possible to conduct a synthetic control analysis of the impact of government intervention on General Motors (GM) since GM was the largest automobile firm (by sales volume) in the market, making it technically infeasible to construct a weighted linear combination of other firms with a similar level of sales as GM where the weights are positive and sum to one.

operating restrictions and commitments from the companies including, inter alia, tight limits on executive compensation and a requirement to produce a share of vehicles in the U.S. rather than abroad. The Treasury appointed four of nine directors on the Chrysler board, and retained the power to block any large transactions (Office of the Press Secretary 2008).

Although the government's financial support in late 2008 likely prevented a traditional private sector bankruptcy or even a liquidation of Chrysler from occurring (Goolsbee and Krueger 2015), the ongoing effect of active government intervention on the company's subsequent operating performance from 2009 onwards has been debated. Industry experts have highlighted a number of disputed issues where government overrode management recommendations, including the Presidential taskforce's move to rapidly terminate 25% of Chrysler's dealership network,[4] and the decision to make significant wage and pension concessions to protect union employees – both of which could have had detrimental consequences for Chrysler's ongoing performance.[5] On the other hand, access to lower cost financing and improved confidence among new vehicle purchasers about future continuation of warranties may have bolstered profits and sales. The net effect of government intervention on Chrysler's operating performance from January 2009 to May 2011 remains unclear and is an open empirical question.

3.2 Data

The synthetic control method can provide insight into the impact of government intervention by constructing a counterfactual Chrysler based on auto firms that did not receive the 'treatment' of government involvement in organizational decision-making after January 2009. The counterfactual thus assumes that Chrysler was financially in a position to continue operating after receiving TARP funds in late 2008, but did not experience

[4] Chrysler management proposed that dealership rationalization should be a gradual process over five years, ending in 2014. Instead, the Treasury, seeking to achieve visible cost reductions, required the Chrysler to drastically accelerate this process, leading to the termination of 789 dealerships, or 25% of the total, by June 10th, 2010 (SIGTARP 2010). The independent Special Inspector General for TARP was publicly critical of this decision since a large fraction of the terminated dealerships were in rural areas, which still represented a relative market strong-hold for the major Detroit automakers.

[5] Chrysler management initially wanted its 225,000 underfunded pension obligations to be discharged in bankruptcy proceedings as this would maximize flexibility and profitability going forward; however, the government wanted to avoid this because the federal Pension Benefit Guaranty Corporation (PBGC) would have been responsible for covering $2 billion of the shortfall (Walsh, 2009; Whoriskey, 2009). At the time Chrysler emerged from managed bankruptcy proceedings, the United Auto Workers' (UAW) Retiree Fund was granted a 67.7% equity stake in the reorganized firm as compensation for the unfunded pension obligations—thereby protecting the government, and in turn union employees, at the cost of the firm (Anginer and Warburton 2014; Roe and Skeel 2010). The government also blocked reductions in hourly wages for existing union employees, endowing Chrysler with a higher cost structure than would have been likely had the firm pursued a pure private sector refinancing-and-restructuring process, constraining the firm's flexibility in offering lower prices on vehicles.

active government involvement in its management from 2009 to mid 2011.[6]

The dependent variable for our analysis is U.S. monthly light vehicle sales volume by firm (i.e. the total number of new cars, light trucks, and vans sold).[7] Sales are the key top-line performance measure for the automotive industry, commonly quoted in the media and closely scrutinized by policymakers and financial analysts alike as a bellwether of firm health. The monthly frequency of its release makes it a good candidate for study as it provides a test of the immediate impact, if any, of government intervention. Alternative measures of organizational performance such as quarterly return on assets or return on sales could also be utilized in principle, but disaggregated corporate financial data solely for the U.S. geographic market is generally not available.

We obtained monthly data on vehicle sales and other variables for the period January 2005 to December 2012 for each of the 19 major auto companies selling in the U.S. from *Ward's Auto*, an auto industry data provider. The time frame encompasses a 48 month pre-treatment period (from January 2005 to December 2008), a 29 month treatment period (January 2009 to May 2011), and a 19 month post-treatment period (June 2011 to December 2012). We experimented with different pre-treatment windows such as 36 months, though a better synthetic fit is achieved with a 48 month window. The core results are nevertheless robust to alternative pre-treatment time periods. We exclude GM from the pool of eligible control firms since it too experienced government involvement in managerial decision-making, and also Jaguar Land Rover due to missing data on several variables. Hence, the pool of eligible control units for inclusion in the synthetic consisted of 16 firms.

We build from the *Ward's* data, several series of explanatory variables that predict light vehicle sales for any auto company. These include attributes of the vehicles sold, along with attributes of the firms such as measures of production capabilities, scope, financial condition, and past performance. The synthetic control method does not place conditions on the number of predictors required; given its optimization process it will assign low weights to predictor variables in construction of a synthetic if they have little explanatory power.

The first set of independent variables captures, on a monthly basis, vehicle-specific factors that could

---

[6] Insider and official reports document the extensive degree of government intervention in the two auto companies' strategic and operational decision-making (Rattner 2011; SIGTARP 2013). Had the government not established a dedicated executive taskforce with specific objectives, resources, and authority to intervene in Chrysler and GM, it is plausible that government intervention in the companies would have been light-handed, as was the case for other TARP-supported sectors.

[7] In robustness checks, we explored alternative outcome variables disaggregated to capture: (i) total monthly sales of light trucks and vans but not cars (ii) total monthly sales of cars, but not light trucks or vans; and, (iii) total annual fleet sales of light vehicles purchased by governments and rental car firms. The results are consistent with our primary findings.

drive consumer demand, including: average price, average fuel economy, maximum fuel economy, average size of engine, and the average weight of the vehicles sold. The second set of variables captures firm-level strategic and operational factors, including the number of active production platforms; the number of active brands; the number of active series within those brands; the number of market segments in which they compete (i.e. luxury, small car, crossover, etc.); the fraction of sales that are from SUVs, light trucks, and vans; and, the fraction of sales of imported vehicles. Further firm features we incorporate into our set of predictor variables are available on an annual basis only, the total number of employees and leverage ratio (total debt/equity). Finally, we include as an additional predictor of performance, the prior month's sales volume indexed to 100 at the time of the government loans. Appendix A provides data sources and details of the data's construction.

3.3 Results

Table 2 shows the composition of the synthetic Chrysler as constructed by the optimization process that assigns weights to control firms in the donor pool. Among the sixteen control firms that could comprise the synthetic, five receive positive weightings and the remaining eleven receive zero weightings. It is typical in applications of synthetic control that a substantial number of potential control units receive zero weights because they do not make good individual matches on the outcome variable as their other attributes are not sufficiently similar to the focal unit's to make them good comparative cases.

<Insert Table 2 here>

Of the five firms that receive positive weightings—Daimler, Ford, Isuzu, Nissan, and Toyota—intuition helps reconcile the weights. Many auto industry observers would likely select Ford as the closest match to Chrysler had General Motors been ruled out as an option given similarities in the vehicle fleets. Moreover, like Chrysler, Ford had initially requested financial assistance from the government in 2008 but later retracted due to concerns regarding the structure and oversight of the loan conditions. These facts make it unsurprising that Ford receives the largest weight in the synthetic at 0.664. That the weighting is not 1.0 illustrates that a combination of auto firms (including Ford) creates a better synthetic representation of Chrysler than just Ford alone.

Toyota and Nissan have diverse product offerings, multiple brands and production platforms, extensive networks of production facilities in the United States, and heavier weighting towards SUVs, light trucks, and vans than most other large Asia- and Europe- based manufacturers. Toyota and Nissan thus receive relatively high weights of 0.077 and 0.169, respectively, in the synthetic Chrysler.

Isuzu may at first seem like an odd company to be included in the synthetic Chrysler, but understanding why illustrates the objective way in which the synthetic control method selects control firms. Isuzu's inclusion can be reconciled when recognizing that the smaller Asia-based manufacturer was heavily weighted towards SUVs, light trucks, and vans—which were an important part of Chrysler's product portfolio and which were highly sensitive to demand fluctuations. Nevertheless, little weight is ascribed at 0.068 given that the other firms in the synthetic also manufacture heavier vehicles. Daimler enters the mix with a small 0.022 weighting, rounding out the large passenger car end of the market in the synthetic Chrysler.

Table 3 demonstrates how the synthetic Chrysler compares on the predictor attributes to the real Chrysler in the 48-month pre-intervention period. The synthetic's attributes are generally close in mean value to the actual Chrysler's, suggesting that the synthetic is a good match. For those attributes where there is a greater difference between the synthetic and actual mean values, the predictive power of those attributes tends to be lower, as calculated in the attribute weighting matrix, v_m.[8] As a basis of comparison, Table 3 also includes the attributes for Ford in the same 48-month window: in the majority of cases the mean values of the synthetic Chrysler's attributes are closer to Chrysler's actual values than are Ford's, implying that the composite synthetic is a better counterfactual match than a single control firm such as Ford.

<Insert Table 3 here>

The core results of the synthetic control analysis are illustrated graphically in Figure 1, which shows the synthetic Chrysler's sales performance during the pre- and post-intervention periods (January 2005 to December 2012) along with the actual firm's performance. The graphical representation, which is standard in synthetic control studies, enables a visual assessment of the comparison with Chrysler's actual performance.

