Making Causal Inferences in Small Samples using Synthetic Control Methodology: Did Chrysler Benefit from Government Assistance?

Adam R. Fremeth‡, Guy L. F. Holburn*, Brian Kelleher Richter†

15 November 2013

Abstract
We introduce synthetic control analysis to management research. This recently developed statistical methodology overcomes challenges to causal inference in contexts constrained by small samples or few occurrences of the phenomenon of interest. Synthetic control constructs a replica of a focal firm, or other observation unit, based on a weighted combination of untreated firms with similar attributes within the sample population. The method quantifies the magnitude and direction of a treatment effect by comparing the actual performance of a focal unit to its counterfactual replica without treatment. As an illustration, we assess the impact of government intervention in the auto sector on the performance of Chrysler which, following the financial crisis, accepted government support in return for Treasury oversight. The synthetic Chrysler we construct—representing the firm’s estimated performance without government intervention—sold 29% more vehicles in the U.S. than did the actual firm during the intervention period.

Keywords: Synthetic Control, Counterfactual Methodology, Government Intervention, TARP, Automobile Industry, Case Study Methods

‡ Adam R. Fremeth. afremeth@ivey.uwo.ca; Assistant Professor of Business, Economics, and Public Policy; Ivey School of Business, University of Western Ontario, London, ON N6A 3K7 CANADA

* Guy L. F. Holburn. gholburn@ivey.uwo.ca; Associate Professor of Business, Economics, and Public Policy and Suncor Chair in Energy Policy; Ivey School of Business, University of Western Ontario, London, ON N6A 3K7 CANADA

† Brian Kelleher Richter. brian.richter@mccombs.utexas.edu. Phone (USA): 310-709-5745. Web: http://briankrichter.com/. Assistant Professor of Business, Government, and Society; McCombs School of Business; University of Texas at Austin; Austin, TX 78705 USA.

**Fremeth acknowledges support from the Social Science and Humanities Research Council of Canada (SSHRC) Grant R4636A02. We have received valuable comments on this research from Dave Primo, Jeff Timmons, PK Toh, Myles Shaver, Rob Seamans, Claus Rerup, Francine LaFontaine, Brian Hong, Romel Mostafa, Ken Schotts, John deFigueiredo, and Tima Bansal. Prior versions of this paper have been presented at the American Law and Economics Association, Harvard Business School, INFORMS, the Industry Studies Association, and the International Society for New Institutional Economics. Any errors in this draft are our own.
1 INTRODUCTION

Management scholars debate the relative merits of qualitative and quantitative empirical methodologies (Flyvbjerg 2006; Shah & Corley 2006). Given few opportunities to apply experimental research designs, which are the scientific ideal, social scientists are left with observational data requiring them to make tradeoffs between accuracy, generality, and simplicity (Throngate 1976; Weick 1979). In some instances, large samples of comparable statistical units cannot be constructed—precluding traditional quantitative approaches and constraining the methodological choices to qualitative ones. Despite yielding deep insights into causal process mechanisms, it is difficult for qualitative approaches to estimate the magnitude of a phenomenon’s effect. Hence, some questions beg for a new method.

In this paper, we introduce to management research the recently-developed synthetic control methodology (Abadie & Gardeazabal 2003); this quasi-experimental approach rises to March et al.’s (1991) call to learn from samples of one (or fewer), while bridging the quantitative-qualitative methodological divide, and opening possibilities to resolve new questions. The synthetic control technique is particularly valuable in contexts where limited sample sizes or few occurrences of the phenomenon of interest preclude researchers from making strong inferences using traditional regression techniques. The synthetic control method generates a counterfactual statistical unit, i.e. a synthetic clone of the focal unit that behaves as if it was not subject to an intervention or treatment phenomenon that has been applied to the focal unit.\(^1\) This counterfactual unit comes complete with its own data on performance and descriptive attributes.\(^2\) The synthetic is constructed based on a weighted average of other observation units in the population that (i) were not subject to the treatment phenomenon that is the focus of analysis, and (ii) have observable performance and descriptive data. The synthetic unit will closely mimic the focal unit in both its descriptive characteristics and pre-treatment performance. Hence, the synthetic will provide a better counterfactual than any single quantitative observation or qualitative case. Researchers can compare the relative performance of the focal observation unit and the

---

\(^1\) Another statistical technique that attempts to construct a representative counterfactual from existing data is the ‘chop shop’ method used to estimate the diversification discount in the finance literature (Lang & Stulz 1994; Villalonga 2004; Laeven & Levine 2007).

\(^2\) Both synthetic control and simulation methods generate data; however, they differ in that in synthetic control researchers need not make assumptions about the underlying data-generating process via formal mathematical models of agent behavior. In synthetic control, the core assumption about the data generating process is that a treated unit shares the same data-generating process as control units with like attributes.
counterfactual synthetic unit to estimate the magnitude and direction of divergence attributable to the phenomenon of interest.

We apply the new methodology to a management question stemming from U.S. government intervention in the auto industry beginning in late 2008: what is the impact on subsequent performance at Chrysler, or its surviving components, of government intervention and of the company accepting financial assistance in return for ongoing Treasury oversight of managerial actions? Media personalities and politicians have debated the broader topic widely, concocting competing stories for why government intervention either succeeded or failed (Romney 2012; Stewart 2012). Nevertheless, we are not aware of another academic paper that attempts to provide a quantitative estimate of the net effect of the auto assistance programs—perhaps given the inadequacy of traditional quantitative approaches in this context. Sufficiently precise quantitative estimates generated using either differences-in-differences or matched sample methods are precluded due to the small number of firms that are auto manufacturers and the even smaller number of firms that took government assistance. Event studies using stock market data are precluded because the focal firm, Chrysler, was privately-owned. Synthetic control is well suited to our research context given the limited population of firms in the auto industry and the availability of balanced panel data on firm performance measures with a relatively long time-dimension (144 months). We estimate that Chrysler sold 29% fewer vehicles than its synthetic counterfactual—representing the firm’s performance without government involvement—during the period of active government involvement, suggesting that any managerial benefits of government guarantees or cheap capital were outweighed by potential costs. To complement the statistical analysis we examine potential costs in a qualitative exploration of mechanisms later.

Given the limited applicability of large sample statistical methodologies to answer our question, management scholars might be left with few non-qualitative tools. A simple approach would be to conduct a single case study examining the direction of Chrysler’s performance subsequent to accepting government assistance; however, that analysis would be flawed as it lacks any counterfactual. Hence,

---

3 Republican presidential candidate, Mitt Romney (2012) famously quipped: “The president tells us that without his intervention things in Detroit would be worse. I believe that without his intervention things there would be better.”
comparative case analysis would be a more appropriate tool. In that approach, a researcher would want to identify the most similar automobile company that did not accept government assistance to use as the counterfactual evidence.\(^4\) This comparative case could be used to draw conclusions about the direction, and size, of the effect of Chrysler’s acceptance of government assistance. If we were to select the best counterfactual for Chrysler among existing firms, we might choose Ford as it was the only one of the Big Three U.S. automakers that did not take government assistance. A simple graphical comparison of Chrysler’s and Ford’s unit sales in Figure 1 suggests that Chrysler substantially underperformed Ford once it began receiving government assistance.

<Insert Figure 1 Here>

Causal claims drawn from Figure 1 would be impossible to support—even if they were limited to claims about the direction of the effect of government assistance on Chrysler and not its magnitude. Ford would be easy to critique as the only counterfactual, or control unit, for a variety of reasons: despite similarities with Chrysler, Ford sells different vehicles with different attributes at different price points, uses slightly different technology in its manufacturing process, has different technical expertise, and has a different marketing and branding program than does Chrysler. We will quantify these differences between Ford and Chrysler and show what will make an even better comparison, using synthetic control, later in this paper. Any one of the attributes that makes Ford different than Chrysler could have affected the companies’ relative sales in the period after December 2008, making it hard to attribute a performance differential to Chrysler’s acceptance of government assistance alone. That is unless we could compare Chrysler’s performance with a firm’s that was a better match than Ford.

If there were a company that matched Chrysler better along more dimensions than Ford, it would provide a more ideal comparison for evaluating the impact of government assistance on Chrysler’s sales. Since other automakers may be more similar to Chrysler on some dimensions than Ford, our analysis might be more accurate were we to apply the logic of multiple case analyses (Eisenhardt 1989) to

\(^4\) While choosing the next most similar company is one comparative case selection logic that is particularly useful when researchers are trying to make inferences about causal direction and magnitudes, another case selection logic suggests choosing the most extremely different case when trying to make inferences about underlying mechanisms.
construct an average of all valid comparator companies’ performance in the absence of government assistance.

Better yet, there would be a way to take a weighted average of several automakers similar to Chrysler in which we gave heavier weights to companies that are most similar along the most important dimensions for predicting Chrysler’s performance. Synthetic control allows researchers to determine systematically what that weighted average should be and how important various attributes of a company are in predicting performance. Hence, the methodology allows us to construct a counterfactual Chrysler that does not accept government assistance based on a sample of control companies that did not do so. We can then compare Chrysler’s actual performance to its synthetic counterfactual’s. When we do, the estimate of Chrysler’s underperformance is not as severe as what we would have estimated had we used Ford as the sole counterfactual, suggesting that synthetic control methods may introduce less bias than single comparative case studies.

In the next section we discuss in more detail the operation of the synthetic control method, its usage to date in cognate disciplines’ research, its advantages, and its limitations. In section 3, we demonstrate the method’s value to management researchers in our application where we quantify the performance impact of a firm’s decision to accept conditional government assistance. In the final sections we explore reasons that may account for the estimated shortfall in Chrysler’s sales; we also discuss the limitations of our analysis while providing suggestions for future research.

