Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis BSc^2 Economics and Econometrics

The Role of Covariates in Synthetic Control Methods and Specification-Searching

Hendrik Johannes de Boer (527932)

Supervisor:	Wendun Wang
Second assessor:	Jens Klooster
Date final version:	30th June 2023

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam.

Abstract

This study examines the use of covariates in synthetic control methods (SCMs) and their impact on specification-searching opportunities. By incorporating elements from prior research, the paper analyzes the inclusion and exclusion of covariates in SCMs, considering their advantages and drawbacks. Based on Monte Carlo simulations, the study recommends a case-by-case approach for including covariates, particularly in situations with limited pretreatment periods or when covariates exhibit multicollinearity or nonlinear effects. The findings of this paper suggest that when including covariates, carefully selecting a small set of pretreatment outcome lags, distributed across the pretreatment period, is an advisable strategy to improve SCM accuracy. The study provides insights into the implications of covariate inclusion and offers guidance for researchers seeking to address specification-searching challenges in SCMs, but also for researchers aiming to efficiently use covariates in their analysis.

1 Introduction

Comparative case studies are a cornerstone for empirical research in order to discover the effect of a certain treatment or shock. They may be useful in a variety of fields and applications, such as in public policy to measure the effects of a newly enacted tax law; or in environmental economics to measure the effects of climate change on economies. In comparative case studies this is done by estimating the supposed path taken by the aggregate outcome of a unit had the treatment not taken place at all, and comparing it with the realized path of the aggregate outcome. In the past, favoured methods to estimate the treatment effects were difference-in-difference, propensity score matching, or regular regression based methods. A more recent technique, however, suggested in Abadie and Gardeazabal (2003) and developed in Abadie, Diamond and Hainmueller (2010), is the synthetic control method (SCM). This method constructs a comparison unit, called a synthetic control unit (SCU), through a data-driven procedure in which a set of similar units that were unaffected by the treatment are combined in order to create a synthetic comparison unit. For example, in their study of the effect of newly instigated tobacco legislation in California, Abadie et al. (2010) use a combination of 5 states unaffected by the legislation to construct a SCU.

SCM boasts three important advantages. First, it allows for a high degree of transparency, as the constructed counterfactual of interest is a combination of weighted control units. This allows an understanding of which control units contributed to the SCU and hence the similarities and differences between the synthetic counterfactual and the unit subject to intervention. Secondly, as the SCM also uses real, observed data of the control units to estimate treatment effects, there is almost no possibility for extrapolation. This also reduces the abilities for manipulation. Finally, postintervention outcomes are not required to construct the SC weights, this means that researchers are able to construct study designs prior to the intervention and will therefore not be subject to potential biases induced by better findings.

Despite these advantages, Ferman, Pinto and Possebom (2020) highlight a major caveat of the SCM, namely, the lack of consensus on the choice of predictor variables and covariates used to establish the synthetic control weights. Without clear standard guidelines on which sets of predictor variables and covariates to use, a researcher can choose specifications that construct

SCUs in such a way that their results become "statistically significant" despite there being no real effect. Specifications can include both pre-treatment outcomes, linear combinations of pre-treatment outcomes, and covariates. In their paper, Ferman et al. (2020) focus especially on the variations in specifications in terms of their pre-treatment outcomes and pay relatively little attention to the role of covariates. Ultimately, Ferman et al. (2020) recommend using a specification in which all pre-treatment outcome values are used to assign weights, and only using covariates when there is 'a strong prior belief that it is crucial to balance on a specific set of covariates'. Botosaru and Ferman (2019) affirm that an approximate balance on covariates is not always necessary if there is a perfect balance on pre-treatment outcomes, and that in certain cases it may be advantageous not to include them. However, they do emphasize the importance of covariates in being able to reduce biases and the amount of assumptions needed. Additionally, the work of Kaul, Klößner, Pfeifer and Schieler (2022) strongly suggests covariates with predictive power should be matched too and they find that the exclusion of covariates can dramatically influence results in certain cases.

This paper therefore initially replicates the results of Ferman et al. (2020) and their discussion on the opportunities for specification-searching within SCMs. Afterwards, at the hand of the papers of Kaul et al. (2022) and Botosaru and Ferman (2019), as well as the Monte Carlo simulations of Ferman et al. (2020), an investigation is made into the degree to which including covariates in specifications affect the weights assigned in SCMs, the relative performance of SCUs that use covariates, and conclusively, the role covariates have in specification-searching for SCMs. By examining several Monte Carlo simulations for cases where two different DGPs are used in combination with covariate and standard specifications, this paper is able to identify that more opportunities for specification-searching arise when including or excluding covariates in the wrong cases. Additionally, the role of covariates is compared for different pre-treatment periods T_0, also for exceptionally large and small cases. These variations help to define cases where covariates are useful, and cases when they are irrelevant or even obstructive. This paper recommends that the decision to include covariates in an SCM should be handled on a case-by-case basis, depending on the properties of the potential covariates. Additionally, it finds that including covariates is more relevant when T_0 is smaller, and that certain specifications are able to make predictions more effectively when covariates are included. At the same time, it is also found that in some cases covariates can cause more bias or complications. These findings reinforce the contribution made by Ferman et al. (2020) on minimizing the room for the manipulation of results when using SCMs. From a grander perspective, this will help researchers and policymakers make more accurate predictions about the effects of shocks, treatments or policies and thus allow for more effective and reliable decision-making.

The research questions are posed as follows. The main research question will be: *To what extent are there specification-searching opportunities when using synthetic control methods and what role do covariates have in this?* The following subquestions are also constituted in order to complement the main research question above:

1. *According to economic and econometric theory, when should covariates be included in synthetic control methods?*

2. *When including covariates, how does this affect the weights assigned in the SCMs?*

3. *Does omitting or including covariates allow for specification-searching opportunities? And what guidelines should be followed to minimize this possibility?*

The rest of the paper is structured as follows: in Section 2, a comprehensive literature review will cover the economic and econometric developments in SCMs and the role of covariates within it; in Section 3, the relevant theory of papers Ferman et al. (2020), Botosaru and Ferman (2019) and Kaul et al. (2022) will be conjoined to provide a theoretical background of covariates within SCMs; in Section 4 the methods used in the analysis within this paper will be explained; in Section 5, the findings of this paper will be presented and discussed; an empirical case-study will be considered in Section 6; and finally, in Section 7, a conclusion is given.

2 Literature Review (Extension)

The conception of SCMs occurred in the paper by Abadie and Gardeazabal (2003), in which a data-driven procedure is used to construct a SCU. They apply this technique by estimating the effects of terrorism in the Basque region in Spain by constructing a SCU consisting of two unaffected Spanish regions. In their follow up paper, Abadie et al. (2010) build an explicit framework and methodology on SCM and apply it to the effects of tobacco legislation in California. Their research revealed the potential of SCMs by discovering that the anti-tobacco legislation had in fact reduced smoking far more dramatically than concluded in previous studies. This paper revealed the potential of SCMs and motivated further research, also on why the SCMs were more beneficial than existing methods. Abadie, Diamond and Hainmueller (2015) propose that because of the method's ability to explicitly describe the contribution of each of the control units to the SCU, SCMs allow for both qualitative and quantitative analysis when comparing the synthetic unit with the treated unit. Building on this, Abadie (2021) mentions advantages of the SCM over conventional regression methods: it is safeguarded against extrapolation; the discrepancy between the treated unit and the SCU can be transparently assessed; it is relatively more safe against specification-searching; there is transparency on the construction of the counterfactual; and it has sparse properties.

Throughout this period, research was carried out on how weights were assigned to individual control units and what effect this had on the SCU. In the paper of Doudchenko and Imbens (2016), assumptions on the weights were relaxed. While this may allow for more flexibility in carrying out comparative case studies, this also creates more room for discrepancy and inconsistent findings. In fact, even with the assumptions of regular SCMs, the lack of guidance on which predictors to choose results in ambiguity, and does not provide conclusive evidence in individual cases (Dube & Zipperer, 2015). This grey area in SCMs is addressed especially in the paper of Ferman et al. (2020).

In their paper, Ferman et al. (2020) state that manipulation of the results due to the wide choice of predictors that can be used to assign weights is an important issue for research using SCMs. They conclude that because of this the SCM does not retain the property of being solely a data-driven process and gives the researcher some degree of discretionary power. To address this issue, Ferman et al. (2020) theoretically derive conditions under which a subset of common specifications should lead to asymptotically equivalent estimators. Additionally, they

provide insight into when specification-searching opportunities arise, namely when there are limited pre-treatment periods. The authors conclude that to relieve the specification-searching problem they advise a restriction for only a subset of specifications, and to always include the specification in which all pre-treatment outcomes are included as benchmarks. Their research particularly focuses on the role of pre-treatment outcome lags as matching variables, and examine specification-searching opportunities for seven specifications that vary in terms of their pre-treatment outcome values used. The discussion on how covariates may or may not allow for specification-searching opportunities, however, is quite limited.

2.1 Covariates

In the work of Abadie et al. (2010), which developed the SCM, the authors assumed a high degree of balance for both the pre-treatment outcomes as well as the covariates. This allowed them to construct outcomes with a smaller bias. While Ferman et al. (2020) provides substantial insight into the role of pre-treatment outcomes, it is limited in its discussion on covariates. Other literature on the role of covariates in synthetic methods is limited as well, or do not fully appreciate their contribution. For example, Doudchenko and Imbens (2016) merely suggests that covariates play a relatively minor role and that they can be overlooked to a large extent.

In contrast, Gilchrist, Emery, Garoupa and Spruk (2023) state that the effects of a policy can only be accurately measured if the SCU also minimizes the imbalance of covariates. Botosaru and Ferman (2019) provide an extensive theoretical background into the importance of balancing the covariates in the SCU and the real treatment unit. Their paper provides two main contributions: firstly, they suggest that covariates do not always have to be included and should be considered on a case-by-case basis. However, if they are not included then a higher number of pre-treatment periods are required to derive similar bounds. Yet, even then, these bounds are more loose than when a perfect balance on covariates is also imposed. Their second contribution concerns the insufficiency of a perfect balance on pre-treatment outcomes. This is because a perfect balance on lagged outcomes does not ensure an approximate balance for covariates when those covariates have multicollinear effect on potential outcomes, or if their effect on potential outcomes are nonlinear. At the same time, Botosaru and Ferman (2019) warns not to include all pre-treatment outcome lags with relevant covariates due to the findings of Kaul et al. (2022).

Kaul et al. (2022) find that including all pre-treatment outcome values together with covariates renders the covariates irrelevant in the SCM. This occurs because in this case only the pre-treatment fit with respect to the variable of interest is optimized, even when covariates are in fact important in reducing the bias of estimates. In a lot of cases, covariates can be highly influential in reducing the bias in SCMs. This is shown in the Monte Carlo simulations of Kaul et al. (2022), where specifications with a single pre-treatment outcome lag and covariates provided more informative and less biased results than the specification with all pre-treatment outcome lags. The findings of Botosaru and Ferman (2019) and Kaul et al. (2022) indicate that the inclusion or exclusion of covariates can provide differing results. Additionally, recent papers have implemented these new findings when constructing their SCUs. In their paper, Ben-Michael, Feller and Rothstein (2021) suggest a new augmented SCM, for which it is found that including covariates improves the pre-treatment fit. Gilchrist et al. (2023) also includes a reduced set of

pre-treatment outcomes so as to ensure that the outcome variable 'does not swallow the importance of ... covariates'. In summary, these papers therefore suggest that covariates have an important role to play in specification-searching opportunities for SCMs.

While Ferman et al. (2020) do address the role of covariates in their paper, they do not discuss exactly in which cases or why the exclusion or inclusion of covariates allows for specification-searching opportunities. Additionally, in the case where covariates are included, they contradict the findings of Kaul et al. (2022) by stating that 'only specifications that satisfy the asymptotic equivalence assumptions' should be compared, whereas Kaul et al. (2022) and Botosaru and Ferman (2019) state explicitly that specifications with less pre-treatment outcome lags allow the covariates to be more influential in reducing bias. It is also these cases that Ferman et al. (2020) largely overlook. This research paper aims to contribute to the existing econometrics literature by providing a discussion on covariates within the specification-searching framework of Ferman et al. (2020), as well as providing more clarity on when covariates should and should not be included in SCMs.

2.2 Economics Literature on the role of Covariates

When using SCMs, it is fundamental to discern if covariates should be included or not. While this is also influenced by certain econometric factors that will be investigated later in this paper, it ultimately depends on the question: 'Are covariates relevant in determining the outcomes of the dependent variable?' If covariates are deemed to be relevant, the researcher should include them in the specification and attempt to balance them. However, within economics, there are certain cases where covariates are relevant in determining potential outcomes, and other cases where they are less relevant. This is largely influenced by the outcome variable itself, as some outcome variables may be heavily determined by covariates whereas other may not. SCMs are especially relevant in the fields of policy and macroeconomics: "The synthetic control approach... is arguably the most important innovation in the policy evaluation in the last 15 years" (Athey & Imbens, 2017). The economy as a whole is influenced and described by a multitude of factors and indicators; some common examples are gross domestic product (GDP), inflation rates, unemployment figures and investment figures. These indicators tend to have large predictive power in macroeconomic forecasting because they are so intertwined. On the other hand, there are also other macroeconomic indicators that are less relevant depending on the policy evaluation application at hand. For example, consider indicators such as unemployment rates or expenditure on education and their relation to stock prices, though there may be some predictive power, it is likely to be less pronounced. Elliott and Timmermann (2016) argue that predictors should only be included in forecasting applications when 'new predictors have genuine predictive power over the outcome', as due to sparseness only a small subset of covariates will truly have influence within the model. Carriero, Galvão and Kapetanios (2019) also suggest this, indicating that models, albeit not SCMs, including a smaller set of hand-picked predictors perform just as well as models with a large number of predictors.