<Insert Figure 1 here>

Prior to the start of government intervention in January 2009, the performance of the synthetic Chrysler and of the actual company track each other reliably, with only short periods when either the synthetic or the actual company outperform each other by marginal amounts. Moreover, the two series have approximately the same average value in the pre-period. There is a downward trend in both Chrysler's and the synthetic's performance which was indicative of the entire auto industry in years prior to the intervention. Given that the series track each

[8] The five attributes that obtain the greatest weights are average weight, sales volume, average price, segment count, and series count. Our results remain robust to alternative model specifications and to including and excluding specific variables.

other well despite the industry turmoil, the synthetic achieves a good match for Chrysler's actual performance. This supports the synthetic control requirement (see section 2.3) that there is a pool of untreated control units that are sufficiently similar to the focal unit in the pre-treatment period to achieve a high quality synthetic match.

What is notable in Figure 1 is that during the government intervention period Chrysler significantly underperforms its synthetic counterfactual, representing estimated performance of the firm without government involvement. This suggests that Chrysler may have been able to sell more vehicles in the absence of government intervention and oversight. Had the two series instead continued to track each other during the intervention period then we would not have been able to make this inference. Similarly, had the synthetic underperformed Chrysler in practice then we might conclude that government intervention increased subsequent sales.

Figure 2 shows the magnitude of the gap between Chrysler's actual sales and its synthetic's sales. At its widest, Chrysler sold 59,000 fewer vehicles in that month than the counterfactual. Given the level of Chrysler's actual sales during the intervention period, Chrysler appears to have sold approximately 29% fewer vehicles than its synthetic counterfactual over the entire period from January 2009 through its repurchase of government equity stakes in the firm, although the estimated shortfall is quite volatile on a monthly basis. Chrysler's underperformance relative to its synthetic is most marked in the two and half year period from January 2009 through May 2011, after which it gradually diminishes until late 2012. The timing of the reversal is notable since it corresponds with the agreement to end government involvement and control.

<Insert Figure 2 here>

Robustness Tests

To test the robustness of the main synthetic estimate, we conduct several placebo tests and model specification sensitivity checks.

Placebo Tests among Control Units

If the effect of the treatment is material in the case of Chrysler, hypothetical application of the same treatment to control firms should not lead to an equally large performance divergence (or treatment effect). We thus replicate the core analysis, creating synthetics for each of the sixteen control firms, and compare the performance of each to its actual. In Figure 3 the dotted lines represent the gap in performance between the control firms' outcome values and their synthetics', while the solid line represents the gap between Chrysler and its synthetic (as depicted in Figure 2).[9] The performance gap is smaller for all the control firms than it is for Chrysler during the period of government intervention, which is consistent with a causal effect of the treatment on Chrysler's performance.

<Insert Figure 3 here>

The estimates generated from the placebo tests among control units also provide a basis for evaluating the strength of inference quantitatively by constructing p-values based on randomization or exact inference. In a regression setting, the strength of inference is assessed with frequentist p-values calculated under the assumption of a normal distribution of parameters. In synthetic control, a similar statistic can be generated computationally, based on randomization inference principles (Ernst 2004): randomization-based p-values capture the probability of obtaining a result at least as extreme as the one estimated for the focal unit in the event that the treatment was randomly assigned to any observation unit in the population.

To evaluate the relative extremity of the treatment event on each observation unit, we adopt the method of Abadie et al. (2010) in calculating a scale-independent measure of treatment extremity so we can compare observation units directly with each other.[10] We use the ratio of the RMSPE for the treatment period to the RMSPE for the pre-treatment period, calculated for each firm evaluated in the among-units placebo test above and signed for the direction of post-treatment performance. Values of this ratio close to 1 or -1 indicate there are no major changes in how well the synthetic matches the actual unit in the treatment period relative to the pre-treatment period.

Figure 4 presents a dot plot of these treatment extremity measures for the 17 firms in the population (16

---

[9] In the placebo tests among control units we excluded synthetics for control firms with mean-squared prediction errors (MSPE) in the pre-intervention period greater than 5 times that of Chrysler. This is due to these control firms having a poor synthetic fit in the pre-treatment period, which is expected particularly for observation units at the extremes (Abadie et al. 2010). This filter removes noisy observations to clarify the result that placebo interventions among untreated units do not have as large a negative effect on sales. Applying such a filter is standard in studies using synthetic control—and in fact, the filter for exclusions that we apply here at 5 times MSPE is conservative; Abadie et al. (2010), in their study of state-level anti-tobacco laws, exclude control units with MSPE that are only 2 times greater than California which was the focal unit.

[10] We make a minor modification to the Abadie et al. (2010) procedure in that we distinguish between positive and negative estimated treatment effects, multiplying a unit's post/pre RMSPE ratio by -1 when the synthetic outperforms the actual, and by +1 when the opposite is true. This allows us to compare the direction as well as the absolute magnitude of the treatment effect for each unit. Distinguishing between positive and negative effects also enables us to conduct a one-tailed t-test to estimate the probability of observing as large a treatment effect of the same sign as for the focal unit.

control firms plus Chrysler). The majority of the observations are clustered towards the center of the distribution, falling in the range [-2, 2], suggesting that government intervention in the management of Chrysler, or coincident events, had negligible effects on most firms.

The x-axis of the dot plot is bordered by a histogram representing the empirical distribution of the extremity measure. Chrysler occupies the second most negative position with a treatment extremity value of -2.5. Recognizing that there are 17 firms in the population and that Chrysler is second from the bottom of the distribution, the p-value is calculated by taking the number of observations as extreme as Chrysler and dividing by the number of observations in the population—which is analogous to integrating the assumed normal distribution in a regression setting between negative infinity and the observed value of the coefficient to generate a p-value (as in a one-tailed t-test). Hence, if instead government intervention was randomly assigned to any firm in the population, the probability of seeing as large a negative effect as on Chrysler's sales is approximately 12% (≈ 2/17). The impact of government intervention on Chrysler's sales may thus be interpreted as being measurably less than zero with a p-value of 0.12.

<Insert Figure 4 here>

Some care is needed, however, in interpreting p-values in small *n* settings. In the hypothetical case where Chrysler had the most negative extreme response to government intervention, its p-value would be $1/17 \approx 0.06$. If researchers followed a strict $p>0.05$ rule they may reject the null hypothesis and declare the treatment effect to be negligible despite it being the single most extreme post-treatment realization.

Placebo Tests in Time

We also implement an in-time placebo test, replicating the analysis but instead using January 2005 as a hypothetical intervention date instead of January 2009, and a 48-month pre-treatment window. Figure 5 demonstrates that the placebo intervention in January 2005 does not have a dramatic effect on Chrysler's future sales since the synthetic version of the firm continues to track the performance of the actual firm quite closely in the post placebo intervention period. We obtain a similar result (with no divergence) when we have a placebo intervention in January 2004. We also find no immediate effect of placebo interventions in January 2006, January 2007, or January 2008, though the tests are not ideal given the 48-month pre-treatment window overlaps with the baseline analysis window. The absence of divergence in these placebo tests supports causal interpretation of the effect of government intervention after January 2009.

<Insert Figure 5 here>

Leave-one-out Tests

The potential for results to be biased by the inclusion of a particular control unit may be a concern if one of the synthetic's components experienced either indirect spillover effects from the focal unit's treatment, or a large, unrelated shock in the post-treatment period. Here, consumer switching from Chrysler to Ford, for instance, would inflate the estimate of the counterfactual's sales during the treatment period. Synthetic control analysis assumes that spillover effects on control units are negligible or offsetting, and that any shocks to outcomes in the post-treatment period are relatively minor or are common to all units, including the treated. Leave-one-out tests help examine whether or not such assumptions are valid.

To perform the leave-one-out test, we construct five alternative versions of the synthetic Chrysler that exclude each of the firms that comprise the baseline synthetic Chrysler on a one-by-one basis. Hence, each of the five alternative synthetic Chryslers is constructed from the population of auto manufacturers excluding either Daimler, Ford, Isuzu, Nissan, or Toyota. Figure 6 presents a graph displaying the differences between the new leave-one-out synthetics and the actual Chrysler (five dashed lines), and also the difference between the baseline synthetic and the actual Chrysler (solid blue line).

<Insert Figure 6 here>

The leave-one-out synthetics and the baseline synthetic track each other closely overall. (It can be difficult to distinguish all five dashed lines in Figure 6 because the differences between any single leave-one-out synthetic and the baseline synthetic are relatively small). The average sales gap between these alternative synthetic Chryslers and the actual firm over the intervention period running from January 2009 through May 2011 ranges from 28% for the synthetic that excludes Ford to 30% for the synthetic that excludes Toyota. The synthetics leaving out Daimler, Isuzu, and Nissan, respectively, fall within this range, but closer to the 29% baseline reduction in sales. Hence, any bias introduced by the inclusion of a particular control firm in construction of the synthetic Chrysler appears to be relatively small. This provides some support for the assumptions that treatment spillover effects on included control units, and post-treatment shocks, are not major confounding concerns in the interpretation of synthetic estimates.