2 **THE SYNTHETIC CONTROL METHOD**

Synthetic control methods are beginning to gain popularity in academic fields including political economy, international economics, public policy, political science, and law—but have yet to take root in management.\(^5\) It has been used to measure the effects of: (i) domestic ETA terrorism on regional growth within the Basque region of Spain (Abadie & Gardeazabal 2003); (ii) economic liberalization on real GDP growth (Billmeier & Nannini 2013); (iii) a new tobacco tax policy in California on cigarette sales (Abadie et al. 2010); (iv) reunification of Germany on per-capita wealth in the former West Germany.

\(^5\) Marx et. al (2009) is the only work we are aware of in management even referencing synthetic control; however, they use a difference-in-difference methodology for their core analysis and only mention synthetic control in passing.
(Abadie et al. 2011); and (v) gun control laws on crime in several U.S. states (Donohue & Aneja 2012). Synthetic control quantifies the magnitude and direction of each of these effects studied by researchers in cognate disciplines and provides visual evidence by graphing divergence between the outcomes realized by focal units and their synthetic counterfactuals after key events.

Despite the breadth of topics studied in other disciplines, all of the questions researched using synthetic control are: (i) focused around a single event or application of a treatment phenomenon; (ii) have theoretical ambiguity that can only be resolved by identifying the direction and magnitude of the treatment effect with relative precision; and (iii) cannot be answered with case analyses due to the lack of obvious stand-alone counterfactuals among the few control units. Given that management researchers often share these initial conditions, opportunities abound to apply the method in this discipline. Diving deeper into how the technique works, its advantages, and its limitations should make its utility for management researchers even more explicit.

2.1 Overview of the Technique

The synthetic control technique creates a counterfactual observation unit whose performance can be compared to a focal observation unit that has undergone any sort of treatment. In a managerial context, these units are industries, firms, corporate divisions, or employees.

The counterfactual observation unit, or synthetic control unit, is constructed as a weighted average of untreated comparison or control units. The technique maximizes the ability of the synthetic unit to 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 attributes of the focal unit that are determined to be good predictors of the selected outcome variable and (ii) values of the same attributes in the synthetic unit. In essence, the underlying algorithm calculates positive or null weights 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. Those post-treatment outcomes can 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.

To implement the method, we first must build the pool of all potential comparison or control units—i.e. all similar statistical units that did not receive the treatment and for which data can be collected on \( k \) attributes which are potential predictor variables (\( X \)) for the outcome variable (\( Y \)) in question. The synthetic control technique subjects the comparison units’ predictor variables’ attribute data in the pre-treatment period to a dual optimization process that minimizes:

\[
\sum_{m=1}^{k} v_m (X_{1m} - X_{0m}W)^2
\]

by selecting the optimal values of \( W \) and \( v_m \)—where \( X_{1m} \) is the value of the \( m \)-th attribute of the focal unit; \( X_{0m} \) is a \( 1 \times j \) vector containing the values of the \( m \)-th predictor attribute of each of the \( j \) potential comparison or control units; \( W \) is a vector of weights on control units; 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. 2011).\(^6\) This optimization process minimizes prediction error between the actual and the synthetic in the pre-treatment period.

\( Y_1 \) is the observed outcome data for the focal, treated, unit. \( Y_0W \) is the synthetically generated outcome data for the synthetic unit in both the pre- and post-treatment periods; more simply, \( Y_0W \) is the weighted average of outcome variables for the included control units.

If there are no important omitted predictor variables then a reliable synthetic match will have been created such that \( Y_1 - Y_0W \)—or the distance between the actual unit’s outcome variable and the synthetic unit’s outcome variable—will be small in the pre-intervention period (Abadie et al. 2010, 2011). This is particularly likely when the pre-intervention period is sufficiently long.

Whether the value of \( Y_1 - Y_0W \) remains the same size, becomes increasingly positive, or increasingly negative in the post-treatment period allows us to make inferences about the direction and

\(^6\) More precisely, \( v_m = \arg\min_{v_m} \{ (Z_1 - Z_0W^*)(Z_1 - Z_0W^*)' \} \) where \( Z_0 \) is a \( m \times j \) vector containing values of predictor attributes for the \( J \) potential control firms, where \( W^* \) is the value of \( W \) that minimize the equation above.
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 pseudo p-value from the results of placebo tests and further validated through robustness checks aimed at falsifying underlying assumptions.

Abadie et al. (2010, 2011) provide a wealth of additional technical details and proofs supporting the underlying synthetic control methodology. Among other things, Abadie et al. (2011) 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 regression coefficients need not be restricted to doing so. Software is available in R, Matlab, and STATA to implement synthetic control methods.7

2.2 Inference in Synthetic Control

Established best practices for assessing the validity of causal inference in any statistical analysis can be applied to synthetic control. As prima facie checks, placebo-tests help ensure that results are not spurious. More important, however, are falsification tests aimed at testing underlying identification assumptions in the statistical model to ensure that the method accords well with the realities of the data in a particular context.

Despite the similar means of assessing the validity of causal inference in synthetic control models, assessing our confidence in the extremity of the outcomes we estimate in synthetic control requires a different approach from calculating frequentist p-values in traditional regression settings. Due to the small samples used in synthetic control, frequentist distributional assumptions fail to hold, so we must rely instead on the validity of randomization or permutation inference to construct, computationally, pseudo p-values which have a similar interpretation.

In this sub-section, we first discuss placebo tests, then the construction of pseudo p-values which follow naturally from the former, and finally falsification tests, all as applied to synthetic control.

7 Programs are available for download on Jens Hainmüller’s website at: http://www.mit.edu/~jhainm/synthpage.html.
2.2.1 Placebo Tests

Placebo tests are beginning to be widely used in the social sciences as a basic check on inference (e.g. Bertrand et. al 2004). The premise is simple: replicating the synthetic control analysis should not generate a significant divergence between synthetic and actual outcomes in the absence of treatment. In-time placebo tests examine whether synthetic control analysis produces large estimated effects when hypothetical treatments are applied at times other than the actual treatment date. Across-unit placebo tests are analogous, but the hypothetical treatments are applied to units that were not subject to the intervention on the actual treatment date instead. These tests bolster confidence in inferred causality when they do not produce gaps between the observed and synthetic outcomes that are as large as those for the focal unit at the actual treatment date.

2.2.2 Construction of Pseudo P-Values based on Permutation

Creating a synthetic for each untreated control unit in the population—as in the across-unit placebo test—enables researchers to ascertain whether the estimated treatment effect for the focal unit is of unique magnitude and direction. These synthetics constructed for untreated control units also provide the basis for calculating pseudo p-values. A ratio of the treatment period prediction errors to the pretreatment prediction errors on these 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 distributions of these extremity measures allows researchers to compute pseudo 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 pseudo p-values can be interpreted, nevertheless, in much the same way as p-values in traditional regression settings.

2.2.3 Falsification Tests

A final way to check the validity of inferences is by attempting to falsify the underlying assumptions of the statistical model (Popper 1959; Leamer 2010). There are two primary assumptions embedded in synthetic control analysis: (i) that the event in question only has a marked effect on the outcome \( Y_1 \) of the focal unit and not on the outcomes of the controls units \( Y_0 \) included in the synthetic; and, (ii) that the structural relationship between the included control units and the synthetic estimated in
the pre-treatment period \((W)\) can be treated as if it were approximately stable over time. *Leave-one-out tests* attempt to falsify the first of these assumptions, while *out-of-sample* tests attempt to falsify the latter.

The primary question we can ask to falsify the assumption about whether the treatment event has a marked effect on included control units’ outcomes, which could bias our estimate of the synthetic, is: does the synthetic’s performance change dramatically if we exclude particular control units from the pool of those eligible? The leave-one-out test replicates the baseline synthetic analysis several times, but excludes each control unit that was originally included in the baseline synthetic from the pool of those eligible to comprise the new synthetics on a one-by-one basis. If the performance of each of these new synthetics mirrors the baseline synthetic then: it is unlikely that the results are biased by the inclusion of any single control unit, and it is unlikely that any single included control unit’s outcome was altered markedly by the treatment event or by a by concomitant event in the post-treatment period. Verifying that included control units were largely unaffected by the treatment event allows us to interpret our synthetic control results as being the direct effect of the treatment on the focal unit only.

Leave-one-out tests cannot fully resolve whether a treatment event alters *status quo* outcomes among all potential control units since the focus is on one unit at a time; however, examining the distribution of treatment extremity measures (used to calculate pseudo p-values) may help. When the distribution is symmetric—i.e. when roughly half of the across-unit placebos outperform their respective synthetics and roughly half underperform—we can surmise that the average control unit performed as expected, or more precisely as its synthetic predicted it would, in the treatment period. Even if we cannot conclude that the treatment event did not affect the entire population of eligible control units, synthetic control may still be a valid approach; however, we have to be especially careful in our interpretation of the results in these cases because the total treatment effect estimated would reflect both: (i) the direct effect of the treatment event on the focal unit, and (ii) the indirect effect of the treatment event on the focal unit through spillovers on the population of control units. The presence of indirect effects is
plausible in any setting with interdependencies among units in the population; in strategic management contexts, these could include potential competitor reactions to the treatment event.

A final falsification question to ask is aimed at the other key assumption: can we safely assume the structural relationship \( W \) we estimate between included control units and the synthetic unit, as an intermediate step, remains relatively stable in the post-treatment period? Out-of-sample tests help address this concern. In these, an alternate synthetic is constructed where the range of pre-treatment data used to select weights on control units does not run up to the treatment date. Rather, the data used to construct the synthetic in out-of-sample tests is drawn from an earlier period, which is analogous to having training and evaluation periods in a regression-based forecasting setting. If the performance of out-of-sample synthetics mirrors that of a baseline synthetic then we can more credibly state that our results are unlikely to be driven by changes in underlying structural relationships.