For previous applications using the SCM (see Table 1), the number of covariates varies. Given that researchers decide to use covariates in their synthetic control approach, it depends on the economic context which covariates should be included and how many. For the published

articles in Table 1, most researchers use less than 15 predictors, as is supported by other papers (Elliott & Timmermann, 2016; Carriero et al., 2019). For example, in the paper by Billmeier and Nannicini (2013), the authors opt for six covariates to determine GDP per capita: investment as a share of GDP, population growth, secondary school enrollment, average inflation rate, and a democracy dummy. Additionally, in macroeconomics, the state of the economy is in constant flux, meaning that all leading indicators are constantly changing values. Though there are many combined indexes (Stock & Watson, 1989), covariates are usually included individually because the variations over time are not always proportional. Additionally, Botosaru and Ferman (2019) advise omitting irrelevant covariates as their inclusion enables multicollinearity to be present and therefore requires the SCM to adopt stronger assumptions. As a result, within the SCM framework, researchers should carefully select relevant covariates based on the economic context of the research question at hand and omit those covariates which are less relevant. One example of a selection process is by consulting economic theory on the specified topic. For example, when forecasting inflation one may consult literature on the Phillips curve (Phillips, 1958). Another example of an effective selection process is given by Gilchrist et al. (2023), wherein the researchers analyze the predictive power of covariates by analyzing their contribution to the SCU after initial simulations. Those covariates with low contributions are then discarded from the model.

Table 1: Specification Summary on Published Articles using SCM

Published Article	T_0	r	Lags	J
Abadie and Gardeazabal (2003)	10	11	Mean	16
Kleven et al. (2013)	11	3	Mean	14
DeAngelo and Hansen (2014)	37	14	Mean	46
Lindo and Packham (2017)	6	0	-1, -3, -5	38
Dustmann et al. (2017)	6	5	All	85
Cunningham and Shah (2018)	18	5	10^a	50
Montalvo (2011)	4	2	0, -1	32
Hinrichs (2012)	9	0	All	3-7
Billmeier and Nannicini (2013)	2-32	5	All	4-62
Cavallo et al. (2013)	11	7	First Half	53
Bohn et al. (2014)	9	42	All	45
Gobillon and Magnac (2016)	8	0	All	135
Smith (2015)	10-43	2	0, -2, -4, -6	7-32
Zou (2018)	2	6	All	2429
Walker et al. (2014)	7	0	Mean	36
Eren and Ozbeklik (2016)	19	7	Even Lags	28
Bartel et al. (2018)	5	11	All, Mean	49

Notes: T_0 is the number of pre-treatment outcomes available in the dataset, r is the number of covariates included in the SCM, Lags are the set of pre-treatment outcomes included in the SCM and J is the number of control units in the donor pool. The articles named above have been retrieved from various journals: American Economic Review, American Economic Journal-Economic Policy, American Economic Journal-Applied Economics, Quarterly Journal of Economics, Review of Economic Studies, Review of Economics and Statistics, Journal of Development Economics, Journal of Labor Economics, and Journal of Policy Analysis and Management. a: The specific outcome lags used in this specification are 0, -1, -2, -7, -8, -9, -11, -14, -15, -16

2.3 Economic Relevance of the SCM

The economic relevance of the SCM is undisputed. It provides numerous advantages over existing methods such as difference-in-difference or propensity score matching. These advantages

stimulate the application of this method to problems in various economic contexts. The superiority of the SCM is most evident in contexts where a comparable counterfactual or control group is not immediately available, such as in policy evaluation (Athey & Imbens, 2017). By constructing a comparison unit with closely correlated characteristics, the control group in the SCM is more reliable than it would be for other methods. In the field of policy, this is extremely useful, because matching is often not possible; it is often not possible to find a control unit when using matching which has identical characteristics. For the difference-in-differences method the same issue holds: it is difficult to know if the counterfactual has the same functional form as the control group and therefore this method relies heavily on assumptions (Athey & Imbens, 2017) and extrapolation.

While the SCM still makes some assumptions and is certainly exposed to potential biases, it is an effective method to largely avoid this. This allows it to construct reliable and data-driven counterfactuals and therefore does not rely as much on assumptions. At the same time, it also does not require identical control units, which are often hard to find, but uses a combination of similar control units. From an economic perspective, the benefits of the SCM can be understood within the framework of the potential outcomes framework (Imbens & Rubin, 2010).

The first advantage of a well-constructed SCU is that causal effects can be defined before specifying the assignment mechanism, and without making functional form or distributional assumptions (Imbens & Wooldridge, 2009). Secondly, SCM provides more clarity on causal inference when using an effective SCU; the construction of an accurate counterfactual means that the research is able to effectively decipher the causality inferred by explicit treatments by analyzing the alternate potential outcome using the control unit. A final advantage from an economic perspective is that there is limited room for manipulation as it is a data-driven method. This allows for economic inferences to be made transparently (Abadie et al., 2010).

These advantages make the SCM useful for applications in many different economic contexts. A short description of the economic application of some of the papers in Table 1 is given below. It is a recurrent theme that SCMs are applied to contentious topics. While these topics were formerly subject to interpretation, the accuracy and transparency of SCMs allow for more conclusive research. Dustmann, Fasani, Frattini, Minale and Schönberg (2017) measure the effect that refugee crises and economic migration have on the labour market. They do this using the SCM in which they estimate gaps in employment probabilities. The paper emphasizes the risk that immigration has on economic stability and therefore calls for different policy measures, such as increased coordination at a multilateral level.

Cunningham and Shah (2018) uses the SCMs to evaluate health policy, they measure the impact of decriminalizing indoor prostition on the number of contractions of sexually transmitted diseases. They find that rape offences as well as incidences of female gonorrhoea fell. Cavallo, Galiani, Noy and Pantano (2013) use SCMs to estimate the effect of natural disasters on GDP. Additionally, they are able to include covariates to control for certain factors such as political changes proceeding the shocks.

Finally, SCMs have also been applied to educational contexts. Hinrichs (2012) finds that affirmative action bans reduce minority enrollment at selective US colleges. The SCM is used here to construct counterfactual for particular selective US colleges. This is an example of how

SCMs can be applied at local levels as well. There is therefore ample ground to support the use of SCMs in economic research.

3 Theoretical Background

The framework of the theoretical component of this paper is established in the following section, this theory will be used as a basis for evaluating the results in Section 5.

3.1 Synthetic Controls and Specification Searching (Replication)

The SCM of Abadie et al. (2010) constructs a counterfactual in the case where there is only one unit that is treated. It does this using data from a donor pool of untreated control units. The method then uses a weighted average of these selected units to estimate treatment effects for every post-treatment period.

Let us suppose that the duration of interest spans $T \in \mathbb{N}$, and that there is available data for $(J+1) \in \mathbb{N}$ units. For $j \in 1, ..., J+1$, only $j = 1$ is treated in some time period T_0. We denote $Y^0_{j,t}$ as the potential outcome that would be observed for unit j in period $t \in 1, ..., T$ had no treatment taken place. We denote $Y^1_{j,t}$ as the potential outcome under treatment. This results in the following model:

$$\begin{cases} Y^0_{j,t} = \delta_t + \theta_t Z_j + \lambda_t \mu_j + \varepsilon_{j,t} \\ Y^1_{j,t} = \alpha_{j,t} + Y^0_{j,t} \end{cases} \tag{1}$$

Here, δ_t is an unknown common factor with constant factor loadings across units: (θ_t) is a $(1 \times r)$ vector of unknown parameters; Z_j is a $(1 \times r)$ vector of observed covariates which are not affected by the treatment; λ_t has dimensions $(1 \times (F-r))$ vector of common factors; μ_j is a $(1 \times (F-r))$ vector of unknown factor loadings; and $\varepsilon_{j,t}$ are unobserved transitory shocks. From (1), it can be derived that the treatment effect is $\alpha_{j,t} := Y^1_{j,t} - Y^0_{j,t}$. $Y_{j,t}$ is the observed outcome. As we are interested in the treatment effect after T_0, we aim to identify $\alpha_{1,T_0+1}, ..., \alpha_{1,T}$. To do this, we require the potential outcome $Y^1_{1,t}$, which is observed for periods $t > T_0$, as well as the potential outcome $Y^0_{1,t}$, which acts as the counterfactual and needs to be estimated.

$\mathbf{Y_j} := [Y_{j,1}...Y_{j,T_0}]'$ is the vector containing pretreatment observed outcomes for each $j \in (1, ..., J+1)$. To predict outcomes $Y^0_{j,t}$ for $t > T_0$, we use F predictor variables, which are either covariates or linear combinations of observed pretreatment outcomes in $\mathbf{Y_j}$. These predictors are included in the $(F \times 1)$-vector $\mathbf{X_j}$. We let $\mathbf{Y}_0 = [\mathbf{Y}_2|\cdots|\mathbf{Y_{J+1}}]$ be a $(T_0 \times J)$-matrix containing all the pretreatment observed outcomes for each unit int he donor pool. Also, we let $\mathbf{X}_0 = [\mathbf{X}_2|\cdots|\mathbf{X_{J+1}}]$ be a $(F \times J)$-matrix containing values of all the predictors for each unit in the donor pool.

Initially for the sake of simplicity, we assume that the predictors in $\mathbf{X_j}$ are balanced already. Then, the aim of the SCM is to construct a counterfactual using a weighted average of the control units, $\hat{Y}^0_{1,t} := \sum_{j=2}^{J+1} \hat{w}_j Y_{j,t}$.

The weights $\hat{\mathbf{W}} = [\hat{w}_2, ..., \hat{w}_{j+1}]' := \hat{\mathbf{W}}(\hat{\mathbf{V}}) \in \mathbb{R}^J$ are assigned through a nested minimization

problem, also called the inner-optimization:

$$\hat{\mathbf{W}}(\mathbf{V}) := \underset{\mathbf{W}\in W}{arg\,min}(\mathbf{X}_1 - \mathbf{X}_0\mathbf{W})'\mathbf{V}(\mathbf{X_1} - \mathbf{X_0}\mathbf{W}) \tag{2}$$

where the sum of weights must be equal to one, and all weights must be non-negative, as stated in the following assumption.

$$\mathbf{W} \in \{w_2, ..., w_{J+1} | \sum_{j=2}^{J+1} w_j = 1 | \text{ and } w_j \geq 0\} \tag{3}$$

Solving Equation (2) yields $\hat{\mathbf{W}}$, where the individual weights $\hat{w_2}, ..., \hat{w_{j+1}}$ indicate the respective composition of each control unit in the synthetic control unit. Here, $\mathbf{V}$ is a diagonal positive semi-definite matrix of dimension $(F \times F)$, and is defined by the outer-optimization:

$$\hat{\mathbf{V}} := \underset{\mathbf{V}}{argmin}(\mathbf{Y_1} - \mathbf{Y_0}\hat{\mathbf{W}}(\mathbf{V}))'(\mathbf{Y_1} - \mathbf{Y_0}\hat{\mathbf{W}}(\mathbf{V})). \tag{4}$$

In (4), $\hat{\mathbf{V}}$ is a matrix containing weights indicating the relative importance of each covariate and pre-treatment outcomes included in $\mathbf{X_0}$. A data-driven optimization procedure which solves (4) is used to determine $\hat{\mathbf{V}}$. Equation (4) essentially determines the value of V for which the mean squared prediction error (MSPE) is minimized for $\mathbf{Y}$ (where $\mathbf{Y}$ is a matrix containing both $\mathbf{Y_1}$ and $\mathbf{Y_0}$). Thereafter, $\hat{\mathbf{V}}$ is used to construct $\hat{\mathbf{W}}(\hat{\mathbf{V}})$. The resulting weight matrix is used to construct the synthetic counterfactual $\hat{Y}^0_{1,t}$. For each period $t \in 1, ..., T$, we can use the observed outcomes of unit 1 ($Y_{1,t}$) to construct the synthetic control estimator (SCE) of $\alpha_{1,t}$. The estimator, also called the estimated gap, is estimated as $\hat{\alpha}_{1,t} := Y_{1,t} - \hat{Y}^0_{1,t}$. This is an estimation of the treatment effect in period t. This is the general framework of the SCM.

A significant issue highlighted in the paper of Ferman et al. (2020) is the lack of consensus on the predictors chosen for $\mathbf{X_j}$. Since the conception of the SCM, a wide spectrum of specifications has been used in $\mathbf{X_j}$ in papers using SCMs. These specifications vary widely in terms of the selection and number of covariates and pretreatment outcome lags used. As a result, the outcomes of the optimization procedure (4) and the nested minimization problem (2) are determined, at least partially, by the choice of predictors in $\mathbf{X_0}$. By being able to influence the weight matrix $\hat{\mathbf{W}}(\hat{\mathbf{V}})$, it is possible to influence the resulting outcomes of the synthetic counterfactual $\hat{Y}^0_{1,t}$ such that the SCE $\hat{\alpha}_{1,t}$ becomes significant for the test at hand. This allows room for manipulation of the results and undermines the 'transparency' advantage of SCMs.

To illustrate the prevalence of this issue, Ferman et al. (2020) assemble a wide variety of papers using SCMs and highlight the variety of specifications used. This table (Table 1) was shown in the previous section. Table 1 indicates the sheer amount of variation in terms of specifications. This begs the question: how influential is the choice of specifications and do they allow for 'significant' treatment results despite their being no actual significant treatment effect? To answer this, Ferman et al. (2020) investigate under which conditions specification are asymptotically equivalent, albeit with a focus on the role of outcome lags. Their findings will be covered in Subsection 3.2. To help answer this question comprehensively, the theory of Botosaru and Ferman (2019) and Kaul et al. (2022) will be used to provide insight into how the choice of

covariates may contribute to the issue of specification-searching. This is covered in Subsection 3.3.