While leave-one-out analyses provide tests of the sensitivity of the results to the composition of the control group – which is especially valuable if one control unit accounts for a significant fraction of the synthetic

– some caution is warranted since the focus is on omitting just one unit at a time. In this case, any significant shifts in consumer vehicle preferences during the treatment period, for instance towards competitors with more fuel efficient vehicles or towards Japanese marques, may have changed the relationship between Chrysler and the synthetic firms, potentially diminishing the synthetic's role as an accurate counterfactual.

Out-of-Sample Tests

Out-of-sample tests help assess whether the weights on control units in the synthetic are reasonably stable in the pre-treatment period, and hence whether the assumption of a stable structural relationship between the focal unit and the synthetic's control units is valid. We implement an out-of-sample test using data in the pre-treatment period from 2001-2004 – rather than data from 2005-2008 as in the baseline – to construct a new synthetic Chrysler (see Figure 7) in the focal period. We observe a sales divergence pattern similar to those in Figures 1 and 2, suggesting that the main results are relatively robust to structural weightings estimated in different pre-treatment time frames. This alleviates concerns about structural breaks altering sales patterns within the industry before 2009, and supports the assumption of stability in the pre-treatment period structural relationship.

<Insert Figure 7 here>

In sum, the synthetic control analysis presented here provides a novel quantitative estimate of the impact of ongoing government intervention on Chryslers' sales performance during its period of ownership and control, based on a composite counterfactual of a small number of firms that did not experience government intervention. The shortfall estimate suggests that, despite likely ensuring financial viability of Chrysler through the TARP financial package in late 2008, had the firm been able to remain a going-concern without government intervention for 29 months, Chrysler could have sold substantially more vehicles. In-depth case study analysis, including comparisons with firms such as Ford and Nissan that are included in the synthetic, may identify specific causal process mechanisms that explain this finding, and also allow exploration of the effect of government involvement on other dimensions of organizational performance.

4 APPLICATION 2: TOYOTA'S UNINTENDED ACCELERATION CRISIS

To further assess the utility of synthetic control for organizational research, we apply it to a second research question in the auto sector – the impact of Toyota's unintended acceleration crisis and vehicle recall during 2009-2010 on sales of its top sub-brand (Camry) – that illustrates how the method can provide novel insights even in 'large *n*' data contexts by estimating individual unit-specific rather than average treatment effects, and thereby complement regression-based statistical techniques.

Unlike U.S. government intervention in automobile firms, vehicle design flaws and brand recalls have been relatively frequent events over the last few decades (Davidson and Worrell 1992; Chen et al. 2009), facilitating traditional statistical analyses of their organizational impact. Event studies of stock price reactions to recall announcements have been a common methodological approach, collectively yielding mixed results, ranging from negative (Jarrell and Peltzman 1985), to no systematic aggregate effect (Bromiley and Marcus 1989), to short-term negative reactions that are subsequently reversed (Govindaraj et al. 2004).[11] Other studies have leveraged panel data on sales and market shares to estimate fixed effect regression and D-i-D models of the effect of product recalls (Rhee and Haunschild 2006). For instance, Lui and Shankar (2015) use data on 672 U.S. vehicle recall events, categorized by severity and media coverage, affecting 95 million cars from 1997 to 2002, to analyze the impact on brand equity and sales volumes. In a difference-in-differences approach, Freedman et al. (2012) estimate the effect of 212 recalls in the toy industry during 2007 on monthly category sales.

A common feature of these studies is that they pool large numbers of recall events in order to statistically estimate *average* treatment effects. Even after controlling for measurable recall characteristics such as severity, coefficient estimates reflect the average impact both across multiple events and over multiple time periods. While researchers often seek to establish the average effect of a phenomenon, in other cases interest may focus on the impact of a specific event that differs in various ways from others, making it by definition a rare(r) event that does not readily lend itself to statistical analysis. Researchers occasionally couple statistical analysis with in-depth case studies of such unique events to develop further insights into causal relationships (for instance, Freedman et al. (2012) include case studies of three exceptionally high profile recalls, noting that "data limitations preclude us from using regression techniques" (p. 511)). By constructing a purpose-built counterfactual around a single treatment event for a single organization, synthetic control has the potential, if key assumptions are satisfied, to generate new insights about specific rather than average treatment effects of product recalls.

4.1 Background

In August 2009 Toyota came to national prominence as news media reported on a high speed Toyota

[11] Jarrell and Peltzman analyze 116 recalls from 1967-1981, Bromiley and Marcus 119 recalls from 1967 – 1983, and Govindaraj et al. one recall in 2000.

vehicle crash in San Diego that allegedly was caused by unintended vehicle acceleration (Bunkley 2010). The issue gained momentum when it subsequently emerged that other consumers had previously complained about unintended acceleration in Toyota vehicles, which had prompted the National Highway Traffic Safety Administration (NHTSA) to open an investigation two years earlier (Camuffo and Weber 2012). In November 2009, Toyota initiated one of the most extensive recalls in U.S. automotive history, involving 4.2 million vehicles and seven of Toyota's sub-brands, including its top seller, the Camry. As a major manufacturer, Toyota had experienced recalls before (169 in total from 1967 to 2008), though they were all significantly smaller in magnitude, with an average recall size of 80,000 vehicles (NHTSA 2015). Toyota's initial defensive reaction to the acceleration crisis led to considerable public criticism (Liker and Ogden 2011), and it was estimated that the recall led to a 19% drop in the firm's stock price over a ten-day event window (Gokhale et al. 2014).

Despite the magnitude of the recall and the wave of negative media reporting, it was less clear what would be the consequences for ongoing sale volumes of Toyota vehicles overall and of its sub-brands. The 19% fall in the stock price may have reflected investor expectations of greater compliance costs with regulations governing accelerator technology, rather than future reductions in sales, or it may have represented a financial market over-reaction to negative public sentiment about the company's future prospects. Toyota's long-built brand reputation for vehicle quality could have insulated ongoing sales, particularly of its leading sub-brands, from such a negative shock, especially since there was no hard evidence of a technical flaw. Consumer Reports regularly ranked Toyota as first in brand perceptions among vehicle manufacturers, and Kelley Blue Book rated Toyota as first for vehicle resale values. The Camry, which accounted for 20 percent of Toyota's sales (by volume), had been awarded the Motor Trend Car of the Year prize in 2007, and had received a rare 5 star rating from the NHTSA. It was thus plausible that Camry sales volumes could either have fallen for a period following the recall announcement – relative to expected levels in the absence of the recall – or they could have remained relatively unaffected if consumers had discounted the negative media stories or responded to Toyota's enhanced promotional schemes (the latter of which could have even pushed sales volumes upwards).

Even though Toyota, and the broader auto industry, had experienced many vehicle recalls prior to 2009, the relative extremity of the November recall event and the unique market position of the Camry sub-brand make regression-based statistical methods less appropriate since these would estimate an 'average' recall effect across multiple sub-brands. Synthetic control, on the other hand, is a possible candidate for analyzing the impact of this type of rare episode (assuming a good synthetic match can be achieved) since the counterfactual is constructed specifically for the focal unit during the focal time period.

4.2 Data

To estimate the impact of the November 2009 recall on Camry sales we use monthly data, obtained from *Ward's Auto*, at the sub-brand level (e.g. Toyota Camry, Honda Accord, Volkswagen Jetta) for the period from November 2005 to December 2011, giving 48 months of pre-treatment and 25 months of post-treatment observations. The synthetic counterfactual we construct therefore assumes that the Camry sub-brand did not experience the recall event in November 2009.

The donor control pool for constructing the synthetic Camry consists of all vehicle sub-brands sold in the U.S. market by automobile manufacturers other than Toyota (or Lexus, a Toyota sub-brand). We omit vehicles that sold fewer than 2,000 vehicles on average per month (mainly high performance luxury sports cars), and also Chrysler and General Motors sub-brands due to their significant restructuring during the same time period. The resulting donor pool consists of 70 untreated sub-brands. For each sub-brand, we gathered data from *Ward's Auto* on monthly sales as well as predictor attributes, including MSRP, fuel economy, vehicle weight, vehicle length, vehicle width, engine size, engine displacement, and the number of valves. In addition, we included a 12 month moving average of sales indexed to 100 at the time of the recall.

4.3 Results

The synthetic Camry, as constructed by the optimization function (section 2.1), consists of three control sub-brands, each with a different weighting reflecting the relative contribution to the synthetic counterfactual: the Honda Civic (53%), Honda Accord (27%), and Ford F-150 (20%). While it might have been expected that Japanese sub-brands would enter the synthetic, the inclusion of the Ford F-150, a lightweight truck, may be more of a surprise. However, like the Camry, the Ford F-150 is a mass-market vehicle with large monthly sales.