2.3 Advantages

The synthetic control method has a number of advantages relative to existing techniques. In some cases, synthetic control empowers researchers to answer questions that existing methods cannot feasibly address. In other situations, where existing methods work, the synthetic control method can provide improvements to estimates by overcoming challenges that could bias results.

The primary advantage of synthetic control vis-à-vis traditional regression techniques is that the method is feasible when only one observation unit, or only a few units, within a (potentially limited) population receives the treatment. Most important managerial questions preclude experimental designs, limiting researchers to observational data that frequently contains few good direct counterfactuals, but more often none whatsoever (March et al. 2001; Runde & de Ronde’s 2010). In situations such as these, regression methods cannot be applied. Regression methods require variation in key variables across multiple observation units making estimation using those techniques infeasible in application to these types of questions. Nevertheless, in each of the situations above only one unit varies significantly from

---

8 For instance, in Abadie et al.’s 2011 analysis of the impact of the collapse of the Berlin Wall and reunification on West German GDP, a synthetic West Germany is constructed from five countries, Austria, Switzerland, the United States, the Netherlands, and Japan. Even though the treatment event, reunification, is applied directly only to West Germany, it is likely that untreated countries in the synthetic would have been indirectly affected by the event due to trade and FDI relationships with West Germany— particularly the units that receive the largest weights: Austria (0.47) and Switzerland (0.17), which may have suffered as East and West Germany integrated, since they share not only borders, but also languages with West Germany.
the others. Synthetic control provides a means to address the problem of only a single observation unit being treated.

Another advantage is that synthetic control can feasibly be applied to a variety of outcome variables. This stands in contrast to event study approaches, which are the only regression-based technique that can, in some circumstances, be applied to the analysis of a single treated unit. In event studies, the only possible outcome variable is abnormal financial market returns; in synthetic control the possibilities for outcome variables are only limited by the availability of data and the researchers’ imagination. This is particularly important in management research since scholars are interested in broader measures of performance than abnormal financial returns. Moreover, for some firm-level questions we may need indicators applicable to private firms as well as public ones. Furthermore, we may be interested in managerial outcomes at other levels of analysis. A final advantage relative to event studies is that, using synthetic control, we observe the entire evolution of outcomes after a treatment rather than being constrained by an event-window—enabling us to assess simultaneously short-run and longer-run implications for a particular observation unit.

In addition to providing feasible ways to answer new questions, synthetic control can be used to overcome biases in existing methods, including: researchers’ cognitive biases, omitted variable bias, and endogeneity. Case selection may be biased by researchers’ beliefs about what makes for good comparator groups (Eisenhardt 1989, Eisenhardt and Graebner 2007), even in multiple comparative case studies; this is not a problem in synthetic control because the matching of comparator units is left to an objective mathematical optimization process. Even fixed-effects approaches in regression techniques cannot solve the problem of time-varying unit-level omitted variables bias; however, Abadie et al. (2010) demonstrate that in synthetic control—if a synthetic unit’s outcome data tracks the focal unit’s well over a sufficiently long pre-treatment period—the omission of unobserved variables need not be a concern. Endogeneity biases are probably the biggest challenge facing researchers in any non-experimental design claiming

---

9 Cognitive biases may tempt qualitative researchers into selecting comparator units that are competitors, or as a whole appear similar, to the focal unit, while the synthetic control process disciplines us into choosing comparator units with attributes that predict an outcome measure instead.

10 The intuition is straightforward: only units that are similar in terms of observed and unobserved determinants should produce similar outcomes and trajectories over extended durations. Hence, once a synthetic is established that closely mimics the actual in outcome behavior over an extended pre-treatment period, any discrepancy that arises in the post-treatment period may reasonably be attributed to the treatment itself.
causality; fortunately, synthetic control methods alleviate these concerns both through a quasi-experimental approach (i.e. examining a treatment applied to some units but not others) and through accompanying tests that falsify non-random assignment concerns (Abadie et al. 2011).\footnote{Placebo tests can show whether or not random assignment of the treatment to units other than the focal unit or to the focal unit at different points in time lead to similarly large effects that would undermine our ability to make causal inferences. (Heckman and Hotz 1989)} Hence, synthetic control can also be used as a robustness check on other methods’ findings when researchers cannot alleviate concerns about these common biases.\footnote{In fact, revisiting old questions using the new method as it has fewer potential biases than prior methods, is the basis for Billmeier and Nannicini’s (2013) re-investigation of the effect of economic liberalization on GDP growth.}

### 2.4 Limitations

Synthetic control is not a panacea despite its many advantages—and should only be used when the question and the data are well suited to it.

External validity is an issue limiting researchers’ ability to extrapolate beyond the results—that is, we cannot directly interpret results beyond making simple comparisons between the focal unit and it’s synthetic. Moreover, the counterfactual the synthetic represents may not be the only interesting one. Unfortunately, the methodology does not allow us to estimate directly the behavior of counterfactuals in alternative what-if scenarios; we can only compare the reality of what happened to a single objectively constructed counterfactual representing the world as if the focal unit, and control units, persisted as going-concerns in the absence of everything that did happen. Consequently, synthetic control cannot be used to estimate the gap between the going-concern counterfactual and other scenarios that never occurred. This does not, however, preclude us from drawing conclusions about other scenarios that did not occur if we are willing to make equilibrium assumptions about underlying actors’ behavior. While providing a bridge between large sample statistics and case study methods, synthetic control differs from both. Since the goal is to quantify causal effects, capturing both the direction and magnitude of interventions, synthetic control may appear similar to large sample statistical approaches. In one sense, this comparison is fair because synthetic control shares the same limitations as large sample statistical techniques for theory building. Synthetic control itself does not provide insights into the mechanisms causing any change in the outcome variable (or lack thereof) following the treatment—which is a task to which other techniques like fuzzy set qualitative case analysis might be better suited (Fiss 2007). Hence, qualitative case studies,
which can describe in rich detail a range of mechanisms, are a natural complement for synthetic control analysis.

In a another sense, synthetic control is more similar to comparative case methods than large sample statistics, because an advantage of the technique is that it can be applied to narrow contexts in which a single statistical unit is treated. Nevertheless, it cannot always provide answers. In particular, the method will fail to create suitable synthetic matches for units that are outliers or have extreme (large or small) values on the outcome measure of interest; this is because synthetic control produces better counterfactual matches in the pre-treatment period when a larger subset of the population is a priori more similar to the focal firm. Hence, qualitative case studies may still be better suited when researchers are interested in extremes.

3 APPLICATION: GOVERNMENT INTERVENTION IN CHRYSLER

Synthetic control methodology allows researchers to analyze phenomena that occur in a limited population and/or that apply to only a small number of firms. We apply the method to assess the impact of government intervention in the auto industry in 2008 on Chrysler’s sales. Existing statistical techniques are inappropriate for addressing this scenario given the small number of auto firms and the even smaller number that accepted conditional government assistance. The estimates we generate for the impact of government intervention on Chrysler’s sales using synthetic control contribute to a broader academic debate within the management literature on the role of government in private industry (Mahoney et al. 2009; Kivleniece & Quelin 2012)—in which predictions about the impact of government intervention on firm performance range the spectrum from positive to negative.

We conduct a synthetic control analysis of this scenario in part to demonstrate the applicability of the method to a management question; we also hope that other management researchers will adopt the method and apply it to question in other contexts. We implement our analysis in STATA using the Synth Package—and are making code available to facilitate the learning and adoption process.14

---

13 This is because: (i) the weighting process assigns weights—that sum to the value of one—to comparator units; and (ii) there typically is a reason why the extreme units, in fact, may have a different data generating process than the other units. By construction, it is not possible to mimic the largest or smallest unit within a population, limiting the technique to more moderate or less extreme units.

14 This will be available for download on author websites, along with a version of data we used, upon publication.
3.1 Background

Beginning with the New York Federal Reserve Bank’s emergency loans to Bear Stearns in March 2008, capital markets in the U.S. began to tighten, leaving many firms across sectors in unexpectedly vulnerable positions. Auto manufacturers globally were hurt by this tightening as consumer vehicle purchases are particularly sensitive to the availability of cheap personal credit; moreover, the manufacturers themselves relied heavily on revolving credit to keep operations running. By early September, all three Detroit-based auto manufacturers had requested the United States Congress provide them with temporary financial relief. Aiming to alleviate burgeoning damage to the economy, Congress overcame partisan tensions to enact the Troubled Asset Relief Program (TARP) in October 2008, with the express purpose of injecting funds into the flailing financial sector. Despite the passage of TARP, the Senate and the House could not overcome differences on the terms of an auto industry relief package.

Seeing no imminent legislative agreement, President Bush sidestepped Congress on December 19, 2008; he used his executive authority to order the Treasury Department to extend an initial $13.4B of TARP funds to Chrysler and General Motors, explicitly spelling out some conditions for the assistance, while leaving other conditions to the discretion of Treasury officials who would provide oversight and be responsible for future negotiations and future injections of capital. President Bush’s decision to extend the use of TARP funds beyond the financial sector was controversial and some claimed illegal (Sullum et al. 2009). Nevertheless, he argued that extending loans to auto manufacturers was a suitable use of TARP because it would be irresponsible to allow iconic American firms to go bankrupt and have their constituent parts sold off if that outcome would lead to the loss of American jobs (Office of the Press Secretary 2008).

Despite having initially requested relief from Congress, Ford deemed unacceptable the structure of the ultimate agreement that Chrysler and General Motors made with the Executive Branch. Ford cited overly restrictive conditions and the potential for additional restrictions as reasons for their concern. Ford also worried about consumer perceptions of taking a “bailout” (Dolan 2009).