3.2 Equivalence of Specifications (Replication)

Ferman et al. (2020) examine the effect of $T_0 \to \infty$ on different specifications. A specification s is defined by the set of predictors included in $\mathbf{X}_j(s, T_0)$, this set of predictors can consist of covariates, pre-treatment outcome lags, and/or functions of pre-treatment outcome lags. The number of pretreatment periods is labeled as $L(s, T_0)$ such that $Y_{j,t}$ is included as a predictor when there are T_0 pretreatment periods. Thus, $L(s, T_0)$ indicates the number of pretreatment outcome lags included.

$\hat{\mathbf{W}}(s, T_0)$ is denoted as the weight matrix used when we apply specification s in combination with T_0 pre-intervention periods. We compare the asymptotic behaviour when $T_0 \to \infty$ for two different specifications s and s'. Ferman et al. (2020) investigate under which conditions $\hat{\mathbf{W}}(s, T_0)$ and $\hat{\mathbf{W}}(s', T_0)$ converges to the same $\bar{\mathbf{W}}$ when $T_0 \to \infty$. The sole condition required for this convergence to hold is that the dispersion of different subsequences are stable, so that the second moments converge to the same value. Furthermore, if $L(s, T_0) \to \infty$ when $T_0 \to \infty$, then $\hat{\mathbf{W}}(s, T_0) \to_p \bar{\mathbf{W}} = arg\min_{\mathbf{W} \in W} Q_1(\mathbf{W})$, where $Q_1(W)$ is a continuous and strictly convex function. This results in the two SCEs, $\hat{\alpha}_{s,t}$ and $\hat{\alpha}_{s',t}$, converging in probability to zero ($|\hat{\alpha}_{1t}(s, T_0) - \hat{\alpha}_{1t}(s', T_0)| = o_p(1)$), given that both $L(s, T_0) \to \infty$ and $L(s', T_0) \to \infty$ when $T_0 \to \infty$. This is because the inner optimization (2) assigns positive weights to the pre-treatment outcome lags in this case. Thus, asymptotically, synthetic control weights for all specifications will be assigned by minimizing an average of a function of the pre-treatment outcomes that are included as predictors.

This theoretical result only holds when $L(s, T_0) \to \infty$ and $T \to \infty$. When T_0 is finite, this result does not hold. In most practical applications, there is a limit to the number of pre-treatment observations. Additionally, it is not always given that $L(s, t) \to \infty$ when $T_0 \to \infty$, as some specifications used in papers only include a specific amount of predictors or only use the mean. Thus, in most practical applications, synthetic control weights will not converge to $\bar{\mathbf{W}}$.

3.3 Covariates (Extension)

This subsection details the role of covariates within the SCM framework, as well as special cases that are relevant for the discussion in Section 5. As stated earlier, the aim of the SCM is to construct a counterfactual :

$$\hat{Y}^0_{1,t} := \sum_{j=2}^{J+1} \hat{w}_j Y_{j,t} \tag{5}$$

where $\hat{\mathbf{W}}$ is the optimal weight matrix. However, Ferman et al. (2020) require only Assumption (5), whereas the original paper on SCM by Abadie et al. (2010) assume a $\hat{\mathbf{W}}$ which satisfies the assumption of a perfect balance on the vector of observed covariates Z_j for unit j as well:

$$Z_1 = \sum_{j=2}^{J+1} \hat{w}_j Z_j \tag{6}$$

where Z_1 is the $(1 \times r)$ vector containing the covariates of the treated unit. Additionally, let $\mathbf{Z_0}$ be the $(J \times r)$ matrix of observed covariates for the J control units.

3.3.1 Omitting the balance on covariates

Botosaru and Ferman (2019) describe the effect which omitting Assumption (6), a perfect balance on covariates, has on the bounds. Omitting Assumption (6) requires us to introduce an assumption of linear independence for the $(1 \times F)$ row vector of effects of the observed and unobserved covariates ($\gamma_t \equiv (\theta_t, \lambda_t)$), where $\xi(T_0)$ is the smallest eigenvalue of $\frac{1}{T_0}\sum_{t=-T_0+1}^{0} \gamma_t'\gamma_t$:

$$\xi(T_0) > 0 \tag{7}$$

Here, $\xi(T_0)$ measures the degree of linear independence in the pre-treatment period of the effects of the observed and unobserved covariates. It is therefore also a required condition that $T_0 \geq F$.

If the effects of the observed and unobserved covariates are linearly independent as in Assumption (7), and a few minor assumptions described in Appendix A hold, it is possible to bound the degree of imbalance for the covariates given a perfect balance on pre-treatment outcome lags, thus allowing bounds to be derived for the SCE. An incomplete but sufficient version of this outcome is seen below, as outlined in Botosaru and Ferman (2019) Proposition 3.1:

$$|\mathbb{E}[\hat{\alpha}_{1,t}^*] - \alpha_{1,t}| \leq \frac{C_\alpha \bar{\gamma}(T_0)^2}{\xi(T_0)} max\left\{\frac{\bar{m}_p(T_0)^{1/p}}{T_0^{1-1/p}}, \frac{\bar{m}_2(T_0)^{1/2}}{T_0^{1/2}}\right\} \tag{8}$$

where $\bar{m}_p(T_0)$ can be interpreted as a measure for the scale of the transitory shocks, and $\bar{\gamma}(T_0)$ can be interpreted as the effect of the (un)observed covariate with the largest absolute shock (Appendix A). The main takeaways of (8) is that these bounds are increasing with a higher scale of transitory shocks $\varepsilon_{j,t}$ (through $\bar{m}_p(T_0)$), decreasing with the number of pre-treatment outcomes T_0 and decreasing with the degree of linear independence $\xi(T_0)$. This indicates that when covariates are linearly independent, they can be beneficial in improving the accuracy of forecasts in SCMs.

In the case that Assumption (7) does not fully hold, the following relaxed assumption is required to construct bounds. $\bar{m}_p(T_0)$ and $\bar{\gamma}(T_0)$ are required to be bounded, and there are positive constants $\tilde{\xi}$ and $\bar{T}$ for which $\xi(T_0) \geq \tilde{\xi}$ for all $T_0 > \bar{T}$, as a result the bounds asymptote to zero as T_0 increases. This indicates that when there is linear independence for the effects of observed and unobserved on potential outcomes, observed and unobserved covariates must be balanced in order to provide a good balance for pre-treatment outcomes when T_0 is larger.

However, when both Assumptions (5) and (6) are satisfied, then Assumption (7) can be replaced by $\dot{\xi}(T_0) > 0$. Here $\dot{\xi}(T_0)$ is the smallest eigenvalue of $\frac{1}{T_0}\sum_{t=-T_0+1}^{0} \lambda_t'\lambda_t$, and thus only requires the unobserved covariates to be linearly independent. Therefore, it may sometimes only be possible to construct bounds given both Assumptions (5) and (6), and not when only Assumption (5) holds. Otherwise, linear independence of the effects of both unobserved and observed covariates on potential outcomes is required, which in turn requires at least $T_0 \geq F$. If both assumptions hold, meaning covariates are also balanced, only the linear independence of the unobserved common factors is required, which requires minimally $T_0 \geq (F - r)$. As a result,

when covariates are not balanced, a higher amount of pretreatment periods T_0 is required to ensure bounds can be constructed.

3.3.2 Reducing bias

When both Assumptions (5) and (6) hold, the bounds will be tighter than when only Assumption (5) is valid. This is because Assumption (6) eliminates any potential bias arising from unbalanced observed covariates. As a result, researchers should aim to find weights that satisfy both of these assumptions, as otherwise, stronger assumptions are required to guarantee the existence of bounds on the bias of the SCE as described earlier. Often however, due to empirical limitations, relevant covariates may not all be observed and additionally, weights that balance both the pretreatment outcome lags as well as the covariates may not exist. Nevertheless, a perfect balance in pretreatment outcome lags can often imply an approximate balance for covariates. This is a result established in the paper by Botosaru and Ferman (2019) in their Proposition 3.1. The exceptions to this rule are when there exists significant multicollinearity such that Assumption (7) does not hold, when included covariates are irrelevant, leading to multicollinearity, and when covariates Z_j have a nonlinear effect on outcomes $Y_{j,t}$. In these cases, while bounds can still be constructed, biases are reduced substantially when there is also a balance on covariates.

Using derivations of Kaul et al. (2022), we assess the importance of covariates in reducing bias. Let $\tilde{Y}_{j,t}$ be the hypothetical value of $Y_{j,t}$ if it was not influenced by covariates. Therefore, usually it is the case that $Y_{j,t} = \tilde{Y}_{j,t} + \theta_t Z_j$ for some $\theta_t \in \mathbb{R}^{1\times r}$, and Z_j are the covariate values of unit j. The difference between the treated unit and its synthetic counterpart in terms of the dependent variable is then

$$Y_{1,t} - \sum_{j=2}^{J+1} w_j Y_{j,t} = \tilde{Y}_{1,t} - \sum_{j=2}^{J+1} w_j \tilde{Y}_{j,t} + \theta_t \left(Z_1 - \sum_{j=2}^{J+1} w_j Z_j \right) \tag{9}$$

Covariates eliminate a bias with a magnitude of $\theta_t \left(Z_1 - \sum_{j=2}^{J+1} w_j Z_j \right)$, and thus play an important role in accurate forecasting. In extension, it is therefore possible to select a specification without covariates, which allows for a significant bias in the SCE.

The inclusion of covariates is not always beneficial however. For predictor weights to be correctly assigned to covariates, not all pretreatment outcome lags should be included (Kaul et al., 2022). By not including all pretreatment outcome lags, one prevents the asymptotic equivalence result in Section 3.2 from holding. Additionally, including all pretreatment outcomes in the specification may also construct an SCE that is better at capturing unobserved components effects than a specification with covariates but a lower number of pretreatment outcome variables (Botosaru & Ferman, 2019). It should be considered however, that this is only relevant for relatively large T_0 and many empirical datasets are often constrained in terms of length.

Depending on the case therefore, biases can arise due to the choice of including covariates or including all pretreatment outcome lags. To add onto this statement, there exist specification-searching opportunities in the choice to include covariates or including all pretreatment lags. The reason why this trade-off exists revolves around how weights are assigned in the SCM.

3.3.3 The influence of covariates on weights

It is often impossible for the weights to be non-negative and to sum to one (Assumption (3)) and also for a perfect balance to hold for both pretreatment outcome lags (Assumption (5)) as well as covariates (Assumption (6)). As a result, an optimization procedure is used using a data-driven method to satisfy these conditions as closely as possible. In Section 3.1, it is explained that this optimization procedure consists of two steps: one is the inner optimization (2) and the other is the outer optimization (4). The outer-optimization determines $\hat{\mathbf{V}}$, this predictor weight matrix considers individual predictors' predictive power for the outcome variable $\mathbf{Y}$. This is because some predictors might be more informative for predicting outcomes, such as the first pretreatment lag.

A crucial point to understand about how $\hat{\mathbf{V}}$ is determined comes from the paper of Kaul et al. (2022). Following the original methodology of Abadie et al. (2010) for the outer optimization (4), $\hat{\mathbf{V}}$ is determined by minimizing the MSPE of the outcome variable $\mathbf{Y}$ over the pre-intervention periods. We can partition the predictor weight matrix $\mathbf{V}$ as follows:

$$\mathbf{V} = \begin{pmatrix} \mathbf{V_Z} & 0 \\ 0 & \mathbf{V_Y} \end{pmatrix}$$

where $\mathbf{V_Z}$ contains the weights for the covariates and $\mathbf{V_Y}$ contains the weights for the pretreatment outcomes. In the outer optimization (4), $\mathbf{V_Y}$ is often more relevant as only the optimal pretreatment fit of the outcome variable is considered, not of the covariates. The degree of relevance depends on how well the included pretreatment outcome lags are able to provide a balance as in (5). Naturally, the more pretreatment outcome lags, the more likely that a perfect balance can be found. Using an optimal $\hat{\mathbf{V}}$ derived from (4), it is possible to calculate $\hat{\mathbf{W}}(\hat{\mathbf{V}})$ using the inner optimization (2). By partitioning the data matrix $\mathbf{X}$ into $\mathbf{X_1} = \begin{pmatrix} Z_1 \\ \mathbf{Y_1} \end{pmatrix}$ and $\mathbf{X_0} = \begin{pmatrix} \mathbf{Z_0} \\ \mathbf{Y_0} \end{pmatrix}$, respectively, we can reconstruct inner optimization (2) as outlined by Kaul et al. (2022), in which a weight matrix $\hat{\mathbf{W}}$ is found so that the predictors of the SCU and unit of interest are most closely aligned:

$$\sqrt{(Z_1 - \mathbf{Z_0}\mathbf{W})'\mathbf{V}_Z(Z_1 - \mathbf{Z_0}\mathbf{W}) + (\mathbf{Y}_1 - \mathbf{Y}_0\mathbf{W})'\mathbf{V_Y}(\mathbf{Y}_1 - \mathbf{Y}_0\mathbf{W})} \overset{\mathbf{W}}{\to} min \tag{10}$$

As can be seen by the first component in optimization (10), the inner optimization does take the covariates into account. However, it is the outer optimization (4) that determines $\hat{\mathbf{V}}$.

We now examine two cases which are of interest to us. In the case that we determine $\hat{\mathbf{V}}$ by using all pretreatment outcome lags, a near perfect balance for the pretreatment outcome values in outer optimization (4) is usually attained. This results in a weight matrix $\hat{\mathbf{W}}$ that is optimal in the sense that it produces the best fit for the outcome variable $\mathbf{Y}$:

$$\hat{\mathbf{W}} := argmin_{\mathbf{W}}(\mathbf{Y}_1 - \mathbf{Y}_0\mathbf{W})'(\mathbf{Y}_1 - \mathbf{Y}_0\mathbf{W})$$

The implication of this optimal weight matrix is that a balance in covariates is no longer needed

to optimize the inner optimization (2), and it is therefore optimal to set $\hat{\mathbf{V}}_{\mathbf{Z}}$ equal to zero so that optimization (10) is minimized. Therefore, (10) yields, in its limit, a solution which effectively assigns only weights to the other predictors, and not to the covariates, as the first component in (10) is effectively ignored. Given that there are multiple solutions for predictor weights $\mathbf{V}$ which yield the same optimal weight matrix $\hat{\mathbf{W}}$, researchers could be prone to selecting a $\mathbf{V}$ with positive weights for the covariates, while these still correspond to the same weight matrix $\hat{\mathbf{W}}$. This could lead to the incorrect conclusion that covariates are considered in the optimization procedure, while they are in fact ignored.