<Insert Table 4 here>

Table 5 presents the comparison between the synthetic and the actual Camry on the values of the predictor variables during the pre-treatment period. The majority of the mean values are closely matched, differing by less than 1%. One exception is the Sales index, though this attribute receives a small weighting (in the v_m attribute matrix) of 0.4%. As a basis of comparison, we have also included in the third column of Table 5 the sales-weighted average for these attributes for the 70 untreated sub-brands in the donor pool. Eight of the ten attributes of this sales-weighted comparator have at least a 10% differential from the Camry's attributes, notably monthly sales volumes and the two engine characteristics, which received large weightings in the attribute matrix—suggesting that the synthetic is a better comparator case than a weighted average.

<Insert Table 5 here>

The core results of the synthetic analysis are presented graphically in Figure 8, which plots the monthly sales of the synthetic and actual Camrys over the pre- and post-treatment periods. During the four year pre-treatment period, despite significant volatility in sales, the synthetic Camry tracks the actual quite well. Average monthly sales are very similar (with the synthetic marginally outperforming), and the two series usually move in the same direction. Both the synthetic and actual Camry experience steep declines of approximately 50% during 2008, followed by rapid rebounds in 2009. The closeness of the match supports the requirement in synthetic control that there is a sufficient pool of untreated units that are similar on key attributes to the focal treated unit.

During the post-recall period, Figures 8 and 9 demonstrate there is little systematic difference between the synthetic and actual Camry monthly sales. In early 2010 there is a significant drop in actual Camry sales – which could potentially be attributed to the impact of the recall – but this is also matched by a fall in the synthetic's sales that is almost as large in magnitude. The two series continue to track each other throughout 2010 and 2011 with some small month-to-month differences, but without any sustained divergence that might be expected had the recall created such a severe shock that Toyota could not mitigate its effects. Hence, after creating a good match between the synthetic counterfactual and the actual Camry in the pre-treatment period, it does not appear that the recall had a significant overall impact on Camry sales volumes. This finding is consistent with qualitative case study research that suggests Toyota's reactive marketing campaign and price discounts might have been effective in offsetting the potentially negative effects of the recall (Liker and Ogden 2011; Camuffo and Weber 2012).

<Insert Figures 8 and 9 here>

Robustness Tests

We test the robustness of the core finding using leave-one-out tests, out-of-sample tests, and placebo tests.

Leave-one-out tests

Given that only three of the 70 potential control units have positive weights in the synthetic Camry, we may be concerned about the impact of any indirect spillover effects from the focal treated unit on untreated units in the synthetic. We test the sensitivity of the estimated performance of the baseline synthetic to its structural composition by using leave-one-out tests, which omit one of the three control units from the donor pool in constructing alternative synthetics. We construct three alternative versions of the synthetic Camry, excluding on a one-by-one basis each of the Honda Civic, Honda Accord, and Ford F-150 from the donor pool. Figure 10 displays the differences between each of the three leave-one-out synthetics and Camry actual sales (three dashed lines), and also the difference between the baseline synthetic's sales and actual Camry sales (solid line).

During the pre-treatment period, the new synthetics track the actual Camry relatively well, though the match is not quite as strong as with the baseline synthetic as there are some periods with greater divergence. On average, however, monthly sales of the leave-one-out synthetics are close to that of the baseline synthetic. In the post-recall period, the leave-one-out synthetics perform similarly to the baseline: there is little sustained divergence from Camry's actual sales, which is consistent with the conclusion from the baseline analysis. As a result, there does not appear to be any systematic bias introduced by the inclusion of a particular control firm in the construction of the synthetic Camry. This offers support for the assumption (see section 2.3) that post-treatment indirect spillover effects or exogenous shocks to control units included in the synthetic, are not significant concerns when assessing the synthetic's performance.

<Insert Figure 10 here>

Out-of-sample tests

One potential concern about applying synthetic control analysis at the sub-brand level is that shifting consumer tastes or developments in vehicle technologies could alter the structural relationship between the Camry and its synthetic's control units during the pre- or post-treatment periods. We thus implement an out-of-sample test to examine the assumption of stability in the relationship during the pre-treatment period. We construct an alternative synthetic Camry based on data from January 2001 to October 2005, rather than from November 2005 to October 2009 as in the baseline analysis.

When we graph (available upon request) the difference in sales between this new synthetic Camry and actual sales from 2005 to 2011, we find, as in the case of the original synthetic, there is a good fit between the new synthetic and the actual in the pre-treatment period, with no clear divergence following the recall in November 2009. The similarity between the two synthetics suggests there was a stable relationship between the Camry and donor control units during the pre-treatment period, which reduces (though does not eliminate) worries about structural changes that would undermine interpretation of the synthetic as a plausible counterfactual. The assumption of a stable structural relationship between the actual Camry and the synthetic's control units in the pre-treatment period thus does not appear unreasonable.

Placebo Tests

As a final check on robustness we implement in-time and among-unit placebo tests, the latter of which we use to construct randomization-based p-values.

For the in-time placebo test, we hypothetically assume that the recall occurred in November 2006, 36 months earlier than the actual date. Accordingly, we construct a new synthetic Camry using pre-treatment data over a 48-month period from November 2002 to October 2006. Although the new synthetic match is not as strong as the baseline in the pre-treatment period, there is no significant sustained divergence in performance from actual sales after the hypothetical treatment date, as anticipated (figure available upon request).

We also implement among-unit placebo tests that apply a hypothetical recall in November 2009 to each of the 70 sub-brands in the donor control pool, expecting that hypothetical treatment should not create significant divergence in post-treatment performance for the majority of units. As in Figure 3, Figure 11 graphs the difference between synthetic and actual sales for each placebo sub-brand (70 dashed lines) and the same for the original baseline Camry synthetic (solid blue line).[12] Although it is not possible to distinguish individual placebo synthetics' lines in the figure given the large number, it is clear that in each time period some of the potential control sub-brands' performance exhibits greater divergence from its own synthetic than does the Camry sub-brand. This suggests that any deviation in the synthetic Camry's sales relative to the actual Camry's sales is not sufficiently distinct to draw inferences from.

<Insert Figure 11 here>

As with the Chrysler synthetic analysis, we can use these placebo synthetics to assess where the Camry falls in the distribution of estimated treatment effects. For each placebo synthetic, we calculate the ratio of the post-treatment RMSPE to the pre-treatment RMSPE as a scale-independent treatment extremity measure signed by the direction of the effect. Figure 12 displays these as a dot plot for all 70 donor pool synthetics and for the Camry synthetic (we only label the Camry given space constraints). The majority of the observations fall in the

[12] As in Figure 3, we exclude placebo synthetics for control sub-brands with mean-squared prediction errors (MSPE) in the pre-intervention period greater than 5 times that of the Camry.

range of [-2, +2], with a few sub-brands falling outside. The Camry ratio of 0.8 is close to the middle of the range, implying that any positive change in sales of the Camry in the post recall period, relative to its synthetic, was not meaningfully distinct from the pre-recall period.

<Insert Figure 12 here>

The absence of a statistically significant effect is confirmed by constructing p-values based on permutation inference across these placebos' treatment extremity measures. Since 19 of 71 sub-brands have a greater extremity measure than Camry, the p-value for a more positive effect on Camry sales in the post-treatment period is 0.27. Since 51 of the 71 sub-brands have a lower positive or even negative effect, the p-value for a more negative effect on Camry sales in the post-treatment period is 0.73. Neither of these one-sided test values is statistically significant at conventional levels. Hence, the change in Camry sales after the recall does not appear to be measurably distinct from the changes in other auto sub-brands where recalls did not occur.

Overall, the synthetic control analysis, which creates a counterfactual specifically for the Camry based on a weighted combination of three sub-brands not subject to the recall, does not find evidence of a significant impact of the November 2009 recall event on subsequent sales volumes. Further analysis we conducted (available upon request) finds that the absence of a measurable recall effect is not unique to the Camry but applies also to other Toyota sub-brands where we were able to construct a well-fitted synthetic match, some of which had been recalled (e.g. Avalon) and others that had not (e.g. Corolla). These results strengthen confidence in the null finding for the Camry sub-brand, and complement organization-level event study findings (Gokhale et al. 2014) and qualitative case study analyses (Liker and Ogden 2011).

5 CONCLUSION

Qualitative and quantitative empirical methodologies play a central role in advancing organizational research, enabling scholars to test and refine theoretically-motivated relationships, and to move beyond simple correlations towards causal inferences. In this paper we have assessed a new addition for the organizational researcher's toolkit, synthetic control, which fills an existing gap by estimating the impact of a phenomenon or event in samples with limited numbers of observations, and by estimating individual unit-specific treatment affects rather than average treatment effects. It has the ability to quantify the impact of a single occurrence of an event by constructing a unique counterfactual unit, which makes it analogous to event studies, albeit with flexibility in outcome variables beyond abnormal stock price returns. Difference-in-differences and matched sampling methods also quantify treatment effects, though they demand more extensive data samples for implementing statistical tests of significance, and estimates are for average effects within the sample population. Synthetic control thus has a specific methodological niche, which allows organizational researchers to address questions in particular contexts where existing empirical techniques are restricted or not well suited.