One initial condition of the assistance program gave the Treasury “the power to block any large transactions” (Office of the Press Secretary 2008). Later negotiations between the recipient firms and the
Treasury would, among other things, lead to specific limits on executive compensation and force both companies through a managed bankruptcy process to facilitate restructuring in a way such that both the U.S. government and a United Auto Worker (UAW) union trust would take equity stakes in the firm, as would Italian auto manufacturer Fiat. While these later conditions were part of the government assistance program, they were unknown to Chrysler at the time the firm accepted the assistance because they did not know exactly how the incoming Obama administration’s Treasury would proceed.

Chrysler and the government would eventually come to an agreement to end the government assistance program, six year earlier than expected, in May 2011 (Treasury Press Center 2011). At that point, the government agreed to sell its remaining 6 percent stake in the firm to Fiat, effectively relinquishing not only its visible investment in the firm, but also its control.

How General Motors and Chrysler would have performed without being under the aegis of the conditional government assistance program is an open empirical question. Would the firms have done better, worse, or about the same without government intervention and oversight? As discussed above, synthetic control offers a means to estimate what the counterfactual performance of Chrysler, or its surviving components, would have been had the firm not accepted the assistance package President Bush offered and President Obama administered.

### 3.2 Data

The population of firms that form the basis for constructing a synthetic Chrysler consists of the 19 major domestic and foreign commercial automotive firms selling into the United States from January 2001 to December 2012. The primary data source is Ward’s Auto, which is an industry-specific commercial data provider that collects monthly data typically used by automakers, dealers, parts suppliers, and the financial community. It provides comprehensive coverage of all automotive firms over the panel, including items such as sales, technical specifications, production capabilities, retail pricing, brands, and fleet composition. We supplement the automotive data with corporate level information from Compustat. Since Chrysler was a privately held company over a period of the sample we consulted SEC filings to supplement the Compustat information for the firm’s leverage and employees data.
firm that closely matches the actual firm on most dimensions—ultimately giving us confidence in the quality of the match.

### 3.2.1 Outcome Variable

We focus on U.S. monthly light vehicle sales volume by firm (i.e. the total number of new cars, light trucks, and vans sold). Sales are a key top-line performance measure for the automotive industry, commonly quoted in the media and closely scrutinized by policymakers and financial analysts alike. The monthly frequency of its release makes it a good candidate for study as it provides immediate insight into the impact of the assistance program on Chrysler. Nevertheless, monthly sales data are quite volatile (as seen in Figure 1), given periodic corporate promotions and seasonal trends in consumer behavior. A final reason we use sales as a performance measure (and one reason we use synthetic control methodology), is that Chrysler was a private company at the time it accepted government assistance, precluding the use of financial event study methods.

### 3.2.2 Predictor Variables

We build from *Wards* 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 dual minimization process it will assign low weights to predictor variables in construction of a synthetic if they have little explanatory power, conditional on unit weights ($W$).

The first set of independent variables captures, on a monthly basis, vehicle-specific factors that could 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 capture strategic and operational factors that distinguish firms, 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...

---

16 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 primary findings.
are from SUVs, light trucks, and vans; and, the fraction of sales of imported vehicles. We are able to calculate all of the above variables from *Wards* on a monthly basis. Further firm features we incorporate into our set of predictor variables are available on an annual basis only in *Compustat*; these are the total number of employees and leverage ratio (total debt/equity). Finally, we construct as additional predictors of performance measures of past performance; these include a twelve-month moving average of sales volume and the level of that moving average indexed to 100 at the time of the government loans. Appendix A provides data sources and details of the data’s construction.

### 3.3 Analysis

Other than selecting outcome variables and predictor variables, the remaining decisions researchers must make in conducting synthetic control are choosing (i) the intervention date and (ii) the length of the pre-intervention window over which to minimize prediction error.

President Bush authorized the disbursement of government funds to the auto sector in late December 2008 and President Obama began regularly monitoring their usage in early February 2009. Hence, we use the mid-point of these two dates, January 2009, as the intervention or treatment date.\(^\text{17}\)

Choosing an appropriate pre-treatment window requires selecting a sufficiently long period over which to minimize prediction errors. We found that the synthetic Chrysler’s behavior closely matched that of the actual Chrysler when using a 48 month window—and that such a window minimized the root mean squared prediction error (RMPSE) in the pre-intervention period to create the best match between the synthetic firm and the actual firm before the government intervention.\(^\text{18}\)

#### 3.3.1 Core Results

Table 1 provides summary statistics of the attributes of the synthetic Chrysler we construct along with actual data on Chrysler and Ford, all in the pre-intervention period. It shows that the synthetic Chrysler compares well to the actual Chrysler in the period prior to the firm’s acceptance of government assistance. It further illustrates that the synthetic Chrysler is a closer match to the actual Chrysler in this

---

\(^{17}\) We also tried other months around January 2009 as a robustness check and find little substantive difference in the results.

\(^{18}\) We experimented with other windows over which to minimize RMSPE and settled on 48 months as appropriate because (i) longer windows did little to improve the fit and in many cases led to marginally higher RMSPE and (ii) shorter windows always led to a worse fit, such that there is a 5% improvement in the RMSPE moving to 48 months from 36 months. Our results are nevertheless robust to different window lengths where the RMSPE is not necessarily minimized.
period than is Ford as a comparative case. For ten of fifteen of the observable attributes, i.e. in the
majority of cases, the synthetic Chrysler is a closer match to the actual Chrysler than Ford.

For those attributes in Table 1 where the distance between the attributes’ value in the synthetic
Chrysler and the actual Chrysler is relatively high, the predictive power of those attributes tends to be
lower—as calculated in the attribute weight matrix, \( v_m \).\(^{19}\) Note that in synthetic control analyses we do
not need to impose assumptions on which predictor variables might matter most, as sometimes must be
done in comparative case analysis, because synthetic control follows an objective, data-driven process.

<Insert Table 1 Here>

Table 2 shows which other auto manufacturing firms are potential control units that could
comprise the synthetic Chrysler. It also shows what weights (\( W \)) these control companies receive in
construction of the synthetic Chrysler. Recall that these weights are determined by how well they
replicate Chrysler’s actual performance in the pre-government intervention period to create a match.
When these weights are applied to post-intervention data, they generate counterfactual performance as
embodied in the synthetic Chrysler that did not accept government assistance.

Of the eighteen auto manufacturers other than Chrysler that are potential matches in a synthetic,
two must be excluded as matches in the synthetic Chrysler for technical reasons. General Motors also
accepted government assistance when Chrysler did, meaning it was subject to the same treatment at the
same time, implying that its performance after the intervention at Chrysler would not be a valid
counterfactual. Jaguar Land Rover does not have leverage or employee data, attributes on which we
created matches; hence, it must be excluded as well.\(^{20}\)

Among the remaining sixteen firms which could comprise the synthetic, five receive positive
weightings and the remaining eleven receive zero weight. It is typical in applications of synthetic control
that a substantial number of potential control units, or other auto manufacturers in this case, receive zero

\(^{19}\) 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. The \( v_m \) matrix is constructed so that it minimizes RMSPE conditional on unit weightings \( W \), creating the best possible match between the synthetic and actual firms in the pre-treatment period—and hence it does not represent the attributes with the highest global predictive power, but rather the highest predictive power conditional on \( W \).

\(^{20}\) Due to its unique organizational structure Jaguar-Land Rover did not report separate balance sheet or operational statements. This made it infeasible to include the leverage ratio or the number of employees for this firm which prevented us from including it as a potential control unit.
weights because they do not make good individual matches on the outcome variable and none of their other attributes are sufficiently similar to the focal unit’s.

Of the five firms that receive positive weightings—Daimler, Ford, Isuzu, Nissan, and Toyota—intuition helps reconcile the weights the companies receive. Most auto industry observers would select Ford as the closest match to Chrysler had General Motors been ruled out as an option. Moreover, Ford had initially requested help from the government but later distanced itself from that proposal over concerns regarding the imposed structure and oversight of the loan conditions. These facts make it unsurprising that Ford receives the largest weight in the synthetic at 0.664.

If industry observers had to go beyond the Big Three, many would select Toyota and Nissan as the next closest matches to Chrysler given the firms’ diverse product offerings, multiple brands and production platforms, vast 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. This again makes it unsurprising that Toyota and Nissan 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 include in the synthetic Chrysler, but understanding why helps illustrate an advantage of the programmatic 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 extremely heavily weighted towards SUVs, light trucks, and vans—which were an important part of Chrysler’s pre-government assistance 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 manufacturer heavier vehicles.

Daimler enters the mix with a miniscule 0.022 weighting—rounding out the large passenger car end of the market in the synthetic Chrysler mix.

<Insert Table 2 Here>

Using the weights on the companies in Table 2 we generate the synthetic firm’s performance and compare it to Chrysler’s actual performance in Figure 2. The figure illustrates how the synthetic Chrysler would have performed relative to the actual firm—from the beginning of the pre-treatment period in
January 2005 through December 2012, giving us 48 months to observe the effects of government intervention on Chrysler’s performance.

Prior to Chrysler’s acceptance of government funds in January 2009, the performance of the synthetic Chrysler and of the actual company track each other reliably with only short periods where 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 also a downward trend in both Chrysler’s and the synthetic’s performance which was indicative of the entire auto industry in the several years prior to the intervention, as the U.S. economy began to slow starting in 2005 and as fuel prices began a rapid ascent in 2007. Given that the series track each other well despite the turmoil in the industry, the synthetic achieves a good match for Chrysler’s actual performance.

What is most striking in Figure 2, however, is that during the government intervention period Chrysler significantly underperforms its synthetic counterfactual, representing performance of the firm had the government not intervened in the industry and had it not accepted an offer of conditional assistance. This suggests that Chrysler would 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 could conclude that government intervention increased subsequent sales.