In the other case, where not all pretreatment outcomes are used, there may not be an optimal pretreatment fit in the outer optimization (4). This allows for positive covariate weights in $\mathbf{V}$ and therefore also requires the inner optimization (2) to effectively take into account the role of covariates when solving for $\hat{\mathbf{W}}$. For this $\hat{\mathbf{W}}$, the covariates will have predictive power. It should also be noted that the matrix $\mathbf{V}$ should not be chosen in an ad hoc fashion. This would assign weights to covariates which are not necessarily required to be balanced.

3.4 Summary

A summary of the main theoretical points discussed in this section is given below. Initially, the framework of the SCM of Abadie et al. (2010) was described in detail. Using this framework, the equivalence of specifications theorem found in Ferman et al. (2020) was described. Next, we discussed the consequences of omitting the assumption of a perfect balance on covariates. This requires stronger assumptions about linear independence of the effect of observed and unobserved covariates on potential outcomes, and therefore also a requirement of a longer pretreatment span T_0. Thereafter, an analysis on the covariates' ability to reduce biases was made. Covariates can reduce bias in predictions, especially when multicollinearity or nonlinearity exists. However, to reap these benefits the number of pretreatment lags included in the specification has to be reduced, which can lead to biases as well. The reason for this trade-off lies in how the assignation of weights for matrices $\mathbf{W}$ and $\mathbf{V}$ in the data-driven optimization process takes place. It is found that it is often difficult to balance both covariates and pretreatment outcomes because using specifications that include large sets of pretreatment outcomes could lead to covariates being ignored in the optimization process, causing biases.

Therefore, a researcher has to determine case-by-case what the costs and benefits are in the trade-off between balancing covariates and including all pretreatment outcomes. On one hand, the exclusion of large numbers of pretreatment outcome lags can cause biases if they are able to capture the effects of unobserved components, especially if these unobserved components are important in explaining the outcome variable. On the other, covariates may be able to reduce biases, reduce the required assumptions and are simply required in the cases of multicollinear and nonlinear effects on the outcome variable. We evaluate the influence of these theoretical findings in the Monte Carlo simulations and empirical application.

4 Data and Methodology

4.1 Test-Statistics (Replication)

To be able to quantify the prevalence of specification-searching opportunities, test statistics that describe the opportunity for specification-searching of different SCUs are required. In their paper, Abadie et al. (2015) use the root mean squared prediction error (root-MSPE) between the observed outcome of a unit and the outcome of the synthetic counterfactual. For the pre-intervention period, it is labeled the root mean squared error (root-MSE), where they generalize the unit being treated instead of assuming unit $j = 1$ is being treated:

$$root - MSE_i = \left(\frac{1}{T_0} \sum_{t=1}^{T_0} \left(Y_{i,t} - \sum_{j=1, j \neq i}^{J+1} \hat{w}_j Y_{j,t} \right)^2 \right)^{1/2} \tag{11}$$

where i is the unit of interest. In the case where they would want to measure the post-intervention period root-MSPE, they would take the average for the period $T_0 + 1, ..., T$ instead. When interpreting the post-intervention period root-MSPE, it is important to consider the pre-treatment fit of the synthetic counterfactual. If the pre-treatment fit is poor, equivalently seen in a higher root-MSE for the pre-treatment period, then a higher post-treatment root-MSPE is not indicative of a strong treatment effect. This is because the counterfactual is not representative, as stipulated by the high pre-intervention root-MSE. Ideally, a SCM case-study for a unit observing a strong treatment effect should have a small pre-intervention root-MSE and a high post-intervention root-MSPE. To combine these two statistics into one, Abadie et al. (2015) computes the ratio of the mean squared prediction errors (ratio-MSPE) as a test statistic. This statistic increases with a higher post-intervention MSPE and decreases with a higher pre-intervention MSE and thus favours a SCU with a good pre-intervention fit.

$$ratio - MSPE_i := \frac{\frac{\sum_{t=T_0+1}^{T} \left(Y_{i,t} - \sum_{j=1, j \neq i}^{J+1} \hat{w}_j Y_{j,t} \right)^2}{(T - T_0)}}{\frac{\sum_{t=1}^{T_0} \left(Y_{i,t} - \sum_{j=1, j \neq i}^{J+1} \hat{w}_j Y_{j,t} \right)^2}{T_0}} \tag{12}$$

To evaluate the probabilities of obtaining a certain treatment effect $\hat{\alpha}_{j,t}$, Ferman et al. (2020) make use of placebo studies, which are widely used when applying SCMs. More specifically, they use "in-space placebos" in which they measure the SCE for each $j \in \{1, ..., J + 1\}$ at time t. By calculating the SCE for every potential control a distribution of placebo effects can be constructed. They then test the confidence of the SCE using a p-value computation taken from Abadie et al. (2015):

$$p := \frac{\sum_{j=1}^{J+1} \mathbb{I}[ratio - MSPE_j \geq ratio - MSPE_1]}{J + 1} \tag{13}$$

where unit 1 is the treated unit. This p-value measure does not have a clear statistic interpretation, but is a good indicator of how 'unique' the treatment effect is of the treated unit. It does this by calculating the fraction of placebo effects that have a greater than or equal effect compared to the treated unit. The null hypothesis of no treatment effect is therefore rejected if

the p-value is less than a significance level which is chosen beforehand, and is often influenced by the amount of control units J. Here, the p-value is described as the probability of obtaining an estimated treatment effect as least as large as that of the unit of interest when the intervention is randomly assigned. Thus, when there is a low p-value, it is highly unlikely that the effect is not caused by the treatment and thus a significant effect can be concluded.

We know that $ratio - MSPE_j$ will remain asymptotically invariant to changes in the synthetic control specification when $T_0 \to \infty$, given that $L(s, T_0) \to \infty$ for these specifications. Thus, as long as only synthetic control specifications that satisfy these properties are used, the test decision in the placebo test is asymptotically irrelevant. This implies that in this case, the possibilities for specification searching are asymptotically irrelevant.This also holds when covariates are included and makes the SCM unique.

4.2 Monte Carlo Simulations (Replication)

In this paper, a part of the Monte Carlo simulations run in Ferman et al. (2020) are replicated. This is done to grasp the importance of the theoretical results in the previous section. The simulations should inform us whether values of T_0 used in common practical applications are large enough for the asymptotic equivalence results to hold. Additionally, insight is given into specification searching opportunities in the scenario where the condition of $L(s, T_0) \to \infty$ when $T_0 \to \infty$ does not hold.

Specifically, we test the null hypothesis of no effect, as described in the previous section, given a set of specifications, which will be explained in Section 4.2.2. This placebo test should by construction provide a rejection rate of α percent when examining one specification. This is because the null tests for an α rejection rate. In this paper we examine the probability that at least one of the specifications rejects the null hypothesis at the α-percent-significance level. This occurs when the different specifications construct inconsistent SCEs. This would result in an incorrect placebo distribution for the control units and thus a higher p-value for the placebo test, such that the null hypothesis for that specification is rejected. If all specifications lead to the same SCE, then the probability to find a specification which rejects the null hypothesis of no effect will be close to α. Alternatively, if S specifications lead to S widely varying estimators, the probability of rejecting the null hypothesis for at least one specification will be large.

4.2.1 Data Generating Processes

In the Monte Carlo simulation, we generate 200 data sets using a linear factor model data generating process (DGP) where units are divided into groups that each follow different stationary time trends.

$$Y_{j,t}^0 = \delta_t + \lambda_t^k + \epsilon_{j,t} \tag{14}$$

For this simulation we use $J + 1 = 20$ units and $k = 1, ..., K$ trends indicating which trend λ_t^k is being followed, where λ_t^k is normally distributed following an AR(1) process with a 0.5 autocorrelation parameter. For $j = 1, 2$, $k = 1$, such that $j = 1, 2$ follow trend λ_t^1, for $j = 3, 4$, $k = 2$ and so on. Additionally, $\delta_t \sim N(0, 1)$ and $\epsilon_{j,t} \sim N(0, 0.1)$.

For each simulation, there are 10 post-treatment periods so that $T - T_0 = 10$. To understand

the degree for specification searching opportunities in published papers mentioned in Table 1, we test with pre-intervention periods $T_0 \in \{12, 32\}$. To test for the presence of the asymptotic equivalence result in Section 3.2, we use larger pre-intervention periods, $T_0 \in \{100, 400\}$. In the simulation, there is no treatment effect so that $Y_{j,t} = Y^0_{j,t} = Y^1_{j,t}$ for all t. This is done because we wish to measure the probability of rejecting the null hypothesis of no effect when there is in fact no treatment taking place. Thus, when the term pre-intervention is used, it is referring to a hypothetical intervention period $t = 0$ where no intervention actually takes place. To repeat the Monte Carlo Simulation for the case where time-invariant covariates are included and influence the outcome variable, we use a slightly altered DGP:

$$Y^0_{j,t} = \delta_t + \lambda^k_t + \theta_t Z_i + \epsilon_{j,t} \tag{15}$$

where Z_i for i=1,...,10 and $Z_i = 0$ for $i = 11, ..., 20$. As with the standard DGP, $K = 10$ and λ^k_t is an AR(1) process with 0.5 autocorrelation. Additionally, $\delta_t \sim N(0, 1)$, $\epsilon_{j,t} \sim N(0, 0.1)$ and $\theta_t \sim N(0, 1)$. Here, the first 10 control units are affected by the covariates by a normally distributed random process.

4.2.2 Specifications

The original seven specifications used in the placebo tests used in Ferman et al. (2020) differ in terms of the number of pre-treatment outcome values used, but also the choice of pre-treatment outcome values. The first 5 specifications satisfy the condition that $L(s, T_0) \to \infty$ when $T_0 \to \infty$, specifications 6 and 7 do not. This means that only the first 5 specifications will be eligible to achieve the asymptotic equivalence result. Therefore, the accumulated results for both sets of specifications $S = 1, .., 7$ and $S' = 1, ..., 5$ will be presented. Initially, no covariates are included in the tests. The elements of each specification are described below:

1. All pre-intervention outcome values: $\mathbf{X}_j = [Y_{j,1}, ..., Y_{j,T_0}]'$
2. The most recent three-fourths pre-intervention outcome values: $\mathbf{X}_j = \left[Y_{j,1}, ..., Y_{j,\frac{3T_0}{4}}\right]'$
3. The most recent half of pre-intervention outcome values: $\mathbf{X}_j = \left[Y_{j,1}, ..., Y_{j,\frac{T_0}{2}}\right]'$
4. All odd pre-treatment outcome values: $\mathbf{X}_j = \left[Y_{j,1}, Y_{j,3}, ..., Y_{j,(T_0-3)}, Y_{j,(T_0-1)}\right]'$
5. All even pre-treatment outcome values: $\mathbf{X}_j = \left[Y_{j,2}, Y_{j,4}, ..., Y_{j,(T_0-2)}, Y_{j,T_0}\right]'$
6. The mean of all pre-intervention outcomes: $\mathbf{X}_j = \left[\frac{\sum_{t=1}^{T_0} Y_{j,t}}{T_0}\right]$
7. The first, middle and last pre-intervention outcome values $\mathbf{X}_j = \mathbf{X}_j = \left[Y_{j,1}, Y_{j,\frac{T_0}{2}}, Y_{j,T_0}\right]'$

4.2.3 Testing procedure (Also extension)

Ferman et al. (2020) use this set of specifications to measure the possibility for specification searching. These specifications will henceforth be referred to as the 'standard' specifications. A

placebo test is carried out for each specification at the 5 percent significance level; if the ratio-MSPE of the treated unit is found to be the largest among the 20 units, the null hypothesis of no treatment effect is rejected. When this is the case, the specification of interest indicates the presence of a treatment effect despite no treatment effect actually existing. We are interested in the cumulative case; thus where at least one specification in the set above indicates a treatment effect for the placebo unit, where there is in fact no treatment effect. We repeat this process for each of the 20 treated units, and repeat this simulation 50 times for each $T_0 = \{12, 32, 100, 400\}$. As a result, we are able to find the empirical probability of rejecting the null of no treatment effect in at least one specification, as in Equation (13). In other words, the values shown in Table 2 in the results section represent the probability of finding a treatment effect even when this is not the case, and thus the possibility for specification searching. In the case that all specifications consistently construct identical SCUs, the probability that the 'treated' unit will have the largest ratio-MSPE for one of the specifications is very small, namely 5 percent. In the case that all specification construct a different SCE each, it is more likely that the ratio-MSPE of the 'treated' unit will be the largest of all units in at least one of the 7 specifications. The specific testing procedure above refers to the 'standard DGP ; standard specification' case.

The difference in specifications for the time invariant case is minimal but important. In the time-invariant case we use each of the seven specifications mentioned above but we also include variable Z_i as an economic predictor. This is influential when the DGP including covariates is used in the Monte Carlo simulation. This is because the 7 specifications listed above are able to use the information given by the predictor Z_i. The availability of predictor Z_i in the specification helps identify whether the unit of interest is likely to be affected by covariates or not. When the treatment 'occurs', this information will determine if the SCU will experience an additional effect of degree θ_t or not. Henceforth, these specifications will be referred to as the 'covariate specifications'. The same testing procedure is used as for the 'standard DGP ; standard specifications' case, only now with the covariate specifications on a DGP incl. covariates. This testing procedure is labeled the 'DGP incl. covariates ; covariate specifications' case.