We illustrate the method's application and approach to robustness testing in two organizational contexts where ex ante predictions about the magnitude and direction of the phenomena's effects were uncertain. Our finding of a negative impact of active government intervention in Chrysler's management, following the financial rescue in late 2008, on vehicle sales volumes provides the first quantitative estimate in debates about the auto bailout, which have largely focused on the related question of whether Chrysler would have survived in the absence of government financial support (Goolsbee and Krueger 2015). Our synthetic counterfactual analysis suggests that, conditional on its survival, Chrysler may have sold more vehicles from 2009 to 2011 without direct government involvement in management decisions. This finding contributes to an extensive literature on the performance implications of government intervention in business (Dewenter and Malatesta 2001). Whereas empirical studies in this literature have examined the impact of government ownership on organization profitability, operating costs, and asset investment levels, this study is one of the first analyses, as far as we are aware, that addresses the effect on unit sales.

Our second application of synthetic control suggests that the consequences of Toyota's unintended acceleration crisis for Camry sales were not as dramatic as media reporting might have implied, and that Toyota was able to weather the storm. A natural extension would be to replicate the synthetic control analysis for each of Toyota's recalled sub-brands, which could yield insights into attributes associated with greater or lesser sensitivity to the recall. We attempted this but good synthetic matches could not be achieved for other recalled vehicles, nor for Toyota overall, which illustrates limitations of the technique.

Like all methodologies, synthetic control comes with a series of assumptions and caveats, which determine the situations in which it is applicable, and which temper interpretation of estimated results. Robustness testing is particularly important given that inferences are based on small data samples. Nevertheless, careful implementation of synthetic control has the potential to yield novel empirical insights when judiciously applied to selected organizational research contexts.

REFERENCES

Abadie A, Diamond A, Hainmueller J (2010) Synthetic Control Methods for Comparative Case Studies: Estimating the Efect of California's Tobacco Control Program. *J. Amer. Statist. Assoc*. 105(490):493-505.

Abadie A, Diamond A, Hainmueller J (2015) Comparative Politics and the Synthetic Control Method. *Amer. J. Political Sci*. 59(2):495-510.

Abadie A, Gardeazabal J (2003) The Economic Costs of Conflict: A Case Study of the Basque Country. *Amer. Econom. Rev*. 93(1):112-132.

Anginer DA, Warburton AJ (2014) The Chrysler Effect: The Impact of Government Intervention on Borrowing Costs. *J. Banking Finance* 40:62-79.

Bertrand M, Duflo E, Mullainathan S (2004) How Much Should We Trust Differences-in-Differences Estimators. *Quart. J. Econom.* 119(1):249-275.

Billmeier A, Nannicini T (2013) Assessing Economic Liberalization Episodes: A Synthetic Control Approach. *Rev. Econom. Statist.* 95(3):983-1001

Black B (2010) The U.S. as "Reluctant Shareholder": Government, Business and the Law. *Entrepreneurial Bus. Law J.* 5(2):561-593

Bromiley P, Marcus AA (1989) The Deterrent to Dubious Corporate Behavior: Profitability, Probability, and Safety Recalls. *Strategic Management J.* 10(3):233-250.

Bunkley N (2010) Toyota Settles over California Deaths. *New York Times* (Sept 18).

Camuffo A, Weber DR (2012) The Toyota Way and the Crisis: A New Industrial Divide? Ciravegna L, ed. *Sustaining Industrial Competitiveness After the Crisis* (Palgrave MacMillan, UK), 57-103.

Card D, Krueger A (1994) Minimum Wages and Employment: A case study of the fast food industry in New Jersey and Pennsylvania. *Amer. Econom. Rev.* 84(4):772-784.

Chen Y, Ganesan S, Liu Y (2009) Does a firm's product recall strategy affect its financial value? An examination of strategic alternative during product-harm crisis. *J. Marketing*, 73(6):214-226.

Chrysler (2012) Proxy Statement. http://pdf.secdatabase.com/299/0001193125-13-096234.pdf.

Dahlander L, O'Mahony S (2011). Progressing to the Center. *Organ. Sci.* 22(4):961-979.

Davidson WN, Worrell DL (1992) The Effect of Product Recall Announcements on Shareholder Wealth. *Strategic Management J.* 13(6):467-473.

De Figueiredo RJP, Meyer P, Rawley, E (2013) Inherited Agglomeration Effects in Hedge Fund Spawns. *Strategic Management J.* 34(7):843-862.

Dehejia RH, Wahba S (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *Rev. Econom. Statis*. 84(1):151-161.

Dewenter KL, Malatesta PH (2001) State-Owned and Privately Owned Firms: An Empirical Analysis of Profitability, Leverage, and Labor Intensity. *Amer. Econom. Rev.*, 91(1):320-334.

Donohue JJ, Aneja A (2011) New Evidence in the 'More Guns, Less Crime' Debate: A Synthetic Controls Approach Working Paper: Stanford Law School, Stanford CA.

Eisenhardt KM (1989) Building Theories from Case Study Research. *Acad. Management Rev.* 14(4):532-550.

Ernst MD. (2004) Permutation Methods: A Basis for Exact Inference. *Statist. Sci.* 19(4):676-685.

Fabrizio K (2012) Institutions, Capabilities, and Contracts: Make or Buy in the Electric Utility Industry. *Organ. Sci.* 23(5):1264-1281.

Freedman S, Kearney M, Lederman M (2012) Product Recalls, Imperfect Information, and Spillover Effects: Lessons from the Consumer Response to the 2007 Toy Recalls. *Rev. Econom. Statist.* 94(2):499-516.

Gibbert M, Ruigrok W, Wicki B (2008) What Passes as a Rigorous Case Study? *Strategic Management J.* 29:1465–1474

Gokhale J, Brooks RM, Tremblay VJ (2014) The Effect on Stockholder Wealth of Product Recalls and Government Action: The Case of Toyota's Accelerator Pedal Recall. *Quart. Rev. Econom. Finance*, 54(4):521-528.

Goolsbee AD, Krueger AB (2015) A Retrospective Look at Rescuing and Restructuring General Motors and Chrysler. *J. Econom. Perspect.* 29(2):3-24.

Govindaraj S, Jaggi B, Lin B (2004) Market Overreaction to Product Recall Revisited—The Case of Firestone Tires and the Ford Explorer. *Rev. Quant. Finance Accounting*, 23(1):31-54.

Jarrell G, Peltzman S (1985) The Impact of Product Recalls on the Wealth of Sellers. *J. Political Econom.* 93(3): 512-536.

Kivleniece I, Quelin BV (2012) Creating and Capturing Value in Public-Private Ties: A Private Actor's Perspective. *Acad. Management Rev.* 37(2):272-299.

Liker JK, Ogden TN (2011) *Toyota Under Fire* (McGraw Hill: New York).

Lui Y, Shankar V (2015) The Dynamic Impact of Product-Harm Crises on Brand Preference and Advertising Effectiveness: An Empirical Analysis of the Automobile Industry. *Management Sci.* Forthcoming.

Mahoney J, McGahan A, Pitelis C (2009) The Interdependence of Private and Public Interests. *Organ. Sci.* 20(6):1034-1052.

March JG, Sproull LS, Tamuz M (1991) Learning from Samples of One of Fewer. *Organ. Sci.* (2):1,1-13.

McWilliams A, Siegel D (1997) Event Studies in Management Research: Theoretical and Empirical Issues. *Acad. Management J.* 40(3):626-657.

NHTSA (2015) *Recalls Database.* http://www-odi.nhtsa.dot.gov/downloads/flatfiles.cfm.

Office of the Press Secretary of the White House (2008) *Fact Sheet: Financing Assistance to Facilitate the Restructuring of Auto Manufacturers to Attain Financial Viability*. (Dec 19).

Rawley E, Simcoe TS (2013) Information Technology, Productivity, and Asset Ownership: Evidence from Taxicab Fleets. *Organ. Sci.* 24(3):831-845.

Rattner S (2011) *Overhaul. An insider's account of the Obama administration's emergency rescue of the auto industry*. (Houghton Mifflin Harcourt Publishing Company, New York).

Rhee M, Haunschild PR (2006). The Liability of Good Reputation: A Study of Product Recalls in the U.S. Automobile Industry. *Organ. Sci.* 17(1):101-117.

Roe MJ, Skeel D (2010) Assessing the Chrysler Bankruptcy. *Michigan Law Rev.* 108:721-771.

Rosenbaum PR (2002a) *Observational Studies* (2nd ed.) (Springer, New York).

Rosenbaum PR (2002b) Covariance Adjustment in Randomized Experiments and Observational Studies. *Statist. Sci.* 17(3):286–327.

Runde J, de Rond M (2010) Evaluating causal explanations of specific events. *Organ. Stud.* 31(4):431-450.

Shah SK, Corley KG (2006) Building Better Theory by Bridging the Quantitative-Qualitative Divide. *J. Management Stud.* 43(8):1821-1835.

SIGTARP (Special Inspector General for the Troubled Asset Relief Program) (2010). *Factors Affecting the Decisions of General Motors and Chrysler to Reduce their Dealership Networks*. SIGTARP-10-008.

SIGTARP (Special Inspector General for the Troubled Asset Relief Program). (2013) *Treasury's role in the decision for GM to provide pension payments to Delphi employees*. SIGTARP-13-003.