We are interested in understanding not just the direction of the effect of Chrysler’s acceptance of government assistance, but also the magnitude of the effect. Figure 3 shows the size of the gap between the actual firm’s sales and the synthetic firm’s sales. At the gap’s widest, Chrysler sold 59,000 fewer vehicles in that month than the counterfactual synthetic suggests it would have otherwise. Given the level of Chrysler’s actual sales during the intervention period, Chrysler appears to have sold 29% fewer vehicles than its synthetic counterfactual during the period from its acceptance of government assistance through its repurchase of government equity stakes in the firm.

The underperformance of the Chrysler 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. Chrysler’s improved performance upon exiting the government program helps us pinpoint causal mechanisms as the performance reversal implies that the harmfulness of such factors’ should fade as active government involvement draws to an end. This leads to further insights about the dynamic between private firms and public institutions more broadly.

<Insert Figure 2 Here>

<Insert Figure 3 Here>

3.3.2 Robustness and Strength of Inference

Robustness checks are particularly important to conduct here since we are making causal inferences based on a small \( n \) dataset. We run placebo tests to check whether or not as large a result would have manifested itself had the intervention occurred at a different point in time or among untreated units. We then construct pseudo p-values to quantify the strength of our inference, leveraging the forced random assignment of the treatment across-units from the placebo tests. Finally, we attempt to falsify two critical underlying assumptions. Our core result survives each of these checks on the inference.

Placebo Tests in Time

One way to check for robustness of results is to introduce a false intervention at a different point in time. In this case, there should be no effect of the false timing of Chrysler accepting government assistance if the timing of the effect is causal.\(^{21}\)

In Figure 4 we treat the intervention as if it had occurred 4 years (or 48 months) earlier in time, when we know it did not occur then in reality.\(^{22}\) To implement this placebo in-time test, we simply re-run our analysis on Chrysler using January 2005 as the intervention date instead of January 2009; we continue to use a 48 month window prior to the intervention date on which to minimize prediction error.

Figure 4 shows 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

\(^{21}\) Abadie, et al. (2011) conduct in-time placebo tests with respect to the date in which Germany was re-unified in an analysis of its effects on West German GDP per capita, showing that there was no divergence around “hypothetical” re-unifications a decade or two earlier.

\(^{22}\) We choose to implement our placebo-in time test 4 years prior to the actual intervention because (i) our main synthetic is constructed using data over a 4 year period prior to the intervention, and (ii) the ideal in-time placebo tests use data from a pre-treatment period that does not overlap with the range of data used to construct the main synthetic (4 years in our case) as this ensures new, rather than duplicate, information is used to generate the placebo in-time synthetic.
actual firm quite closely in the post placebo intervention period. The absence of real divergence improves our confidence in our core result: Chrysler’s acceptance of government assistance in January 2009 caused the firm’s sales to deteriorate by more than they would have if the firm had not accepted it at that date.23

<Insert Figure 4 Here>

Placebo Tests among Untreated Statistical Units

Another way to conduct placebo tests is to run them among untreated statistical units. The purpose is to check, if the intervention had been falsely applied to each of the control units rather than to the focal unit, whether an unexpectedly large negative reaction occurs at any of the control units. If the effect of the treatment on the focal unit is causal in the case of Chrysler, then we would not expect an application of the treatment to untreated units to lead to an equally large negative divergence (or treatment effect).24

Figure 5 illustrates the gap in performance between the actual and the synthetic versions of each of the sixteen possible control firms. In it, synthetics for each control firm are created as if they had received an intervention in January 2009 just as Chrysler did in reality. We continue to use a 48 month pre-treatment window over which to minimize pre-treatment prediction error.25 The solid line represents the gap in performance between the actual and synthetic Chrysler as depicted in Figure 3. Figure 5 demonstrates that the gap in Chrysler’s performance relative to its synthetic’s during the period of active government intervention in the firm from January 2009 through May 2011 is far greater than the gap for any of the control units to which a placebo intervention was applied. As early as June 2011, however, other firms’ synthetics perform relatively worse than Chrysler’s, suggesting that Chrysler recovered some of its lost sales following the government’s agreement with the firm to exit early. This further bolsters our causal interpretation of the effect of Chrysler accepting and retaining conditional government financial assistance as having a more significant detriment on Chrysler’s sales than at potential control firms, which received similar hypothetical treatments.

23 We get a very similar result (with no divergence) when we have a placebo intervention in January 2004 which exhausts our dataset as our panel only goes back to January 2000. We also find no immediate effect of placebo interventions in January 2006, January 2007, or January 2008, though the tests are less than ideal given the data overlaps with our baseline analysis.

24 Abadie et al. (2010) do this with respect to the state in which anti-tobacco legislation was passed in an analysis of the impact of a California law’s effects on cigarette consumption by hypothetically assuming every other state passed a similar law at the same time.

25 Note that Chrysler is not in the pool of potential control firms for constructing placebos across-units.
In Figure 6, we present the results of the placebo test, among untreated firms, excluding synthetics for control firms with mean-squared prediction errors (MSPE) in the pre-intervention period greater than 5 times that of Chrysler—because several control firms in Figure 5 have a poor synthetic fit.26 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.27

Assessing the Strength of the Inference

The data generated in running the placebo-among-units test also provides a basis for evaluating the strength of inferences quantitatively via constructing pseudo p-values based on randomization or exact inference.28 In a traditional regression setting, we assess the strength of inferences with p-values based on frequentist inference that capture the fraction of observations that are at least as extreme as a specified data point under what is assumed to be a normal distribution of all possible repeat observations. In a synthetic control setting, we can generate a similar statistic computationally, appealing instead to randomization inference principals (Ernst 2004): pseudo 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 were randomly assigned to any observation unit in the population.

To evaluate the relative extremity of the treatment event on each observation unit—which is the first step in calculating a pseudo p-value—we first calculate a scale-independent measure reflecting treatment extremity so we can compare observation units directly with each other. We use the ratio of the root mean square prediction error (RMSPE) for the treatment period to the RMSPE for the pre-treatment period. We construct this measure for each firm in the population evaluated in the placebo-among-firms

---

26 Poor fits in the pre-prediction period for some statistical units can be expected in synthetic control analysis, particularly for observation units at the extremes (Abadie et al. 2010).

27 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.

28 Abadie et al. (2010) are the first to calculate such a metric but do not refer to it as a p-value until Abadie et al. (2011). They note the technique is based on Rosenbaum’s (2002a, 2002b) methods for permutation inference in randomized experiments or observational studies.
test above. Because positive extremes may be interpreted very differently from negative extremes, we want this extremity measure to take into account whether the underlying treatment effect was positive or negative on average; multiplying the ratios of treatment period to pre-treatment period RMSPEs by -1 for observation units where the mean treatment effect is negative brings this directional information back into the measure. We can interpret values of this measure close to 1 or -1 as indicating there are no major changes in how well the synthetic matches the actual unit in the treatment period relative to the pre-treatment period.

In Figure 7, we display a dot plot of our treatment extremity measure for the 17 firms in our population, 16 of which are control firms and 1 of which is the focal firm, Chrysler. The majority of the observations are clustered towards the center of the distribution, falling in the range [-2, 2] indicating that the government assistance program, or other coincident events, had negligible effects on most auto manufacturers.

The x-axis of the dot plot in Figure 7 is bordered by a histogram representing the empirical probability distribution of the extremity measure. In traditional regression settings, frequentist inferences are made based on p-values calculated based on assumptions about where a given coefficient falls in what is assumed to be a continuous, normal distribution; in synthetic control (and other permutation inference) settings we do not need to make asymptotic assumptions about continuous distributions as we can make exact inferences about where a given outcome falls in an empirical distribution of all possible permutations of the treatment as our dataset includes the entire, albeit small, sample population. 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 (represented by the histogram on the axis border), we can readily calculate our pseudo p-value as $\frac{2}{17} \approx 0.12$ by taking the number of observations as extreme as what we observe and dividing by the number of observations in the population—which is analogous to integrating the assumed normal

---

29 In principle, we could also calculate treatment extremity values using multiple in-time placebo tests and construct analogous pseudo p-values to determine if the timing of the divergence was significantly different in one period than in another period. We are unlikely to gain much insight from this since in our case it is clear when the actual government intervention started. It is fairly clear from comparing the in-time placebo tests we ran (in Figure 4) to the actual event (in Figure 2) that there is a larger divergence between the synthetic and actual Chrysler around the actual event. Nevertheless, constructing treatment extremity measures based on different in-time placebo tests could be useful if it was unclear when a treatment event, or other structural break, occurred.
distribution in a regression setting between negative infinity and the observed value of the coefficient to generate a p-value.\textsuperscript{30} Hence, the probability of seeing as large a negative effect of accepting and maintaining government assistance on sales as we saw at Chrysler is approximately 12% if instead the government assistance were randomly assigned to any firm in the population. While in traditional regression settings many researchers would reject null hypotheses when p-values are greater than a specified level (typically p>0.05), we need to be particularly careful in synthetic control settings to calibrate our interpretation of p-values (Sellke et al. 2001). The negative effect of government assistance on Chrysler’s sales can still be interpreted as being measurably greater than zero in this instance with a pseudo p-value of 0.12.\textsuperscript{31}

Another advantage of constructing the empirical distribution of treatment extremity values is that it allows us to evaluate whether or not a treatment event had spillover effects, which potentially altered the status quo post-treatment outcomes of all potential control units. In our case, half of the control firms underperformed their own synthetics and half of the control firms outperformed their own synthetics—that is, the average control firm performed as its synthetic predicted it would have. This suggests that if there were any spillover effects from the intervention in Chrysler on other auto firms’ sales, then the effects did not systematically benefit or harm the average firm.\textsuperscript{32} Moreover, the control firms included in the baseline synthetic Chrysler are clustered tightly above-and-below the center of the treatment extremity distribution—suggesting that our core results are unlikely to be significantly biased as a result of spillovers.