The comparison of these two cases allows for insights concerning the difference in specification searching opportunities when the 'correct' set of specifications is applied to to its corresponding DGP. Such insights could be that one case has reduced specification-searching opportunities for shorter pretreatment periods T_0. Nevertheless, it is also important to compare the cases where these two sets of specifications are compared for the same DGP. In effect, this will allow us to see the consequences concerning specification-searching opportunities of including covariates in a given specification. This brings us to a third testing procedure: the 'DGP incl. covariates ; standard specifications' case. In this case, the set of covariate specifications have more information to identify the effects of covariates, and are thus able to detect the effect θ_t after the treatment period. On the other hand, the set of standard specifications do not have this information. Finally, a fourth testing procedure is the 'standard DGP; convariate specifications' case. This case is used to determine if irrelevant covariates cause additional bias, as they do not carry any information.

In addition to the comparison of specification-searching opportunities for these three different cases, the root-MSPE (Equation (11)) is also calculated for each individual specification. This

will show which specifications produce the most accurate SCEs. We evaluate the differences for root-MSPEs instead of ratio-MSPEs because comparing ratio-MSPEs would only allow us to measure the differences in pretreatment to post-treatment accuracy ratios, whereas root-MSPEs allow us to identify the best-performing specifications. Essentially, this will allow us to further dissect why specification-searching opportunities are more or less present in certain cases, but they will also allow us to see which specific specifications perform best in different cases.

To complement the information provided by the root-MSPEs in different cases, the predictor weight matrix $\hat{\mathbf{V}}$ will be used to identify the average fraction of weights that is assigned to the covariates for each experiment. This is done by first calculating the fraction of weights in matrix $\hat{\mathbf{V}}$ assigned to covariates, repeated for the time periods $T_0 = \{12, 32\}$ and for the seven covariate specifications. The average of all these fractions in the Monte Carlo simulation is taken to give an indication of how much importance is given to covariates in each of the experiments.

Finally, the above tests for specification-searching opportunities and root-MSPEs will be repeated for the case of $T_0 = 4$. This is done because the results of Kaul et al. (2022) and Botosaru and Ferman (2019) both imply that covariates have a more important part to play when the pretreatment periods T_0 are shorter. $T_0 = 4$ is chosen as any shorter pretreatment period would result in specifications 1-7 overlapping. These results as a whole will complement the literature review and theoretical background in allowing for a more in-depth examination of the role of covariates in the simulations of Ferman et al. (2020).

5 Results and Discussion

5.1 Monte Carlo Simulation

5.1.1 Specification-searching opportunities (Replication and Extension)

We first investigate the existence of the asymptotic equivalence theorem for specification-searching opportunities, both for the standard and time-invariant case. The probabilities of rejecting the null hypothesis of no treatment effect for the standard case, at the 5% and 10% significance levels, are given in Table 2 for different numbers of pre-treatment outcome lags. This is done using the DGP in Equation (14).

Following the asymptotic equivalence theorem presented in section 3.2, the possibilities for specification searching should diminish as the length of the pre-treatment period T_0 increases. This result is somewhat evident based on the results in Table 2, Panel A; for $T_0 = 12$, the probabilities at the 5% and 10% level are 15.3% and 26.0% respectively; for $T_0 = 32$, the probabilities are 14.5% and 25.0%; for $T_0 = 100$, the probabilities are 14.0% and 25.1%; and for $T_0 = 400$, the probabilities are 13.7% and 25.3%. The results are all above the 5% level that would be expected for the theoretical case where there are no specification searching opportunities. Therefore, we find that specification searching opportunities are significant when $T_0 = 12$, which is an empirically common number of pre-treatment outcome lags. Even for empirically uncommon levels such as $T_0 = 400$, the results indicate that specification-searching opportunities are still prevalent.

From the asymptotic equivalence theorem in Section 3.2 it can be derived that those specifications that do not meet the conditions for this equivalence result are most likely responsible

Table 2: Specification searching opportunities for the 'standard DGP; standard specifications' case

	5% test	10% test
Panel A: *Specifications 1 to 7*		
$T_0 = 12$	0.153	0.260
	(0.012)	(0.015)
$T_0 = 32$	0.145	0.250
	(0.011)	(0.014)
$T_0 = 100$	0.140	0.251
	(0.011)	(0.014)
$T_0 = 400$	0.137	0.253
	(0.011)	(0.014)
Panel B: *Specifications 1 to 5*		
$T_0 = 12$	0.108	0.194
	(0.009)	(0.012)
$T_0 = 32$	0.099	0.179
	(0.009)	(0.012)
$T_0 = 100$	0.084	0.150
	(0.009)	(0.012)
$T_0 = 400$	0.079	0.140
	(0.009)	(0.012)

Notes: This table presents the probabilities of rejecting the null hypothesis of no treatment effect in at least one specification for the given set of specifications at the 5 and 10 percent significance levels. The results are based on 3,900 observations.

for the most disparate SCEs. Table 2 Panel B shows specification-searching opportunities when specifications 6 and 7 are not included in the set of specifications. Instead, all specifications in this set follow $L(s, T_0) \to \infty$ when $T_0 \to \infty$. Excluding specifications 6 and 7, the results are as follows: for $T_0 = 12$, the probability at the 5% and 10% level are 10.8% and 19.4% respectively; for $T_0 = 32$, the probabilities are 9.9% and 17.9%; for $T_0 = 100$, the probabilities are 8.4% and 15.0%; and for $T_0 = 400$, the probabilities are 7.9% and 14.0%. These results provide evidence that when a specification is used which tends towards $L(s, T_0) \to \infty$ for which the asymptotic equivalence theorem holds, large values of T_0 do indeed attenuate the specification-searching problem in SCMs. Nevertheless, specification searching opportunities are not completely eradicated, especially as even for cases where T_0 is extremely large, as with $T_0 = 400$, there is still some room for specification-searching. The disparity between the results in Panel A and Panel B is because as $L(s, T_0) \to \infty$ for specifications 1 to 5, the weights assigned to each SCE are more correctly allocated. For specifications 6 and 7, this is not the case, and as a result, specifications 6 and 7 allow for divergent SCEs, causing higher p-values.

Table 3 shows the opportunities for specification-searching opportunities for the 'DGP incl. covariates; covariate specifications' case. It can be seen that the p-value also reduces as $T \to \infty$. The p-values also decrease more substantially when specifications 6 and 7 are excluded. This difference between the p-values for $T_0 = 12$ and $T_0 = 400$ for Panel A is 0.023 whereas for Panel B it is 0.041. The reduction in p-values as T_0 increases is twice as large, whereas the difference in the number of specifications in each panel is only approximately 30%. This indicates that, as theoretically founded in Section 3.2, the asymptotic equivalence theorem is evident from the results as specifications 6 and 7 do not follow the asymptotic equivalence theorem.

Table 3: Specification searching opportunities for the 'DGP incl. covariates; covariate specifications' case

	5% test	10% test
Panel A: *Specifications 1 to 7*		
$T_0 = 12$	0.153	0.243
	(0.011)	(0.013)
$T_0 = 32$	0.129	0.231
	(0.011)	(0.013)
$T_0 = 100$	0.140	0.245
	(0.011)	(0.013)
$T_0 = 400$	0.130	0.229
	(0.011)	(0.013)
Panel B: *Specifications 1 to 5*		
$T_0 = 12$	0.119	0.202
	(0.009)	(0.012)
$T_0 = 32$	0.088	0.170
	(0.009)	(0.012)
$T_0 = 100$	0.089	0.166
	(0.009)	(0.012)
$T_0 = 400$	0.078	0.135
	(0.009)	(0.012)

Notes: This table presents the probabilities of rejecting the null hypothesis of no treatment effect in at least one specification for the given set of specifications at the 5 and 10 percent significance levels. The results are based on 3,900 observations.

These results for the 'standard DGP; standard specifications' and 'DGP incl. covariates; covariate specifications' cases align closely with the original results found in the paper by Ferman et al. (2020). While some results deviate, they are always within one standard deviation of the value found in the paper of Ferman et al. (2020). It is notable that in the reproduced findings in Tables 2 and 3, the standard deviations are larger than in the original paper, often around three to four times as large. This is due to the reduced number of observations used in the reproduced findings, 3,900 to be exact, while in the original paper there are 10,000. These result in a higher and lower standard deviation, respectively. In the reproduced results of Table 3, there are some simulations which yield a p-value that is only just within, or just outside, one standard deviation of the original p-values found in Ferman et al. (2020). This is the case for both panels and significance levels, but only for the pretreatment periods $T_0 = \{12, 32\}$. This bias is likely the result of a reduced number of observations used. As T_0 increases to 100 and 400, there is less bias in the construction of SCUs, due to a better balance being made possible, and thus less room for deviation. This explains why the p-values for higher T_0 deviate less from that of Ferman et al. (2020).

The results above allow for a main conclusion to be drawn from this Monte Carlo simulation: when T_0 is large enough and the asymptotic equivalence conditions ($L(s, T_0) \to \infty$ when $T_0 \to \infty$) hold for the set of specifications, the opportunities for specification-searching decrease substantially. However, in empirical settings these conditions often do not hold and thus specification-searching remains a relevant problem.

Additional deductions on specification-searching opportunities using the results are described below. In Appendix B, it can be seen that when specification searching opportunities are in-

vestigated for the 'DGP incl. covariates: standard specification' and 'standard DGP; covariate specification' case, the p-values do not differ much from the results in Tables 2 and 3. There are two useful results to note which are beneficial for later investigations. One is that when using a standard DGP, specification searching opportunities slightly decrease when covariate specifications are used. Second, when using a DGP including covariates, there are less specification searching opportunities when using standard specifications 1 to 5 for $T_0 = 12$.

In Appendix C, Table 11 shows that specifications 6 and 7 most frequently construct SCUs that produce different SCEs to specification 1 for the covariate case. It also shows that specification 6 tends to deviate from specification 1 more frequently than specification 7, and thus provides significant results most often. This suggests that, in our experiments, using the mean of pretreatment outcome lags as a specification along with covariates, provides opportunities for specification searching. Therefore, caution should be taken when choosing to use specification 6.

5.1.2 Average Root-MSPEs (Extension)

Table 4: Average Root-MSPE for different specifications and DGPs

	Specifications						
	1	2	3	4	5	6	7
Panel A: *Standard DGP; standard specifications*							
$T_0 = 12$	0.156	0.172	0.185	0.167	0.167	0.314	0.224
$T_0 = 32$	0.142	0.147	0.155	0.144	0.144	0.314	0.222
$T_0 = 100$	0.137	0.139	0.141	0.138	0.138	0.315	0.220
$T_0 = 400$	0.136	0.136	0.137	0.136	0.136	0.314	0.219
Panel B: *DGP incl. covariates; covariate specifications*							
$T_0 = 12$	0.155	0.171	0.178	0.158	0.158	0.272	0.182
$T_0 = 32$	0.142	0.147	0.155	0.145	1.145	0.275	0.178
$T_0 = 100$	0.138	0.139	0.141	0.139	0.138	0.269	0.175
$T_0 = 400$	0.137	0.137	0.138	0.137	0.137	0.270	0.175
Panel C: *DGP incl. covariates; standard specifications*							
$T_0 = 12$	0.155	0.173	0.186	0.166	0.166	0.340	0.234
$T_0 = 32$	0.141	0.145	0.152	0.144	0.143	0.346	0.237
Panel D: *Standard DGP; covariate specifications*							
$T_0 = 12$	0.157	0.174	0.176	0.159	0.158	0.271	0.180
$T_0 = 32$	0.141	0.145	0.152	0.144	0.143	0.270	0.176

Notes: This table presents the average root-MSPEs for different combinations of specifications, DGPs and T_0. The results are based on 12,000 observations.

The influence of the asymptotic equivalence theorem is evident from the root-MSPE values as well. Using Table 4, it can be seen that the trend is that as T_0 increases, the SCUs become more accurate and as a result the root-MSPEs reduce, this is evident from all four panels. However, this is not the case for specifications 6 and 7, which retain the same values as T_0 increases. For specifications 6 and 7, $L(s, T_0) \to \infty$ does not hold as $T_0 \to \infty$, and therefore specifications 6 and 7 do not become more accurate. Less disparate SCUs are often implied when root-MSPEs decrease, and as a result, opportunities for specification-searching can be attributed to specifications 6 and 7. This is supported by the findings of Appendix C as well. These findings therefore support the asymptotic equivalence theorem of Ferman et al. (2020), as larger

root-MSPEs indirectly imply more disparate SCEs, and more opportunities for specification-searching.

This table also identifies many of the theoretical findings of Section 3. Firstly, it is evident that covariates can be more relevant for certain specifications. In Panel A, there is a large difference between the root-MSPEs of specifications 1 to 5 (≈ 0.170 for $T_0 = 12$), and 7 (≥ 0.220) for all periods T_0. In Panel B however, especially for smaller values of T_0, the difference between root-MSPEs of 1 to 5 (≈ 0.170 for $T_0 = 12$) and 7 (≥ 0.175) is much smaller. Thus, when covariates are included in a model with low T_0, given that they are relevant, the performance of specification 7 improves relatively more than specifications 1 to 5.