Toh PK, Kim T (2013) Why Put All Your Eggs in One Basket? A Competition-Based View of How Technological Uncertainty Affects a Firm's Technological Specialization. *Organ. Sci.* 24(4):1214-1236.

Whoriskey P (2009) Auto Retirees Brace for Hardship; Pension Benefits May Suffer in Bankruptcy. *Washington Post*. (Apr 24).

Table 1: Comparison of Synthetic Control and other Treatment Effect Estimation Methods

Method	Typical Data Characteristics: Sample Size Required (Number of Units)	Number of Treated Units Required	Type of Dependent Variable	Time Series Duration	Assessment of Reliability of Estimated Treatment Effect	Strengths	Limitations
Synthetic Control	Few (often <50)	One or greater	Single-dimensional	Panel data: multiple periods before and after event	Randomization inference based on placebo tests	• Provides unit-specific quantitative estimate of treatment effect for each treated unit, even in a small population • Addresses potential endogeneity of treatment when good fit between synthetic and actual unit achieved during pre-treatment period • Provides time varying estimate of treatment effect rather than an average effect over time for all units.	• Not always possible to achieve a good pre-treatment fit between synthetic and actual unit • Cannot be applied to 'outlier' units on outcome variable • No standard has emerged in the literature to assess quality of model fit or to compare models with different outcome variables
Comparative Case Study	Very few	Very few	Multi-dimensional	Flexible	Not applicable	• Permits in-depth, multi-faceted exploration of phenomena and causal mechanisms	• Selection of cases can be subjective
Difference-in-Differences	Many (typically > 200)	Few or Many	Single-dimensional	Panel data: multiple periods before and after event	Frequentist inference (T-test of coefficient on treatment variable)	• Controls for both observable factors and unobservable factors common to all units across time or to a specific unit if it is time-invariant • Simple statistical test of estimate's precision, and standard statistical measure of overall model fit (R-squared)	• Large sample of units required • Estimates *average* treatment effects (across units and time) • Assumes no difference in trends between treated and untreated groups in standard model
Matched Sample	Many (typically > 200)	Half the sample in exact method of matching	Single-dimensional	Cross-sectional data (matching at a single point in time) or pooled panel data	Frequentist inference (T-test)	• Addresses potential endogenous selection into treatment by matching on observable attributes • No time dimension to the data is required	• Large sample of units required • Estimates *average* treatment effects (across units only) • Matches may not control for unobserved factors • Estimates can be sensitive to choice of matching algorithm
Event Study	Market or industry	One or greater	Abnormal Stock returns	Panel data: short time window around event	Frequentist inference (T-test)	• Can be applied to a single treatment event • Estimate of treatment effect controls for other common market or industry factors	• Limited to publicly traded firms, and to a single type of dependent variable (stock returns) • Can be difficult to identify the first public reporting of an event

Table 2: Weights of Companies in Synthetic Chrysler

Company	***Weight***	***Company***	***Weight***
BMW	0	Mitsubishi	0
Daimler	0.022	Nissan	0.169
Ford	0.664	Porsche	0
General Motors	-	Saab	0
Honda	0	Subaru	0
Hyundai-Kia	0	Suzuki	0
Isuzu	0.068	Toyota	0.077
Jaguar Land Rover	-	Volkswagen-Audi	0
Mazda	0	Volvo	0

Notes:
(1) General Motors is excluded as a match because it also received government assistance
(2) Jaguar Land Rover is excluded as match because of limited data availability

Table 3: Comparison of Attributes between Chrysler, Ford, and Synthetic Chrysler

	(Averages of Monthly Figures over 48 months prior to acceptance of government assistance)		
Attribute	***Chrysler***	***Synthetic Chrysler***	***Ford***
Price, Average of Vehicles Sold	26843	28241	28384
MPG, Average of Vehicles Sold	19.9	20.4	19.4
MPG, Maximum of Vehicles Sold	31.5	34.3	34.6
Weight (in lbs), Avg. of Vehicles Sold	4181	4175	4422
Engine Size (in L), Avg. of Vehicles Sold	3.95	4.00	4.35
Fraction of Sales in SUV/Truck/Van	0.751	0.612	0.657
Brands, # Active	4.3	2.6	3.0
Platforms, # Active	15.8	15.3	18.1
Segments of Market, # Active	15.3	14.6	16.0
Series of Vehicles, # Active	22.3	21.2	24.3
Fraction of Sales Manufactured Outside North America	0.008	0.103	0.000
Leverage (Debt/Assests)	0.249	0.328	0.391
# of Employees	277211	268829	323050
Sales Volume	178759	193013	250620
Level of Sales Volume, Index (Dec '08=100)	200.1	267.5	191.9

Table 4: Sub-Brands weights in the Synthetic Camry

Subseries	***Weight***	***Subseries***	***Weight***
Acura MDX	0	Kia Sorento	0
Acura TL	0	Kia Spectra	0
Audi A4	0	Kia Sportage	0
BMW 3 Series	0	Lincoln Navigator	0
BMW 5 Series	0	Lincoln Town Car	0
BMW X5	0	Mazda 3	0
Ford Club Wagon	0	Mazda 6	0
Ford Crown Victoria	0	Mazda Tribute	0
Ford Econoline	0	Mercedes C Class	0
Ford Edge	0	Mercedes M Class	0
Ford Escape	0	Mercury Grand Marquis	0
Ford Expedition	0	Mercury Mountaineer	0
Ford Explorer	0	Mercury Sable	0
Ford F-150	0.199	Mini Cooper	0
Ford Focus	0	Mitsubishi Eclipse	0
Ford Fusion	0	Mitsubishi Galant	0
Ford Mustang	0	Mitsubishi Lancer	0
Ford Ranger	0	Nissan Altima	0
Ford Taurus	0	Nissan Frontier	0
Ford Windstar	0	Nissan Maxima	0
Honda Accord	0.272	Nissan Murano	0
Honda Civic	0.529	Nissan Pathfinder	0
Honda CR-V	0	Nissan Quest	0
Honda Element	0	Nissan Rogue	0
Honda Fit	0	Nissan Sentra	0
Honda Odyssey	0	Nissan Titan	0
Honda Pilot	0	Nissan Versa	0
Hyundai Accent	0	Nissan Xterra	0
Hyundai Santa Fe	0	Subara Impreza	0
Hyundai Sonata	0	Subaru Forester	0
Hyundia Elantra	0	Subaru Legacy	0
Infiniti G35	0	Suzuki Xl-7	0
Kia Optima	0	VW Beetle	0
Kia Rio	0	VW Jetta	0
Kia Sedona	0	VW Passat	0

Table 5: Comparison of Attributes between Camry and Synthetic Camry

Attribute	***Camry***	***Synthetic Camry***	***Average of 70 control sub-brands***
Price	24008.17	22740.82	26732.31
MPG, Maximum	30.16	32.02	26.02
Weight (in lbs)	3342.02	3464.15	3856.36
Length (in inches)	190.52	191.80	193.44
Width (in inches)	71.286	71.90	73.13
Engine Size (in L.)	2.83	2.83	3.44
Engine Displacement (in CID)	170.84	172.315	209.07
Number of Valves per Cylinder	4.00	3.75	3.51
Sales Volume	35370.25	35235.49	20230.02
Level of Sales Volume, Index (Dec ’09 =100)	123.73	139.39	228.09

Figure 1: Chrysler's & Synthetic's Sales Volumes

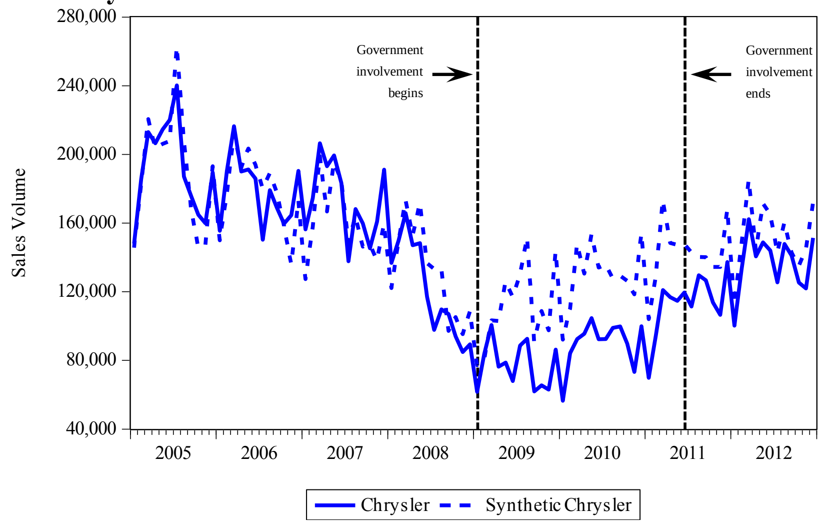


Figure 2: Difference between Chrysler's Actual & Synthetic Sales Volumes

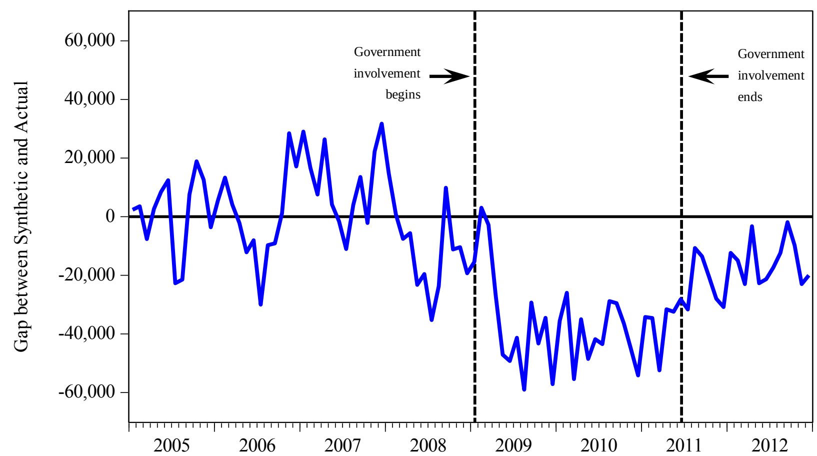


Figure 3: Placebo Test Among Untreated Firms. Difference between Actual & Synthetic Sales Volumes