---

\textsuperscript{30} To gain further insights into when the government intervention had the strongest effects on Chrysler’s sales we considered alternative windows over which the treatment might have had a stronger/weaker effect—as we thought this might better help us understand the underlying mechanisms at play were. We found that Chrysler had the most negative treatment extremity measure among all units when we calculated the treatment extremity measure as if the treatment event occurred only in the period from January 2009 to January 2011 rather than January 2009 to May 2011—giving us a pseudo p-value of 0.06. This is also true for even shorter intervals. Hence, the effects of the government intervention were most acute in the immediate two years following the intervention. When we lengthened the window over which the treatment effect may have occurred to include all post-intervention dates with available data (i.e from January 2009 through December 2012), Chrysler had only the third most negative treatment extremity value, equating to a pseudo p-value of 0.18. Hence, the effect of the government intervention on Chrysler’s sales appears to fade after the agreement to end the government intervention was announced.

\textsuperscript{31} We need to be particularly careful not to make Type II errors by falsely rejecting real treatment effects by calibrating our interpretation of p-values in this synthetic control setting. To illustrate why we need to avoid the p-value fallacy here, consider the hypothetical case where Chrysler had the most negatively extreme relative response to the government assistance applied in January 2009; its p-value would still only be 1/17 ≈ 0.06 in that scenario. If we followed a p>0.05 rule we would reject the null hypothesis and declare the treatment effect to be negligible despite it being the single most extreme post-treatment realization. Moreover, in our data the most-negatively extreme observation unit in the post-January 2009 period was BMW—a firm whose underperformance in this timeframe has been attributed to factors other than government assistance, namely the competitive resurgence of Volkswagen-Audi who took market share from BMW in the luxury segment (Libby 2012). Unsurprisingly, Volkswagen-Audi occupies one of the most extremely positive positions in Figure 7. This could be interpreted as being consistent with policy justifications for the government intervention centered on the sustained viability of the average firm in the industry, including parts suppliers shared by multiple manufacturers, rather than the welfare of recipient firms.
Leave-one-out Tests

While the distribution of the treatment extremity measure examined above suggests our results should not be overly-sensitive to the inclusion of specific control units in the synthetic, leave-one-out tests explicitly check for this possibility. 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 a large, unrelated shock in the post-treatment period. An implicit assumption in conducting synthetic control is that any shocks to outcomes in the post-treatment period are either (i) common to all units including the treated or (ii) minor and offsetting among control units. Leave-one-out tests examine whether or not these assumptions hold.

To perform the leave-one-out test, we construct five alternative versions of the synthetic Chrysler where we exclude each of the firms that comprises the baseline synthetic Chrysler (shown in Figures 2 and 3) on a one-by-one basis. Hence, each of the five alternative synthetic Chryslers represents the synthetic Chrysler constructed from the population of auto manufacturers excluding either Daimler, Ford, Isuzu, Nissan, or Toyota. Figure 8 presents a graph analogous to that shown in Figure 2 displaying the baseline synthetic and the actual Chrysler, but adds five thinner dashed lines representing each of the leave-one-out synthetic Chryslers as well. Figure 9 presents a graph analogous to that shown in Figure 3 displaying the difference between the baseline synthetic and the actual Chrysler, but adds five thinner dashed lines representing the difference between each of the alternative synthetic Chryslers and the actual Chrysler.

Given how closely each of the leave-one-out synthetics and the baseline synthetic track each other on average, it can be difficult to see that there are five lines because there are only minor, and rarely sustained, differences between any single leave-one-out synthetic and the baseline synthetic. The average sales gap between these alternative synthetic Chryslers and the actual firm in 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% reduction in sales we estimated without leaving out any control firms. Hence, any bias introduced by the inclusion of a particular control firm in the
construction of the synthetic Chrysler is likely minimal and of indeterminate sign—changing our estimate that Chrysler underperforms its synthetic in periods of active government intervention by less than 1% in either direction.

<Insert Figure 8 Here>
<Insert Figure 9 Here>

*Out-of-Sample Tests*

A final robustness check aimed at falsifying underlying assumptions includes out-of-sample validation (Abadie et al. 2011). Out-of-sample validation takes data from well before the treatment is applied, leaving a gap before the event, such that there is a training and validation period as in regression-based forecasting, and uses this data to construct a synthetic; this helps alleviate any concerns about the period immediately preceding the event being abnormal and about whether or not there are omitted, time-varying concomitant-variables. More importantly, it helps assess whether or not the weights on control units we estimated using pre-treatment period data are reasonably stable over time. We construct an out-of-sample test using a four year training and four year validation period, such that the weights on control units in the synthetic are based on data from 2001-2004—rather than data from 2005-2008 as in our baseline synthetic—and present the results in Figures 10 and 11. Overall, we observe patterns similar to those in Figures 2 and 3, suggesting that indeed the weightings on control units included in the synthetic are relatively stable over time and that our substantive results are robust to minor changes in the weightings, alleviating concerns about earlier structural breaks altering vehicle sales patterns within the industry.

<Insert Figure 10 Here>
<Insert Figure 11 Here>

4 MECHANISMS EXPLAINING CHRYSLER’S UNDERPERFORMANCE IN THE INTERVENTION PERIOD

While synthetic control can be used to estimate, with relative precision, the magnitude and direction of a treatment effect, it does not directly reveal underlying causal mechanisms. Nevertheless,

---

33 We might worry about either of these concerns if the synthetic constructed using out-of-sample data did not approximate the synthetic using in-sample data; however, it would not necessarily obviate our findings if the reason for the break could be identified and was no longer relevant.
we can build qualitative arguments around the actual circumstances at Chrysler to support existing theories about potential mechanisms. Potential factors explaining why Chrysler sold fewer vehicles during the period in which it maintained active government assistance include: (i) consumer backlash, (ii) corporate governance and executive compensation challenges, (iii) substitution of commercial for political objectives, and (iv) competitor reactions to government involvement.\textsuperscript{34} To the extent that we can demonstrate qualitatively that these mechanisms had dampening effects on Chrysler’s sales when it took government assistance in January 2009—and that those effects lessened or disappeared altogether when active government assistance ended in May 2011—they corroborate our quantitative application of synthetic control.

4.1 Consumer Backlash

The social movement literature predicts resistance to grow against firms when issues they are associated with become salient and sets of actors can unite around shared beliefs (Davis & Thompson 1994). That prediction is consistent with one argument for Chrysler’s underperformance: that consumer backlash against the firm for taking a tax-payer funded “bailout” weighed on sales despite those who believed government backing would assuage consumer’s fears about warranty and maintenance issues (Stewart 2012). Indeed, independent market research found that 22% of likely purchasers planned to avoid Chrysler products because of the firm’s decision to accept government assistance (Vlasic 2012).\textsuperscript{35} Moreover, this mechanism is likely to have moderated following public announcements surrounding the government’s exit from its assistance program at Chrysler, which diminished the salience of the firm being “bailed out”— potentially allowing for sales to return to baseline levels.

4.2 Executive Compensation Challenges

Corporate governance scholars predict that firm performance is a function of executive compensation and related incentive structures (Jensen & Murphy 1990). Hence, any exogenous restrictions imposed on executive compensation could significantly impact performance if attracting,\textsuperscript{34}
\textsuperscript{35}

\textsuperscript{34} We do not find any evidence that Chrysler increased absolute or relative vehicle prices, which could be a potential explanation for the estimated sales volume shortfall. Between 2009 and 2011, Chrysler reduced its MSRP\textsuperscript{s} on average by 3.3% while Ford increased its by 12.0% (authors’ calculations, available on request, based on MSRP data obtained from Wards).

\textsuperscript{35} This finding comes from a survey question included on a broader, ongoing automotive industry market research questionnaire administered in the first quarter of 2012 by CNW Marketing Research of Brandon, Oregon.
retaining, and incenting top managerial talent is critical (Pfeffer 1994). These mechanisms explaining firm performance are consistent with an argument that Chrysler underperforms because government conditions on executive compensation were overly restrictive, making it difficult for the firm to remain competitive in the market for executive talent by tying compensation to vehicle sales. On October 10, 2009, the Treasury’s Special Pay Master announced rules capping cash salaries and severely limiting bonuses for most executives at the seven biggest recipients of TARP funds including Chrysler (Treasury Press Center 2010). Within five months there was over 15% turnover among top management at these firms (Treasury Press Center 2010). Once Chrysler exited TARP in May 2011, however, it was free to compensate its executives at market rates. By fiscal 2012, the board authorized sweeping increases in senior executive pay, on the order of 256% as compared to 2010.\footnote{36 This calculation is based on the compensation data listed in Chrysler’s 2012 10-K filing; it reflects the increase in the average compensation of ‘Named Executive Officers’.}

4.3 Substitution of Commercial for Political Objectives

Property rights theorists have argued that firms in which governments maintain ownership stakes should perform less efficiently than purely privately owned firms (Boardman & Vining 1989; Dewenter & Malatesta 2001; Megginson & Netter 2001). The substitution of commercial for political objectives can result in outcomes that compromise economic efficiency. Allowing Treasury oversight as a condition of accepting TARP funds permitted the U.S. government to directly influence strategic decision-making at Chrysler. Insider accounts have revealed a number of contentious issues where the Presidential Task Force on the Auto Industry prevailed against the preferred approach of Chrysler management (Barofsky 2011; Rattner 2011).

One such issue concerned the pace of auto dealership rationalization, which Chrysler management initially proposed should be a gradual process over five years, ending in 2014. Instead, the Treasury required the firm to drastically accelerate this process, leading to the abrupt 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.
The CEO of General Motors, which also had to cut many dealerships, commented in retrospect that the cuts were “not necessary” (SIGTARP 2010).