In Section 3, it was established that in cases where there are less pretreatment periods, the inclusion of covariates can either cause or eliminate biases. Covariates do eliminate some bias for the case when $T_0 = 12$ when comparing Panel B to Panel C. For $T_0 = 12$, specification 1 of Panel C performs identically to its counterpart in Panel B (0.155), specifications 2 to 5 obtain slightly higher root-MSPEs, and specifications 6 and 7 obtain dramatically higher root-MSPEs. This result highlights that covariates can reduce biases when they are relevant. Furthermore, it emphasizes the importance that covariates have when $L(s, T_0)$ is low. With low $L(s, T_0)$, the weights are likely to be more equally divided between the covariates and pretreatment outcome lags. The inclusion of covariates therefore helps to eliminate bias, as found by Botosaru and Ferman (2019). At the same time, irrelevant covariates may invite multicollinearity among the predictors, causing extra biases. However, the results in Panel D show that including irrelevant covariates in the specifications actually improve the forecasts and reduce average root-MSPEs. This is a surprising result as it contradicts Botosaru and Ferman (2019), but could potentially be attributed to the fact that the covariates in the specification may still provide some information in the SCM. Possibly, the covariates may act as indicators for which unit j is being addressed, as the covariate vector Z_j takes value 1 for $j = 1, ..., 10$ and 0 for $j = 11, ..., 20$.

Section 3 also discussed the increasing irrelevance of covariates as T_0 increases. One of the reasons is because a balance on a large number of pretreatment outcome lags implies an approximate balance on covariates: for $T_0 = 400$, approximately the same root-MSPEs are found for specifications 1 to 5 in Panel A and Panel B. This suggests that the distortionary effect of the covariates in the DGP are overcome even for specifications that barely assign weights to the covariates such as specification 1. Another reason is the different assignation of weights when all covariates are included, the indifference of the results for $T_0 = 32$ in Panel C and Panel B also suggests that as $L(s, T_0)$ increases, covariates become redundant. In fact, the inclusion of covariates at large T_0 may potentially even assign unnecessary weights to the covariate component in the inner optimization (10), causing a bias due to irrelevant covariates as stated in Botosaru and Ferman (2019). The slightly better root-MSPEs of Panel C for $T_0 = 32$ compared to Panel B underscore the dangers of including covariates for large T_0. Thus, the use of more pretreatment outcomes diminishes the relevance and explanatory power of the covariates.

It can also be identified that some specifications have a better forecasting accuracy, and that this is not attributable to differences in T_0 or covariates. A common trend is that specifications with the same number of $L(s, T_0)$, specifications 3, 4 and 5, differ in predictive performance depending on the specific pretreatment outcome lags included. From Table 4 it can be seen that

specifications with lags spread over T_0 (i.e. 4, 5) perform better than those only including the first half (i.e. 3). This suggests specifications with more spread capture more information and have better forecasting accuracy.

The results in Table 4 do at first sight provide some inconsistencies with the results for specification-searching opportunities. As stated before, when using a DGP including covariates, standard specifications 1 to 5 (Panel C) for $T_0 = 12$ provide less specification searching opportunities (Table 9) than covariate specifications (Table 3). This occurs despite the average root-MSPE being higher for the standard specification case, which would lead one to expect more disparate SCUs and thus more specification searching opportunities. However, this could be explained by standard specifications producing more similar SCUs. It was previously mentioned that, for the standard DGP cases, specification-searching opportunities are lower at $T_0 = 32$ when using covariate specifications (Table 10) than when using standard specifications (Table 2). By looking at the average root-MSPEs in Table 4, it can be seen that this occurs despite the average root-MSPEs for specifications 1 to 5 being the same for both types of specifications. The disparity is likely then caused by specifications 6 and 7 which construct more disparate SCUs when the standard specifications are used.

Table 5: Fraction of weights in V-matrix assigned to covariates

	Specifications						
	1	2	3	4	5	6	7
Panel A: *DGP incl. covariates; covariate specifications*							
$T_0 = 12$	0.022	0.124	0.210	0.224	0.233	0.487	0.298
$T_0 = 32$	0.029	0.036	0.106	0.117	0.111	0.411	0.325
Panel B: *Standard DGP; covariate specifications*							
$T_0 = 12$	0.020	0.088	0.177	0.183	0.182	0.547	0.247
$T_0 = 32$	0.028	0.040	0.085	0.082	0.086	0.483	0.290

Notes: This table presents the average fraction of weights assigned to covariates per specification, and T_0. The results are based on 400 iterations of the Monte Carlo simulation.

Table 5 indicates the fraction of weights in the V-matrix assigned to covariates. This is only done for the cases where covariate specifications are used. Panel A indicates that specification 1 barely assigns weights to covariates, supporting the findings of Kaul et al. (2022) that covariates are rendered irrelevant when all pretreatment outcomes are balanced. On the other hand, for specifications 6 and 7, the covariates are much more important indicating that for low $L(s, T_0)$ covariates have relatively more explanatory power. This is also in line with the previous findings of this paper. Finally, as T_0 increases from 12 to 32, covariates are assigned less weights in the V matrix. This is in line with the theory of Kaul et al. (2022) that the pretreatment outcome balance is prioritized in the outer optimization. As a result, covariates attain less weights. The results in this table eliminate the possibility of overlooking covariates, and allow for the differences in root-MSPEs found in Table 4 to be attributed to the covariates.

Panel B confirms that covariates are still influential in determining outcomes even when they are supposedly uninformative. This is once again indicative of the covariates maintaining explanatory power despite the DGP not including covariate effects.

5.1.3 The case for 4 pretreatment periods (Extension)

Table 6: Specification searching using a stationary model with time-invariant covariates for $T_0 = 4$

	5% test	10% test
Specifications 1 to 7		
Stand. DGP; Stand. spec.	0.178	0.288
	(0.011)	(0.014)
DGP incl. cov.; Cov. spec.	0.171	0.274
	(0.011)	(0.014)
DGP incl. cov; Stand. spec.	0.176	0.298
	(0.011)	(0.014)

Notes: This table presents the probabilities of rejecting the null hypothesis of no treatment effect in at least one specification for the given set of specifications at the 5 and 10 percent significance levels. The results are based on 3,000 observations.

Table 7: Average Root-MSPE for standard specifications and standard DGP, for $T_0 = 4$

	Specifications						
	1	2	3	4	5	6	7
Standard DGP; Standard specifications	0.229	0.264	0.288	0.286	0.275	0.316	0.241
DGP incl. covariates; Covariate specification	0.236	0.238	0.251	0.252	0.246	0.282	0.226
DGP incl. covariates ; Standard specification	0.236	0.267	0.308	0.304	0.294	0.343	0.253

Notes: This table presents the average root-MSPEs for different combinations of specifications and DGPs for $T_0 = 4$. The results are based on 3,000 observations.

The results in Tables 6 and 7 help explain the case for covariates when T_0 and therefore also $L(s, T_0)$ is small. Table 6 confirms that specification-searching opportunities are higher for larger T_0 as is suggested by Ferman et al. (2020). It can be seen however, that the 'DGP incl. covariates; covariate specification' case has slightly reduced specification searching opportunities when $T_0 = 4$. This suggests covariates are more important in the construction of similar SCUs, likely because they hold more explanatory power. This suggests covariates should be included to prevent specification searching when T_0 is small. Additionally, looking at Table 7 it can be seen that in the cases where DGP incl. covariates, covariate specifications reduce average root-MSPEs for all specifications compared to the standard specifications. Moreover, there is evidence that it is preferable to not include all pretreatment outcomes in the specification alongside covariates as these render covariates irrelevant as found by Kaul et al. (2022). As seen by the result for specification 7 in the 'DGP incl. covariates; covariate case', including covariates with a specification that does not use all pretreatment outcomes may provide better results (0.226) than when using covariates alongside all pretreatment outcomes (0.236). As T_0 is small, a balance on the pretreatment outcomes does not infer an approximate balance on the covariates as is the case for large T_0. These results highlight how covariates play a more important role when T_0 is small, as is often the case in empirical research.

5.1.4 Discussion (Extension)

Some limitations are discussed here. It is important to address that the covariates in the Monte Carlo simulation have effects of a certain magnitude θ_t. Notably, in the simulation carried out by Ferman et al. (2020), $\mathbb{E}(\theta_t) = 0$ and $Var(\theta_t) = 1$. In reality, covariates may have a stronger or weaker expected effect, or be more or less volatile, causing covariates to become more or less important in the construction of SCUs. At the same time, it is usually not the case that when covariates are applied, that just one covariate is present. When multiple covariates are used, there is also more room for potential multicollinearity that needs to be taken account of. Additionally, according to Botosaru and Ferman (2019), when covariates are in fact irrelevant for the DGP at hand, we would expect biases to increase and specification-searching opportunities to increase, unlike the results of this paper.

It should also be pointed out that due to time constraints some results were not able to be fully completed. The fact that individual simulations were run a total of 100 times means that there exists some uncertainty in the results. Therefore, further research should attempt to replicate these results for a higher number of simulations. Time constraints also prevented certain further investigations. The results above point to some interesting areas for further research. One such area is the predictor weight matrix $\mathbf{V}$, by manually altering the composition of this matrix a more comprehensive picture of the importance of covariates could be achieved. Additionally, an investigation into different DGPs and the existence of specification-searching opportunities subject to these differences could provide more accurate recommendations of which specification-searching opportunities are prevalent in which research applications.

Another limitation is that we have only investigated time-invariant covariates. In cases where covariates have nonlinear effects on the outcome, which can often be the case in economics, a balance on covariates over time might be required as well. This is an interesting topic for future research.

5.1.5 Recommendation (Replication)

To minimize the opportunities for specification-searching, Ferman et al. (2020) provide one main recommendation. They suggest that the specification in which all pre-treatment outcome values are included should be the main specification used in SCMs. This is because this specification is able to minimize the pre-treatment root-MSE and therefore is best suited to provide the most reliable SCE. By reaching a consensus on which specification should be used as a standard, this removes the possibility for arbitrary decisions on the choice of pre-treatment outcome lags that allow for specification-searching.

Additionally, researchers should only deviate from this specification when the researcher believes that a balanced set of covariates is also required. Ferman et al. (2020) recommend that in this case not all pre-treatment outcome lags should be included due to the findings of Kaul et al. (2022). Instead, only specifications that meet the conditions outlined in section 3.2 should be used in this case. They also suggest presenting several different specifications in this case to provide a reliable result, and that specification 1 in section 4.2.2 is always included as a benchmark. However, when including multiple specifications, it is difficult to construct a valid hypothesis test. This is because including a test for which all specifications reject the null would

be overly conservative, and a test rejecting for at least one would lead to over-rejecting the null. Combining all the individual test-statistics also does not come with an obvious point-estimator (Imbens & Rubin, 2015). To solve this issue, two solutions are provided. First, an inference procedure using new test statistics can be used. To do this, we can use a weighted average of the point-estimator which corresponds to the weighted average of the test-statistics used as a test function. Another method to do this is to invert a combination of test-statistics to compute a confidence set containing all treatment-effect functions within a pre-specified class. A second solution is to consider a criterion for choosing a specification. If this criterion is followed, there is no room for subjective choice of specifications by the researcher. This would limit the room for specification searching. Limiting the freedom of researchers in terms of choice of pretreatment outcome values and covariates consequently limits specification-searching opportunities. However, it should be noted that researchers could still find specification-searching opportunities in other dimensions, such as deciding when the treatment takes place.

5.1.6 Revised Recommendations using Covariate Analysis (Extension)

While Ferman et al. (2020) address that extra steps should be taken when including covariates, their explanation is somewhat reductive and does not appreciate the full importance of covariates. In their table of published articles using SCM (Table 1), over three quarters of the papers used covariates. Moreover, those papers using covariates in their specifications differ drastically in the pretreatment outcome lags or linear combinations thereof used. Several use the mean, several use a short set of pretreatment outcome lags and some use all lags. To further explain the consequences of these differences in specifications and reduce the possibilities for specification-searching, some supplementary recommendations concerning covariates are included here.

When researchers design their SCM, they must first decide on whether to include covariates. This should be done based on several conditions. Firstly, how many pretreatment periods T_0 are available to the researcher and how explanatory are pretreatment outcomes relative to covariates? If T_0 is extensive and the application of the research concerns outcome variables that are most accurately predicted by its lags, such as stock prices or inflation, one should opt for the specification including all pretreatment outcome lags. In this case, as we have seen from the results and theory of Botosaru and Ferman (2019), an approximate balance on the covariates is implied, and balancing for covariates may impose unnecessary constraints in the SCM. Additionally, balancing on all pretreatment outcomes may actually allow the effects of unobserved covariates and components to be better incorporated in the SCM. This could be useful in applications where observed covariates are limited in their explanatory power or when unobserved components are deemed influential. It is not always beneficial to include all pretreatment outcomes in the specification, even when T_0 is large. In the case where covariates are multicollinear or when covariates have a nonlinear effect on the outcome, the approximate balance of covariates is not implied by a perfect balance in pretreatment outcomes. In this case, it is essential that a reduced number of pretreatment outcome lags are included in the specification alongside covariates. This ensures that covariates are balanced as well and thus enables us to derive bounds on the bias of the estimator. When T_0 is small however, as is the case for many empirical applications, it is often beneficial to include covariates. Additionally,

the results on the predictor weight $\mathbf{V}$ matrix, and the results of Kaul et al. (2022) indicate that also with low values of T_0, not all pretreatment outcome values should be used if covariates hold substantial explanatory power. This is because covariates have relatively more explanatory power at low T_0, and there are not enough pretreatment outcome lags in the specification to assure an approximate balance on covariates, thus allowing for more bias as indicated by Botosaru and Ferman (2019).

If the decision is made to include covariates, the next consideration is which covariates to include and which other predictors should be included alongside. Covariates that should be included are those with significant explanatory power in that application according to theory or practical experiments as in Gilchrist et al. (2023). Care should be taken to exclude irrelevant covariates as this would cause extra biases as stipulated by Botosaru and Ferman (2019). Using the results of this paper, it is possible to narrow-down the choice for predictors alongside covariates. This should be a small sample spread out over the range of T_0. This small sample is found to be more informative than simply using the mean as a predictor (Table 4), while still assigning a high proportion of weights to the covariates, as seen by the results in Table 5. As T_0 decreases, the role of covariates increases.