(Excluding Firms with MSPE 5x higher than Chrysler's prior to Government Involvement)

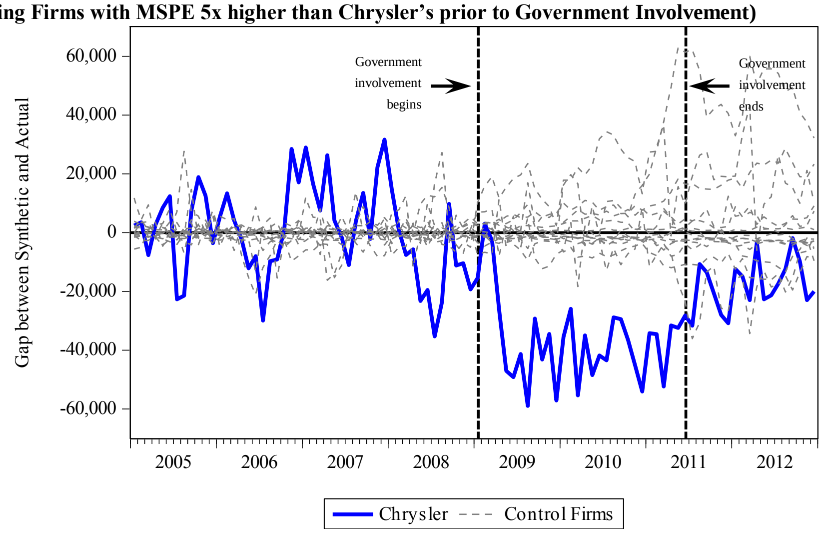


Figure 4: Dot plot of Treatment Extremity Measure, Among-Firm Placebos' Post-Treatment/ Performance Relative to Pre-Treatment Performance