The government also required Chrysler to make significant concessions to the auto unions, notably over pensions and wages—consistent with the head of the Auto Task Force’s view that the government intervention “led to a somewhat kinder outcome than a purely private restructuring would have” (Rattner 2011). 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 that outcome 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 67.69% 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 & Warburton 2010; Roe & 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 (Rattner 2011).

The government’s exit in May 2011 removed the need for Chrysler management to compromise commercial objectives for political ones. Chrysler could now openly challenge the government, which it did in June 2013 when it publicly objected to the government’s request that the automaker recall 2.7 million Jeep vehicles that the National Highway Traffic Safety Administration (NHTSA) claimed were “defective and prone to fires in the event of rear-impact collisions” (Vlasic 2013a). Taking such a position would have been quite precarious under government ownership. Despite Chrysler’s newfound ability to challenge the government, some of the effects of prior disagreements between management and

---

37 Similar arguments for discharging pension liabilities had been made successfully in prior bankruptcy proceedings by United Airlines in 2005 and Bethlehem Steel in 2002.

38 Chrysler eventually backed off its hardline position and agreed that it would voluntarily install trailer hitches on some older model Jeeps, so long as regulators stopped calling the vehicles “defective” (Vlasic 2013b). Nevertheless, it is still difficult to imagine this outcome rather than quick acquiescence to NHTSA demands had the government retained an ownership stake.
their political overseers may have had longstanding effects on operations; for example, closed dealerships were unlikely to be reopened in the short-run and union wage floors were unlikely to fall to the levels Chrysler management initially wanted (especially since the UAW now held a large equity stake in the firm). These factors could have had persistent deleterious effects even after the government’s exit, explaining why Chrysler did not return fully to the level its synthetic suggests it should have.

4.4 Competitor Reactions to Government-backed Firms

Seamans (2012) finds that the emergence of government-owned, or financed competitors, can alter the competitive landscape by causing other firms to compete more vigorously. This may partially help explain Chrysler’s sales shortfall following government intervention. Competitor reactions may account for why some firms that did not accept government assistance or were not offered government assistance marginally outperformed their synthetics during the government intervention period. Ford responded to government support of Chrysler and GM by focusing its own vehicle redesign efforts specifically on segments in which its government-assisted rivals had recently gained market share (Terlep & Ramsey 2011). Moreover, Ford ran a marketing campaign positioning themselves as the only American firm that did not take government assistance. Despite the theoretical plausibility of this mechanism, competitor reactions are unlikely to be the primary driver of our results given that our baseline estimate changes by only one percentage point in the leave-one-out test excluding Ford.

5 Discussion of Chrysler Application

Our analysis provides the first quantitative estimate of how Chrysler would have performed had it declined conditional government assistance in January 2009: specifically, we estimate that the firm would have sold approximately 29% more vehicles during the period in which it actively received government assistance when compared to a synthetic counterfactual that survived without such assistance. The analysis holds up to a battery of robustness checks. Moreover, qualitative explorations of several theoretically grounded mechanisms corroborate the quantitative evidence generated in our synthetic control analysis.

39 In one of Ford’s “Drive One” television advertisements featuring real customers who are put on the spot about their decision to buy from the company, a Ford F-150 owner named Chris explains “I wasn't going to buy another car that was bailed out by our government. I was going to buy from a manufacturer that's standing on their own: win, lose, or draw. That's what America is about is taking the chance to succeed and understanding when you fail that you gotta' pick yourself up and go back to work. Ford is that company for me.” (Bedard 2011)
Despite our quantitative and qualitative efforts, several important questions remain unanswered. Is it realistic to believe that the counterfactual Chrysler that survived without government assistance could have existed, and, if not, can we still learn from it? Can we say anything about the effect of government assistance on General Motors’ sales? And, can we say whether it was a good idea from the government’s perspective to extend the assistance packages in the first place?

5.1 Assessing the Plausibility and Relevance of the Synthetic Counterfactual

Given the realities in late 2008, the synthetic embodies a plausible scenario for what could have happened had Chrysler refused conditional government assistance. The U.S. Treasury had on October 14, 2008 forced the largest financial institutions to take a total of $250 billion to use explicitly to make new investments and loans rather than to write off “toxic” assets (Lander & Dash 2008). This suggests that private capital should have been available to Chrysler (at some interest rate) given that the amount of funds the government had just injected into major financial institutions dwarfed the amount Chrysler took from the government. Moreover, given Chrysler’s stock of firm-specific property, plant, and equipment (Klein 1988), the firm likely would have been worth more to existing creditors when it was producing cars than not producing them (Baird & Morrison 2001; Jackson & Scott 1989)—suggesting that if a bankruptcy had occurred that restructuring while continuing to operate (i.e. pursuing a Chapter 11 bankruptcy) would have been preferable to liquidation (i.e. pursuing a Chapter 7 bankruptcy). Taken together, this argumentation suggests that it is plausible that Chrysler could have survived without government assistance by making do with private sector funds while restructuring under Chapter 11. In fact, we can interpret the synthetic counterfactual as if the firm had pursued a Chapter 11 bankruptcy in a going-concern scenario and liquidated one particularly challenged division, the light trucks/vans/SUVs segment—since in the post-treatment period one control unit that comprises part of the synthetic Chrysler, Isuzu, did shutdown their North American operations in that segment.

The counterfactual Chrysler we estimate using synthetic control assumes that Chrysler was able to continue as a going-concern without accepting government assistance. We have focused on this counterfactual scenario so far, not because it is the only interesting one, but because it is the scenario that synthetic control allows us to evaluate most directly. Despite the plausibility of the going-concern
scenario, some might argue that other potential counterfactual scenarios are worthy of examinations, since embedded in the going-concern scenario are questions about: (i) whether private sector financing was available; (ii) whether the firm would have liquidated entirely, (iii) whether the firm would have liquidated some constituent parts to stay alive; and, (iv) whether Chrysler would have won the concessions it wanted had it pursued an unmanaged Chapter 11 bankruptcy. Different combinations of answers to these questions represent an interesting set of alternative counterfactual scenarios worthy of direct consideration in their own right; however, we are unable to estimate counterfactual performance for each of these alternative scenarios given methodological limitations.

Despite synthetic control’s inability to help us directly evaluate counterfactual scenarios other than the plausible going-concern one—we can use equilibrium logic to make some indirect inferences about Chrysler management’s assessment of the options they faced. If Chrysler management believed they were making the right probabilistic decision with the information they had in December 2008, then they must have made an assessment that the conditional government assistance option was the lesser of several evils—otherwise they would not have self-selected into a scenario that implied a 29% shortfall in sales. Taking this equilibrium logic further, it is likely that Chrysler management updated their assessment several months into the government assistance program and with the benefit of hindsight determined that they had made the wrong decision—because they had not fully anticipated the extent to which factors such as consumer backlash, incentive-based compensation restrictions, and government meddling in strategic decisions would harm the company—otherwise it is difficult to explain why Chrysler pushed to exit government assistance six years ahead of schedule. Hence, one interpretation of the temporary underperformance is that Chrysler management made a strategic and costly error in taking government assistance, but corrected it as the consequences became apparent. Another interpretation is that Chrysler management would have preferred avoiding the costs associated with government assistance entirely.

**5.2 Extending the Analysis to General Motors**

Given our findings and our exploration of potential mechanisms which could explain Chrysler’s underperformance while maintaining government assistance, it is natural to wonder what the result would
have been for General Motors, as they also took government assistance. Unfortunately, the General Motors case is one where the synthetic control method cannot be applied fruitfully. Recall that the method cannot be used to analyze extreme statistical units—and note that General Motors is such a unit as it is 50% larger than any other auto manufacturer. The reason why synthetic control comes up short in these extreme cases is that any weighted average of control units, where the weights must sum to one, can never achieve the same scale as the extreme statistical units. Attempting to create a synthetic in this situation yields one with unacceptable pre-intervention prediction error. While we can’t apply synthetic control effectively to the General Motors’ case, our synthetic control analysis of Chrysler may still inform a qualitative study of that firm. To the extent that we believe, and can confirm, that the same mechanisms that were present at Chrysler were also present at General Motors, then we would expect to find similar results there as well.

5.3 Extrapolating to the Public Policy Debate

While we might be tempted to stretch our finding beyond the managerial question we asked, we are limited in how broadly we can interpret our results. We cannot fully resolve the hotly-debated public policy question: whether or not the government made the right decision in offering assistance to Chrysler. We use synthetic control to analyze how Chrysler’s sales would have evolved had the firm not taken government assistance. The government may, however, have had objectives in addition to maximizing the firms’ sales; namely, policymakers may have been interested in keeping an iconic American firm in business, ensuring the economic viability of Michigan’s economy, protecting existing union employment, and improving re-election prospects. What our analysis can tell us about the policy debate is that if selling more vehicles would have helped achieve policy goals, then the assistance program may have fallen short in this one aspect.

6 Conclusion

In this paper, we introduced the synthetic control method to management research. It is a quantitative tool that can help answer questions existing methodologies cannot answer, particularly when datasets are constrained by few observation units and even fewer that are treated. In an application that

---

40 Abide et al. (2010) note this problem with respect to New Hampshire in their smoking study focused on California.
demonstrates the utility of the method, we asked what impact did government involvement have on Chrysler’s performance following the recent financial crisis? Specifically, we estimated that Chrysler, or its remaining constituent parts, would have sold 29% more vehicles than a synthetic representation of the firm which refused conditional government assistance. Throughout the application, we provided direction for how to apply synthetic control methodology more broadly. In doing so, we demonstrated what concerns researchers should consider in robustness checks when applying the method in a management context. We have also made replication code publicly available to enable other researchers to learn how to implement the method and how to check the robustness of their results.