Ultimately however, it is the decision to include or exclude covariates that allow for the most specification-searching opportunities, especially when T_0 is small. This is evident from the results of this paper, but also from the paper of Kaul et al. (2022), who reassess the findings of Billmeier and Nannicini (2013), uncovering substantial differences in predictions when correctly incorporating covariates in the SCM. Therefore, if uncertain, placebo tests should be carried out for different validation periods using both specifications without covariates and specifications correctly incorporating covariates. Thus, to eliminate the possibilities for specification-searching within a covariate context, researchers should explain why they opted for including or excluding covariates and additionally, if covariates are included, how they ensured covariates were included in the SCM in the correct fashion. This should be done using insights from this paper as well as that of Botosaru and Ferman (2019) and Kaul et al. (2022).

6 Empirical Application (Replication and Extension)

The empirical application for which we measure the possibilities for specification searching revolves around the German Reunification of 1991 and its effect on average GDP per capita. This case is taken as it is a common application used in research concerning SCMs and is extensively discussed in the paper by Abadie et al. (2015) as well as Ferman et al. (2020). The data set runs from 1960-2003. As the Reunification (the treatment) took place in 1991, the pre-treatment period spans from 1960-1990 and the post-treatment period spans from 1991-2003. The pre-treatment period is also split up into a training period of 1971-1980 and a validating period of 1981-1990. The donor pool of control units include 16 Organisation for Economic Co-operation and Development (OECD) countries.

As in Section 4.2, different specifications are used to measure the effect of a treatment at the 10% level on average GDP per capita. Here, 14 specifications are used, the same seven as in Section 4.2.2, with an additional set of these seven specifications that also include a set of 5 covariates. These covariates are trade openness, inflation rate, industry share, schooling, and

investment rate. We present the p-value results in Table 8 below.

Table 8: Specification searching using a stationary model

	Standard	Covariates
Specifications 1 to 7		
1: *All values*	0.059	0.059
2: $\frac{3}{4}$ *of all values*	0.059	0.118
3: $\frac{1}{2}$ *of all values*	0.118	0.059
4: *Odd values*	0.059	0.059
5: *Even values*	0.118	0.059
6: *Mean*	0.588	0.059
7: *Three values*	0.353	0.059

Notes: P-values of rejecting the null hypothesis of no treatment effect, for each type of specification, both for those without covariates and those with.

First and foremost, it is evident from the table above that the researcher would be able to pick a specification for which there is a significant result. Nine specifications are equal to $0.059 \approx \frac{1}{17}$, meaning there is a significant treatment effect, while four specifications are larger, which would result in a rejection of the null hypothesis of a significant treatment effect. These contradictory conclusions are due to the specification-searching issue.

Following the recommendations of Ferman et al. (2020) would lead us to two different strategies for choosing specifications. One strategy, if we believe covariates to be uninformative, is to choose standard specification 1 containing all pre-treatment outcome lags. On the other hand, if we believe that covariates are informative, we look at specifications 1 to 5 containing covariates according to Ferman et al. (2020). There is not a strong consensus due to specification 2, and thus Ferman et al. (2020) suggest using a test statistic which combines the test statistics of specifications 1 to 5. This can be done using a mean of the root-MSPE test statistic which is equal to 0.059, indicating a statistically significant impact on West-Germany's per-capita GDP. Alternatively, a confidence set can be computed, which provides a similar conclusion. These findings therefore strongly suggest that there is a significant and negative treatment effect in the long run.

Following the revised recommendations of this paper however, would suggest that covariates are highly explanatory in this application. This is likely due to the nonlinear relationships existing between covariates and outcomes. As can be seen from Table 8, covariates specifications are more conclusive and coherent than standard specifications. The only anomaly is specification 2. Using the information of this paper, this can potentially be attributed to not assigning enough weights to the covariates due to the inclusion of too many pretreatment outcomes, but at the same time not including enough pretreatment outcomes to imply an approximate balance on covariates. This is not an issue for the standard case as the covariate component of weight matrix $\mathbf{W}$ is likely still given some weight and therefore the inclusion of covariates confounds the SCM. Instead, using standard specification 1 and covariate specifications 7, as well as 4 or 5, could help gain a more complete picture. By then analyzing root-MSPEs for placebo tests, an optimal specification can be chosen. An average of these values indicates an identical conclusion as when using the recommendations of Ferman et al. (2020).

7 Conclusion

This paper has explored the potential of covariates to allow for specification-searching opportunities. It has done this by incorporating elements and theoretical as well as empirical findings from Botosaru and Ferman (2019) and Kaul et al. (2022) into the framework Ferman et al. (2020). This paper finds that the decision to include covariates and in which manner they should be included should be handled on a case-by-case basis. This is because there are benefits and drawbacks both when including covariates and when excluding covariates. To maximise the accuracy of SCMs and remove specification-searching opportunities, this paper advocates the use of covariates in contexts where they have important explanatory power. This is done using results from Monte Carlo simulations based on the work by Ferman et al. (2020). They should especially be considered when the number of pretreatment periods is small, and/or when covariates have multicollinear or nonlinear effects on the outcome variable. These recommendations are made after analysing experiments wherein specification searching opportunity p-values, average root-MSPEs and V matrices are calculated and compared for different types of specifications, DGPS and pretreatment time periods T_0. If the decision is made to include covariates in the SCM, it is essential that not all pretreatment outcome lags are included alongside them. Instead, a small set of pretreatment outcome lags, spread over the span T_0, performs best. Finally, the empirical application on German reunification is used to illustrate how a researcher should approach the choice of including covariates and how to include them. This paper therefore provides an extensive insight into how the choice of including covariates allows for specification-searching opportunities and which procedures should be taken to best avoid this.

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, *59*(2), 391–425.

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

Abadie, A., Diamond, A. & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, *59*(2), 495–510.

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

Athey, S. & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic perspectives*, *31*(2), 3–32.

Ben-Michael, E., Feller, A. & Rothstein, J. (2021). The augmented synthetic control method. *Journal of the American Statistical Association*, *116*(536), 1789–1803.

Billmeier, A. & Nannicini, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *Review of Economics and Statistics*, *95*(3), 983–1001.

Bohn, S., Lofstrom, M. & Raphael, S. (2014). Did the 2007 legal arizona workers act reduce the state's unauthorized immigrant population? *Review of Economics and Statistics*, *96*(2), 258–269.

Botosaru, I. & Ferman, B. (2019). On the role of covariates in the synthetic control method. *The Econometrics Journal*, *22*(2), 117–130.

Carriero, A., Galvão, A. B. & Kapetanios, G. (2019). A comprehensive evaluation of macroeconomic forecasting methods. *International Journal of Forecasting*, *35*(4), 1226–1239.

Cavallo, E., Galiani, S., Noy, I. & Pantano, J. (2013). Catastrophic natural disasters and economic growth. *Review of Economics and Statistics*, *95*(5), 1549–1561.

Cunningham, S. & Shah, M. (2018). Decriminalizing indoor prostitution: Implications for sexual violence and public health. *The Review of Economic Studies*, *85*(3), 1683–1715.

DeAngelo, G. & Hansen, B. (2014). Life and death in the fast lane: Police enforcement and traffic fatalities. *American Economic Journal: Economic Policy*, *6*(2), 231–257.

Doudchenko, N. & Imbens, G. W. (2016). *Balancing, regression, difference-in-differences and synthetic control methods: A synthesis* (Tech. Rep.). National Bureau of Economic Research.

Dube, A. & Zipperer, B. (2015). Pooling multiple case studies using synthetic controls: An application to minimum wage policies.

Dustmann, C., Fasani, F., Frattini, T., Minale, L. & Schönberg, U. (2017). On the economics and politics of refugee migration. *Economic policy*, *32*(91), 497–550.

Elliott, G. & Timmermann, A. (2016). Forecasting in economics and finance. *Annual Review of Economics*, *8*, 81–110.

Eren, O. & Ozbeklik, S. (2016). What do right-to-work laws do? evidence from a synthetic control method analysis. *Journal of Policy Analysis and Management*, *35*(1), 173–194.

Ferman, B., Pinto, C. & Possebom, V. (2020). Cherry picking with synthetic controls. *Journal of Policy Analysis and Management*, *39*(2), 510–532.

Gilchrist, D., Emery, T., Garoupa, N. & Spruk, R. (2023). Synthetic control method: A tool for comparative case studies in economic history. *Journal of Economic Surveys*, *37*(2), 409–445.

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

Hinrichs, P. (2012). The effects of affirmative action bans on college enrollment, educational attainment, and the demographic composition of universities. *Review of Economics and Statistics*, *94*(3), 712–722.

Imbens, G. W. & Rubin, D. B. (2010). Rubin causal model. *Microeconometrics*, 229–241.

Imbens, G. W. & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Imbens, G. W. & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of economic literature*, *47*(1), 5–86.

Kaul, A., Klößner, S., Pfeifer, G. & Schieler, M. (2022). Standard synthetic control methods: The case of using all preintervention outcomes together with covariates. *Journal of Business & Economic Statistics*, *40*(3), 1362–1376.

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

Lindo, J. M. & Packham, A. (2017). How much can expanding access to long-acting reversible contraceptives reduce teen birth rates? *American Economic Journal: Economic Policy*, *9*(3), 348–376.

Montalvo, J. G. (2011). Voting after the bombings: A natural experiment on the effect of terrorist attacks on democratic elections. *Review of Economics and Statistics*, *93*(4), 1146–1154.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861-1957. *economica*, *25*(100), 283–299.

Smith, B. (2015). The resource curse exorcised: Evidence from a panel of countries. *Journal of development economics*, *116*, 57–73.

Stock, J. H. & Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER macroeconomics annual*, *4*, 351–394.

Walker, W., Baccini, A., Schwartzman, S., Ríos, S., Oliveira-Miranda, M. A., Augusto, C., ... others (2014). Forest carbon in amazonia: the unrecognized contribution of indigenous territories and protected natural areas. *Carbon Management*, *5*(5-6), 479–485.

Zou, B. (2018). The local economic impacts of military personnel. *Journal of Labor Economics*, *36*(3), 589–621.

A Additional assumptions and definitions required 3.3

We also require the Assumptions relating to the unobserved transitory shocks $\varepsilon_{j,t}$ by Abadie et al. (2010):

$$\begin{aligned} &(a)\varepsilon_{j,t} \text{ } i.n.i.d.; \\ &(b)\mathbb{E}[\varepsilon_{j,t}|Z_j, \mu_j] = 0; \\ &(c)\mathbb{E}[|\varepsilon_j, t|^p] < \infty, \end{aligned} \tag{16}$$

For some even integer $p \geq 2$, for all $t = -T_0 + 1, ..., 0$ and $i = 2, ..., J + 1$. Additionally, we define

$$\bar{m}_p(T_0) \equiv \underset{i=2,...,J+1}{max} \left(\sum_{s=-T_0+1}^{0} E[|\varepsilon_{j,s}|^p] \right),$$

where $\bar{m}_p(T_0)$ can be interpreted as a measure for the scale of the transitory shock. Also let

$$\bar{\gamma}(T_0) \equiv \underset{t=-T_0+1,...,0,1,...,T;s=1,...,F}{max} \gamma$$

where $\bar{\gamma}(T_0)$ can be interpreted as the effect of the (un)observed covariate with the largest absolute shock.

B Specification searching opportunities

Table 9: Specification searching opportunities for the 'covariate DGP; standard specifications' case

	5% test	10% test
Panel A: *Specifications 1 to 7*		
$T_0 = 12$	0.143	0.254
	(0.011)	(0.013)
$T_0 = 32$	0.132	0.235
	(0.011)	(0.013)
Panel B: *Specifications 1 to 5*		
$T_0 = 12$	0.105	0.188
	(0.009)	(0.012)
$T_0 = 32$	0.086	0.169
	(0.009)	(0.012)

Notes: This table presents the probabilities of rejecting the null hypothesis of no treatment effect in at least one specification for the given set of specifications at the 5 and 10 percent significance levels. The results are based on 3,900 observations.

Table 10: Specification searching opportunities for the 'standard DGP; covariate specifications' case

	5% test	10% test
Panel A: *Specifications 1 to 7*		
$T_0 = 12$	0.152	0.240
	(0.010)	(0.013)
$T_0 = 32$	0.138	0.238
	(0.011)	(0.013)

Notes: This table presents the probabilities of rejecting the null hypothesis of no treatment effect in at least one specification for the given set of specifications at the 5 and 10 percent significance levels. The results are based on 3,900 observations.

C Sets of two

It is possible to gain an indication which specification is responsible for the most disparate SCEs. This can be done by comparing the p-values of sets of two specifications with each other. In this setting, the sets compared all contain specification 1, the supposedly 'optimal' specification, as well as one of the other 6 specifications. This allows these 6 specifications to be compared so as to determine which one of these specifications is most often responsible for an SCE which is significantly different from that of specification 1. The results are shown in Table 11. Please note that the validity of this investigation assumes that specification 1 is 'optimal'. In order to improve the interpretability of these results, only the results at the 5% significance levels have been included.

Table 11: Specification searching using a stationary model with time-invariant covariates - Sets of two Specifications

	Specifications					
	2	3	4	5	6	7
$T_0 = 12$	0.070	0.073	0.073	0.070	0.087	0.080
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)
$T_0 = 32$	0.064	0.067	0.061	0.060	0.081	0.076
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)
$T_0 = 100$	0.062	0.067	0.059	0.067	0.089	0.081
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)
$T_0 = 400$	0.064	0.065	0.062	0.058	0.092	0.078
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)

Notes: This table presents the probabilities of rejecting the null hypothesis of no treatment effect in at least one specification, for each set of two specifications, where each set includes specification 1, results are only shown at the 5 percent significance level. The results are based on 3,900 observations.