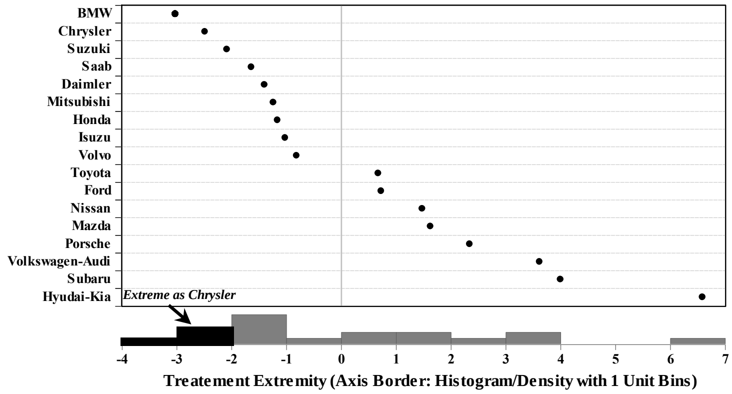


Figure 5: Placebo Test In Time

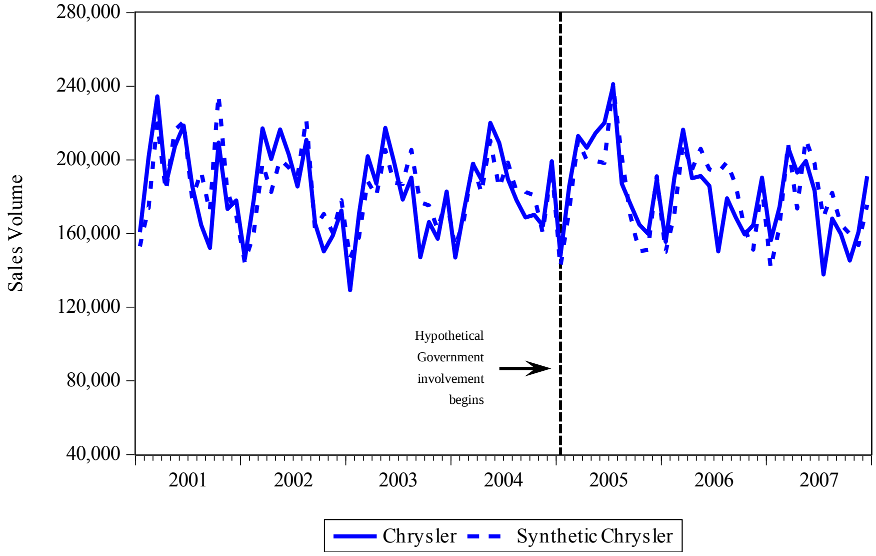


Figure 6: Leave-One-Out Tests, Differences

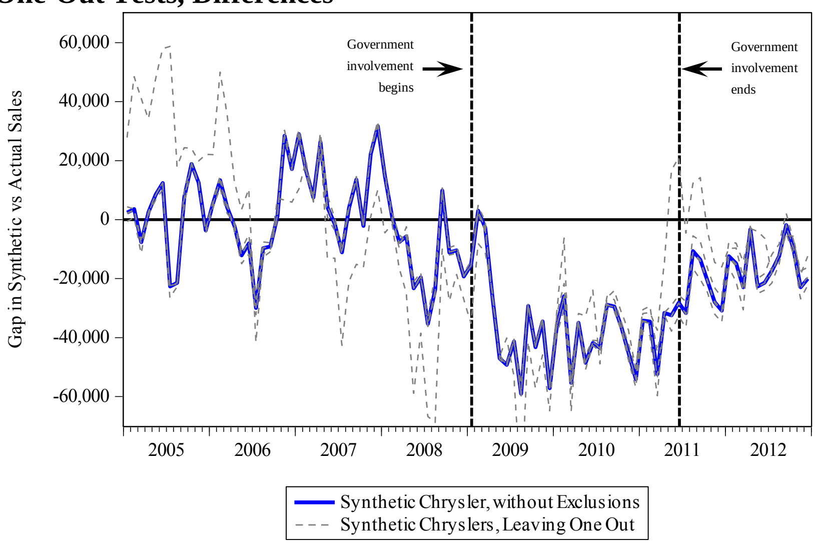


Figure 7: Out-of-Sample Test, Differences

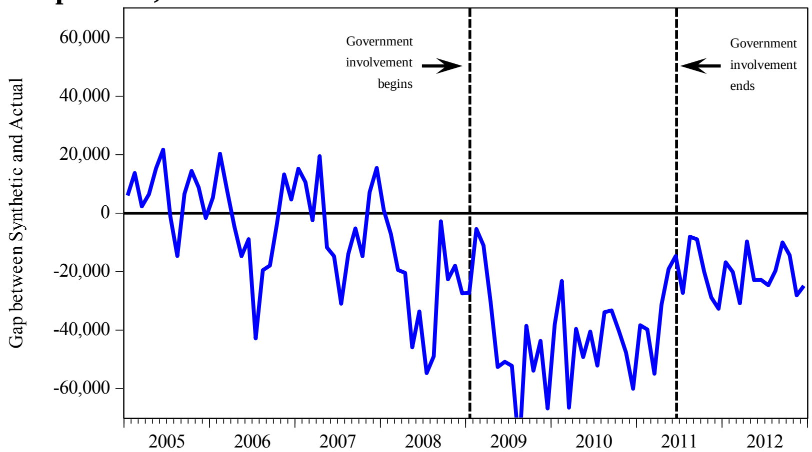


Figure 8: Camry's and Synthetic's Sales Volumes

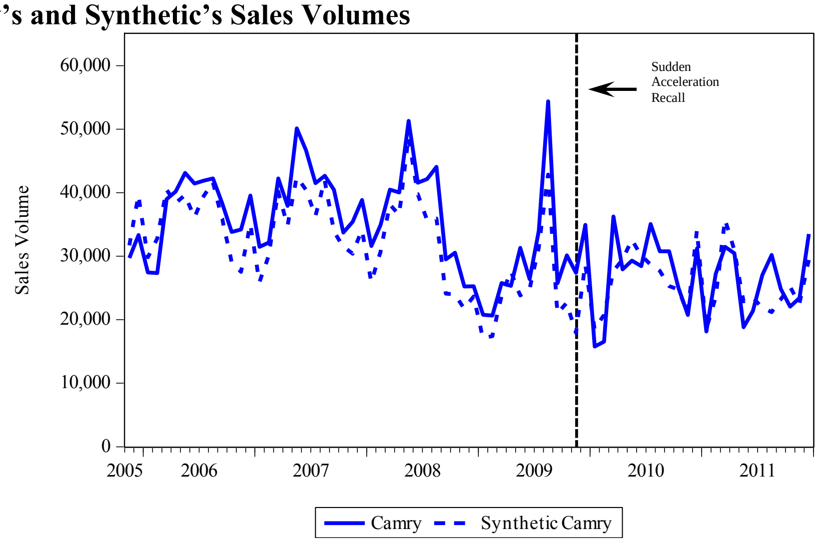


Figure 9: Difference between the Camry's Actual & Synthetic Sales Volumes

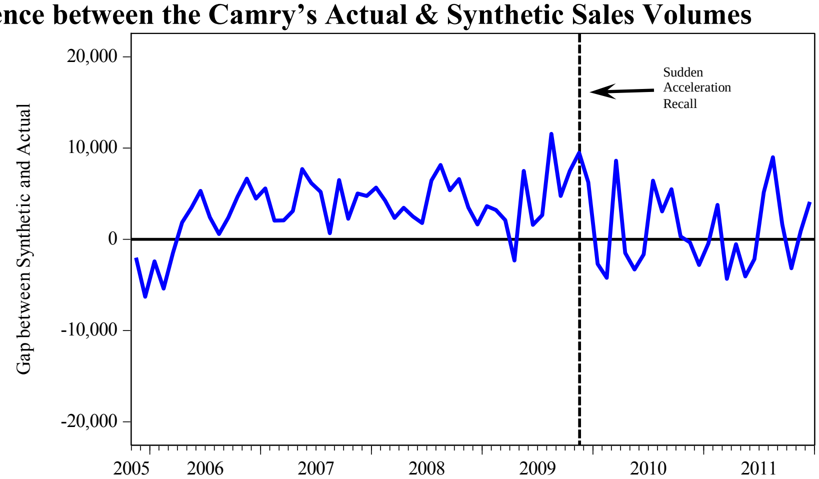


Figure 10: Leave-One-Out Tests, Differences

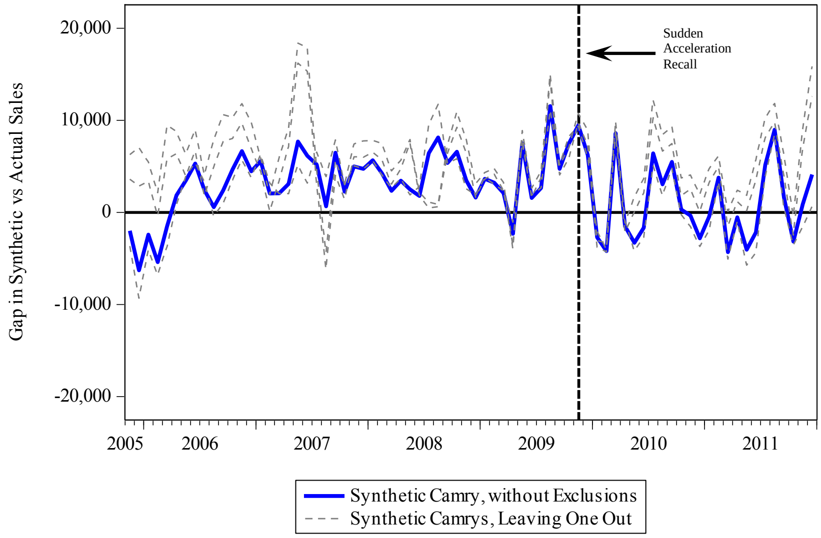


Figure 11: Placebo Test Among-Units

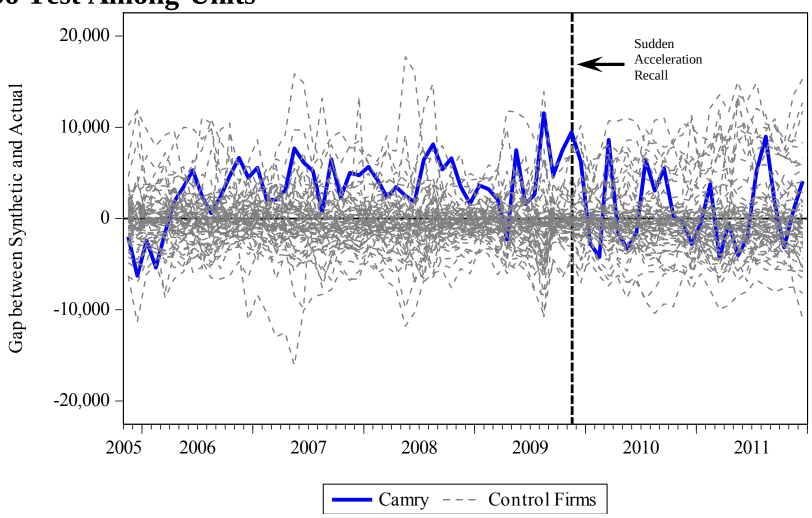
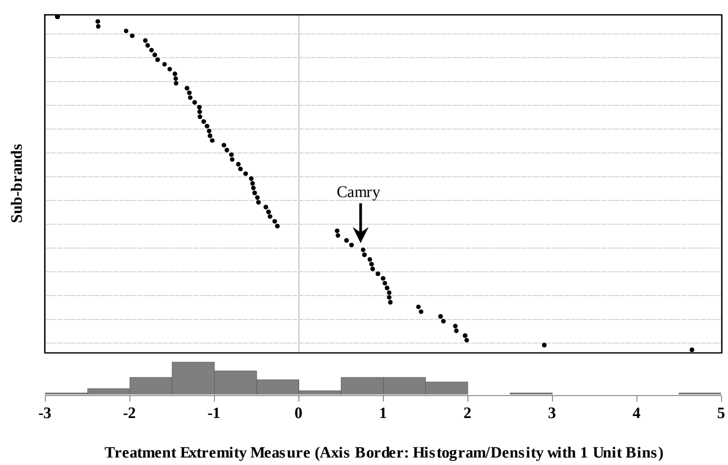


Figure 12: Dot plot of Treatment Extremity Measure, Among Sub-brands Placebos' Post-Treatment Performance Relative to Pre-Treatment Performance



APPENDIX A: DATA DESCRIPTION AND SOURCES

Variable	Description	Source
Light vehicle sales volume	Total units sold by firm of new vehicles in the U.S., monthly	*Ward's Auto: U.S. Model Car Specifications and Prices*
Average price	Average retail price of all vehicles sold (U.S. dollars), monthly	Author's calculations based on data from *Ward's Auto: U.S. Model Car Specifications and Prices*
Average fuel economy	Average fuel economy of all vehicles sold in the U.S. (miles per gallon), monthly	Author's calculations based on data from *Ward's Auto: U.S. Model Car Specifications and Prices*
Maximum fuel economy	Maximum fuel economy of all vehicles sold in the U.S. (miles per gallon), monthly	Author's calculations based on data from *Ward's Auto: U.S. Model Car Specifications and Prices*
Average weight	Average weight of cars sold in the U.S. (lbs.), monthly	Author's calculations based on data from *Ward's Auto: U.S. Model Car Specifications and Prices*
Average engine size	Average size of engines of all vehicles sold in the U.S. (Liters), monthly	Author's calculations based on data from *Ward's Auto: U.S. Model Car Specifications and Prices*
Number of active production platforms	Total number of production platforms for vehicles sold in the U.S., monthly	Author's calculations based on data from *Ward's Auto: U.S. Model Car Specifications and Prices*
Number of active brands	Total number of brands marketed by firm in the U.S. (e.g. In 2011, Chrysler markets the Chrysler, Dodge, Jeep, Ram and Fiat branded vehicles), monthly	Author's calculations based on data from *Ward's Auto: U.S. Model Car Specifications and Prices*
Number of active series	Total number of series marketed within a brand by a firm in the U.S. (e.g. In 2011, Chrysler branded vehicles include four series: 200, 300, Sebring and the Town and Country.), monthly	Author's calculations based on data from *Ward's Auto: U.S. Model Car Specifications and Prices*
Number of market segments	Total number of market segments that a firm competes within in the U.S. (e.g. In 2011, Chrysler competed within 14 different segments, including small specialty, middle specialty, upper small, upper middle, small SUV, small CUV, small van, middle SUV, middle CUV, small pickup, medium duty, large SUV, large regular, and large pickup), monthly	Author's calculations based on data from *Ward's Auto: U.S. Model Car Specifications and Prices*
Fraction of vehicles sold from the SUV, light truck, and van segments	Proportion of total U.S. sales that are made in the larger vehicle segments, monthly	Author's calculations based on data from *Ward's Auto: U.S. Model Car Specifications and Prices*
Fraction of vehicles manufactured outside North America	Proportion of total U.S. sales of imported vehicles, monthly	Author's calculations based on data from *Ward's Auto: U.S. Model Car Specifications and Prices*
Number of employees	Total number of worldwide employees, annual	*Compustat*
Leverage ratio	Total Long-term debt / Total Assets, annual	*Compustat*