The synthetic control method fills an important gap in that it enables management researchers to better analyze the effects of phenomena that occur in limited populations and/or apply to only a small number of observation units. Previously management researchers had limited tools to accurately assess the magnitude and direction of such phenomena’s effects. Nevertheless, this is a particularly important task in empirical settings when competing theories predict big/small or positive/negative effects as was true in our context.

While we applied synthetic control in a single context, examining a firm and its interaction with its external environment, many other management applications remain to be explored in alternative contexts. We considered the firm the unit of analysis; however, the method can also be applied at an industry level, within divisions of firms, or even among individuals employed at firms. The treatment in our example was determined through interactions between a firm and its external environment, although other treatments need not be. The method could be applied to: industry self-regulation; technological breakthroughs at a firm or the adoption of new practices (e.g. ISO 9001); outsourcing decisions made differentially across divisions or facilities; or, modification of individuals’ incentive schemes. It may also be useful to address the growing interest in rare events (Lampel et al. 2009; Rerup 2009) given other approaches’ methodological limitations. Finally, there may be cost-saving practitioner applications to synthetic control where managers can explore new tactics on a sub-set of the firm and measure the associated cost or benefits.
REFERENCES


### APPENDIX: DATA DESCRIPTION AND SOURCES

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light vehicle sales volume</td>
<td>Total units sold by firm of new vehicles in the U.S., monthly</td>
<td>Ward’s Auto: U.S. Model Car Specifications and Prices</td>
</tr>
<tr>
<td>Average price</td>
<td>Average retail price of all vehicles sold (U.S. dollars), monthly</td>
<td>Author’s calculations based on data from Ward’s Auto: U.S. Model Car Specifications and Prices</td>
</tr>
<tr>
<td>Average fuel economy</td>
<td>Average fuel economy of all vehicles sold in the U.S. (miles per gallon), monthly</td>
<td>Author’s calculations based on data from Ward’s Auto: U.S. Model Car Specifications and Prices</td>
</tr>
<tr>
<td>Maximum fuel economy</td>
<td>Maximum fuel economy of all vehicles sold in the U.S. (miles per gallon), monthly</td>
<td>Author’s calculations based on data from Ward’s Auto: U.S. Model Car Specifications and Prices</td>
</tr>
<tr>
<td>Average weight</td>
<td>Average weight of cars sold in the U.S. (lbs.), monthly</td>
<td>Author’s calculations based on data from Ward’s Auto: U.S. Model Car Specifications and Prices</td>
</tr>
<tr>
<td>Average engine size</td>
<td>Average size of engines of all vehicles sold in the U.S. (Liters), monthly</td>
<td>Author’s calculations based on data from Ward’s Auto: U.S. Model Car Specifications and Prices</td>
</tr>
<tr>
<td>Number of active production platforms</td>
<td>Total number of production platforms for vehicles sold in the U.S., monthly</td>
<td>Author’s calculations based on data from Ward’s Auto: U.S. Model Car Specifications and Prices</td>
</tr>
<tr>
<td>Number of active brands</td>
<td>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</td>
<td>Author’s calculations based on data from Ward’s Auto: U.S. Model Car Specifications and Prices</td>
</tr>
<tr>
<td>Number of active series</td>
<td>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</td>
<td>Author’s calculations based on data from Ward’s Auto: U.S. Model Car Specifications and Prices</td>
</tr>
<tr>
<td>Number of market segments</td>
<td>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 van, middle SUV, middle CUV, small pickup, medium duty, large SUV, large regular, and large pickup), monthly</td>
<td>Author’s calculations based on data from Ward’s Auto: U.S. Model Car Specifications and Prices</td>
</tr>
<tr>
<td>Fraction of vehicles sold from the SUV, light truck, and van segments</td>
<td>Proportion of total U.S. sales that are made in the larger vehicle segments, monthly</td>
<td>Author’s calculations based on data from Ward’s Auto: U.S. Model Car Specifications and Prices</td>
</tr>
<tr>
<td>Fraction of vehicles manufactured outside North America</td>
<td>Proportion of total U.S. sales of imported vehicles, monthly</td>
<td>Author’s calculations based on data from Ward’s Auto: U.S. Model Car Specifications and Prices</td>
</tr>
<tr>
<td>Number of employees</td>
<td>Total number of worldwide employees, annual</td>
<td>Compustat</td>
</tr>
<tr>
<td>Leverage ratio</td>
<td>Total Long-term debt / Total Assets, annual</td>
<td>Compustat</td>
</tr>
</tbody>
</table>
Table 1: Comparison of Attributes between Chrysler, Ford, and Synthetic Chrysler in the period prior to Acceptance of Government Assistance

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Chrysler</th>
<th>Synthetic Chrysler</th>
<th>Ford</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price, Average of Vehicles Sold</td>
<td>26843</td>
<td>28241</td>
<td>28384</td>
</tr>
<tr>
<td>MPG, Average of Vehicles Sold</td>
<td>19.9</td>
<td>20.4</td>
<td>19.4</td>
</tr>
<tr>
<td>MPG, Maximum of Vehicles Sold</td>
<td>31.5</td>
<td>34.3</td>
<td>34.6</td>
</tr>
<tr>
<td>Weight (in lbs), Avg. of Vehicles Sold</td>
<td>4181</td>
<td>4175</td>
<td>4422</td>
</tr>
<tr>
<td>Engine Size (in L), Avg. of Vehicles Sold</td>
<td>3.95</td>
<td>4.00</td>
<td>4.35</td>
</tr>
<tr>
<td>Fraction of Sales in SUV/Truck/Van</td>
<td>0.751</td>
<td>0.612</td>
<td>0.657</td>
</tr>
<tr>
<td>Brands, # Active</td>
<td>4.3</td>
<td>2.6</td>
<td>3.0</td>
</tr>
<tr>
<td>Platforms, # Active</td>
<td>15.8</td>
<td>15.3</td>
<td>18.1</td>
</tr>
<tr>
<td>Segments of Market, # Active</td>
<td>15.3</td>
<td>14.6</td>
<td>16.0</td>
</tr>
<tr>
<td>Series of Vehicles, # Active</td>
<td>22.3</td>
<td>21.2</td>
<td>24.3</td>
</tr>
<tr>
<td>Fraction of Sales Manufactured Outside North America</td>
<td>0.008</td>
<td>0.103</td>
<td>0.000</td>
</tr>
<tr>
<td>Leverage (Debt/Assets)</td>
<td>0.249</td>
<td>0.328</td>
<td>0.391</td>
</tr>
<tr>
<td># of Employees</td>
<td>277211</td>
<td>268829</td>
<td>323050</td>
</tr>
<tr>
<td>Sales Volume</td>
<td>178759</td>
<td>193013</td>
<td>250620</td>
</tr>
<tr>
<td>Level of Sales Volume, Index (Dec ’08=100)</td>
<td>200.1</td>
<td>267.5</td>
<td>191.9</td>
</tr>
</tbody>
</table>

Table 2: Weights of Companies in Synthetic Chrysler

<table>
<thead>
<tr>
<th>Company</th>
<th>Weight</th>
<th>Company</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMW</td>
<td>0</td>
<td>Mitsubishi</td>
<td>0</td>
</tr>
<tr>
<td>Daimler</td>
<td>0.022</td>
<td>Nissan</td>
<td>0.169</td>
</tr>
<tr>
<td>Ford</td>
<td>0.664</td>
<td>Porsche</td>
<td>0</td>
</tr>
<tr>
<td>General Motors</td>
<td>-</td>
<td>Saab</td>
<td>0</td>
</tr>
<tr>
<td>Honda</td>
<td>0</td>
<td>Subaru</td>
<td>0</td>
</tr>
<tr>
<td>Hyundai-Kia</td>
<td>0</td>
<td>Suzuki</td>
<td>0</td>
</tr>
<tr>
<td>Isuzu</td>
<td>0.068</td>
<td>Toyota</td>
<td>0.077</td>
</tr>
<tr>
<td>Jaguar Land Rover</td>
<td>-</td>
<td>Volkswagen-Audi</td>
<td>0</td>
</tr>
<tr>
<td>Mazda</td>
<td>0</td>
<td>Volvo</td>
<td>0</td>
</tr>
</tbody>
</table>

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
Figure 1: Chrysler and Ford Sales around Chrysler’s Acceptance of Government Assistance

Figure 2: Chrysler’s & Synthetic’s Sales Volume around Acceptance of Government Assistance
Figure 3: Difference between Chrysler’s Actual & Synthetic Sales Volume around Acceptance of Government Assistance

Figure 4: Placebo Test, In Time. Chrysler & Synthetic’s Sales Volume around Hypothetical January 2005 Acceptance of Government Assistance
Figure 5: Placebo Test, Among Untreated Firms. Difference between Actual & Synthetic Sales Volume around January 2009 Acceptance of Government Assistance

Figure 6: Placebo Test, Among Untreated Firms. Difference between Actual & Synthetic Sales Volume around January 2009 Acceptance of Government Assistance (Excluding Firms with MSPE 5x higher than Chrysler’s prior to Acceptance of Government Assistance)
Figure 7: Dotplot of Treatment Extremity Measure, Among-Firm Placebos’ Post-Treatment/Pre-Payoff Performance Relative to Pre-Treatment Performance (with a Histogram Axis Border)

Figure 8: Leave-One-Out Tests

Chrysler accepts Government Assistance

Government assistance to Chrysler ends

Sales Volume

280,000
240,000
200,000
160,000
120,000
80,000
40,000
2005 2006 2007 2008 2009 2010 2011 2012

Chrysler
Synthetic, without Exclusions
Synthetics, Leaving One Out
Figure 9: Leave-One-Out Tests, Differences

Figure 10: Out-of-Sample Test
Figure 11: Out-of-Sample Test, Differences

Chrysler accepts Government Assistance

Government assistance to Chrysler ends

Gap between Synthetic and Actual