D Programming code

```
Obtaining root-MSPE values in Stata
forval i = 1/7 {
 local varname = "rmspe_32_‘i’"
    gen temp_‘i’ = sqrt(post_‘i’ * 10) / 10 if t0 == 32
   summarize temp_‘i’
```

```
   rename temp_‘i’ ‘varname’
}


################################################################################
# Article: Specification Search with the Synthetic Control Method:  DGP incl. covariates; St
# Authors: Bruno Ferman, Cristine Pinto and Vítor Possebom
# Function: Data Generating Process.
################################################################################
dgp <- function(J, T0, Tf, g) {
  # Parameter to generate our AR processes
  rho <- 0.5
  sd.rho <- sqrt(1-rho^2)
  # Create an empty dataframe.
  data <- as.data.frame(matrix(NA, J*Tf, 4))
  # Create the unit identifiers.
  for (j in 1:J) {
    data[(1 + (j-1)*Tf):(Tf + (j-1)*Tf), 1] <- rep(j, Tf)
  }
  # Create the time identifiers.
  data[, 2] <- rep(seq(from = 1, to  = Tf, by = 1), J)
  # Create the common time trend
  delta <- matrix(rnorm(Tf, mean = 0, sd = 1), Tf, 1)
  # Create the stationary components
  lambda <- matrix(NA, Tf, 10)
  for (k in 1:10) {
    lambda[, k] <- matrix(arima.sim(list(ar = rho), n =Tf, sd=sd.rho))
  }
  # Create the covariates
  Z <- as.numeric(data[, 1] <= 10)
  data[, 4] <- Z
  psi <- matrix(rnorm(Tf, mean = 0, sd = 1), Tf, 1)
  # Create the observed outcome
  for (j in 1:J) {
    # Group indicator
    mu1 <- matrix(c(as.numeric(j <= 2), as.numeric((j > 2) & (j <= 4)),
                    as.numeric((j > 4) & (j <= 6)),
                    as.numeric((j > 6) & (j <= 8)),
                    as.numeric((j > 8) & (j <= 10)),
                    as.numeric((j > 10) & (j <= 12)),
                    as.numeric((j > 12) & (j <= 14)),
                    as.numeric((j > 14) & (j <= 16)),
```

```
                    as.numeric((j > 16) & (j <= 18)),
                    as.numeric(j > 18)), 10, 1)
    # Error term
    epsilon <- matrix(rnorm(Tf, mean = 0, sd = sqrt(0.1)), Tf, 1)
    # Observed outcome
    data[(1 + (j-1)*Tf):(Tf + (j-1)*Tf), 3] <-
      delta + lambda %*% mu1 + psi * as.numeric(j <= 10) + epsilon
  }
  data <- data[, -4]
  return(data)


################################################################################
# Article: Specification Search with the Synthetic Control Method: Standard DGP: covariate s
# Authors: Bruno Ferman, Cristine Pinto and Vítor Possebom (adjusted by Hendrik J. de Boer)
# Function: Data Generating Process.
################################################################################
dgp <- function(J, T0, Tf, g) {
  # Parameter to generate our AR processes
  rho <- 0.5
  sd.rho <- sqrt(1-rho^2)
  # Create an empty dataframe.
  data <- as.data.frame(matrix(NA, J*Tf, 4))
  # Create the unit identifiers.
  for (j in 1:J) {
    data[(1 + (j-1)*Tf):(Tf + (j-1)*Tf), 1] <- rep(j, Tf)
  }
  # Create the time identifiers.
  data[, 2] <- rep(seq(from = 1, to  = Tf, by = 1), J)
  # Create the common time trend
  delta <- matrix(rnorm(Tf, mean = 0, sd = 1), Tf, 1)
  # Create the stationary components
  lambda <- matrix(NA, Tf, 10)
  for (k in 1:10) {
    lambda[, k] <- matrix(arima.sim(list(ar = rho), n =Tf, sd=sd.rho))
  }
  # Create the 2 stationary components when g == 2
  phi <- matrix(NA, Tf, 2)
  for (r in 1:2) {
    phi[, r] <- matrix(arima.sim(list(ar = rho), n =Tf, sd=sd.rho))
  }
  # Create the covariates
```

```
  Z <- as.numeric(data[, 1] <= 10)
  data[, 4] <- Z
#  psi <- matrix(rnorm(Tf, mean = 0, sd = 1), Tf, 1)
  # Create the observed outcome
  for (j in 1:J) {
    # Group indicator
    mu1 <- matrix(c(as.numeric(j <= 2), as.numeric((j > 2) & (j <= 4)),
                    as.numeric((j > 4) & (j <= 6)),
                    as.numeric((j > 6) & (j <= 8)),
                    as.numeric((j > 8) & (j <= 10)),
                    as.numeric((j > 10) & (j <= 12)),
                    as.numeric((j > 12) & (j <= 14)),
                    as.numeric((j > 14) & (j <= 16)),
                    as.numeric((j > 16) & (j <= 18)),
                    as.numeric(j > 18)), 10, 1)
    # Group indicator when g == 2
    mu2 <- matrix(c(as.numeric(j <= 10), as.numeric(j > 10)), 2, 1)
    # Error term
    epsilon <- matrix(rnorm(Tf, mean = 0, sd = sqrt(0.1)), Tf, 1)
    # Observed outcome
    data[(1 + (j-1)*Tf):(Tf + (j-1)*Tf), 3] <-
      delta + as.numeric(g == 10) * lambda %*% mu1 +
      as.numeric(g == 2) * phi %*% mu2 + epsilon
  }
  return(data)
}



################################################################################
# Article: Cherry Picking with the Synthetic Controls
# Authors: Bruno Ferman, Cristine Pinto and Vítor Possebom (adjusted by Hendrik J. de Boer)
# Code objective: This code runs a piece our Monte Carlo Experiment. vexperiment covariates
################################################################################
# Import libraries and functions
library("methods")
library("Synth")
library("doParallel")
source('function_dgp.R', encoding='UTF-8', echo=TRUE)
source('function_RMSPE_12preperiods.R', encoding='UTF-8', echo=TRUE)
source('function_RMSPE_32preperiods.R', encoding='UTF-8', echo=TRUE)
source('function_RMSPE_100preperiods.R', encoding='UTF-8', echo=TRUE)
```

```
source('function_RMSPE_400preperiods.R', encoding='UTF-8', echo=TRUE)
source('function_RMSPE.R', encoding='UTF-8', echo=TRUE)
source('function_SWWU.R', encoding='UTF-8', echo=TRUE)
source('function_VoW.R', encoding='UTF-8', echo=TRUE)
source('function_WA1.R', encoding = 'UTF-8')
source('function_VoGM1.R', encoding = 'UTF-8')
source('function_GA1.R', encoding = 'UTF-8')
source('function_VoGM2.R', encoding = 'UTF-8')
source('function_VoGM3.R', encoding = 'UTF-8')
# Setup parallel backend to use all but one of the cores.
n.cores <- detectCores() - 1
cl <- makeCluster(n.cores)
registerDoParallel(cl)
# Define the parameters.
lround <- 11
uround <- 20
T0 <- 12
g <- 10 # How many groups do we have?
MC <- 5 # How many Monte Carlo repetitions does each round have?
J <- 20 # How many units do we observe?
row_fraction_1_total <- rep(0, 13)
row_fraction_2_total <- rep(0, 10)
row_fraction_3_total <- rep(0, 7)
row_fraction_4_total <- rep(0, 7)
row_fraction_5_total <- rep(0, 7)
row_fraction_6_total <- rep(0, 2)
row_fraction_7_total <- rep(0, 4)

for (r in lround:uround) {
  # Define parameters that change inside the loop.
  t <- ifelse(T0 == 12, 1, ifelse(T0 == 32, 2, ifelse(T0 == 100, 3, 4)))
  seed <- as.numeric(
    paste(r, which(c(2, 10) == g), t, sep = ""))
  Tf <- T0 + 10
  # Run the Monte Carlo function.
  ##############################################################################
  # Set the seed for the random number generator.
  set.seed(seed)
  # Define the null hypothesis of no effect whatsoever.
  nullhypothesis <- matrix(rep(0, Tf), Tf, 1)
  ####################################
  # Create a matrix that will save all my results
```

```
MCresults <- matrix(NA, 20*MC, 160)
# Run MC Monte Carlo repetitions.
counter <- 0
attempt <- 0
while (counter < MC) {
  attempt <- attempt + 1
  # Print the seed number and the number of simulations associated to it that
  # have been already succesfully performerd.
  print("Seed Number:")
  print(seed)
  print("Monte Carlo repetition number: ")
  print(counter)
  print("Number of attempts: ")
  print(attempt)
  # Create a matrix that will temporarily save my results
  MCmatrix <- matrix(NA, 20, 159)
  # Generate the data for each Monte Carlo repetition using a factor model.
  data <- dgp(J, T0, Tf, g)
  # Conduct permutation tests with different specifications.
  synth_results <- foreach(
    j = 1:J, .combine = rbind, .inorder = FALSE, .errorhandling = "remove",
    .packages = "Synth", .verbose = FALSE) %dopar% {
      # Define the control units
      controlunits <- setdiff(1:J, j)
      ############
      # Specification #1: All the Pre-Treatment Outcome Values.
      specification <- 1
      dataprep.out.spec1 <- dataprep(
        foo = data,
        predictors = c("V4"),
        predictors.op = "mean",
        time.predictors.prior = seq(from = 1, to = T0, by = 1),
        special.predictors = list(
          list("V3", seq(from = 1, to = 1, by = 1), "mean"),
          list("V3", seq(from = 2, to = 2, by = 1), "mean"),
          list("V3", seq(from = 3, to = 3, by = 1), "mean"),
          list("V3", seq(from = 4, to = 4, by = 1), "mean"),
          list("V3", seq(from = 5, to = 5, by = 1), "mean"),
          list("V3", seq(from = 6, to = 6, by = 1), "mean"),
          list("V3", seq(from = 7, to = 7, by = 1), "mean"),
          list("V3", seq(from = 8, to = 8, by = 1), "mean"),
          list("V3", seq(from = 9, to = 9, by = 1), "mean"),
```

```
    list("V3", seq(from = 10, to = 10, by = 1), "mean"),
    list("V3", seq(from = 11, to = 11, by = 1), "mean"),
    list("V3", seq(from = T0, to = T0, by = 1), "mean")),
  dependent = "V3",
  unit.variable = "V1",
  time.variable = "V2",
  treatment.identifier = j,
  controls.identifier = controlunits,
  time.optimize.ssr = seq(from = 1, to = T0, by = 1),
  time.plot = seq(from = 1, to = Tf, by = 1))
synth.out.spec1 <- synth(data.prep.obj = dataprep.out.spec1,
                         method = "BFGS")
point.estimates.spec1 <-
  dataprep.out.spec1$Y0plot %*% synth.out.spec1$solution.w
gaps.spec1 <- dataprep.out.spec1$Y1plot - point.estimates.spec1
# Compute the RSMPE test statistic.
RMSPE.spec1 <- RMSPE(gaps.spec1, nullhypothesis, Tf, T0, specification)
# Compute a, b and Rsq
dataframe1 <- as.data.frame(
  cbind(dataprep.out.spec1$Y1plot[1:T0], point.estimates.spec1[1:T0]))
lm1 <- lm(V1 ~ V2, dataframe1)
a1 <- lm1$coefficients[1]
b1 <- lm1$coefficients[2]
rsq1 <- summary(lm1)$r.squared
# Compute RSSR
tempN <- as.matrix(dataframe1$V1 - dataframe1$V2)
tempD <- as.matrix(dataframe1$V1 - mean(dataframe1$V1))
tempN <- t(tempN) %*% tempN
tempD <- t(tempD) %*% tempD
RSSR1 <- 1 - tempN/tempD
############
# Specification #2: The First Three Fourths the Pre-Treatment
specification <- 2
dataprep.out.spec2 <- dataprep(
  foo = data,
  predictors = c("V4"),
  predictors.op = "mean",
  time.predictors.prior = seq(from = 1, to = T0, by = 1),
  special.predictors = list(
    list("V3", seq(from = 1, to = 1, by = 1), "mean"),
    list("V3", seq(from = 2, to = 2, by = 1), "mean"),
    list("V3", seq(from = 3, to = 3, by = 1), "mean"),
```

```
    list("V3", seq(from = 4, to = 4, by = 1), "mean"),
    list("V3", seq(from = 5, to = 5, by = 1), "mean"),
    list("V3", seq(from = 6, to = 6, by = 1), "mean"),
    list("V3", seq(from = 7, to = 7, by = 1), "mean"),
    list("V3", seq(from = 8, to = 8, by = 1), "mean"),
    list("V3", seq(from = 9, to = 9, by = 1), "mean")),
  dependent = "V3",
  unit.variable = "V1",
  time.variable = "V2",
  treatment.identifier = j,
  controls.identifier = controlunits,
  time.optimize.ssr = seq(from = (T0/2 + 1), to = T0, by = 1),
  time.plot = seq(from = 1, to = Tf, by = 1))
synth.out.spec2 <- synth(data.prep.obj = dataprep.out.spec2,
                         method = "BFGS")
point.estimates.spec2 <-
  dataprep.out.spec2$Y0plot %*% synth.out.spec2$solution.w
gaps.spec2 <- dataprep.out.spec2$Y1plot - point.estimates.spec2
# Compute the RSMPE test statistic.
RMSPE.spec2 <- RMSPE(gaps.spec2, nullhypothesis, Tf, T0, specification)
# Compute a, b and Rsq
dataframe2 <- as.data.frame(
  cbind(dataprep.out.spec2$Y1plot[1:T0], point.estimates.spec2[1:T0]))
lm2 <- lm(V1 ~ V2, dataframe2)
a2 <- lm2$coefficients[1]
b2 <- lm2$coefficients[2]
rsq2 <- summary(lm2)$r.squared
# Compute RSSR
tempN <- as.matrix(dataframe2$V1 - dataframe2$V2)
tempD <- as.matrix(dataframe2$V1 - mean(dataframe2$V1))
tempN <- t(tempN) %*% tempN
tempD <- t(tempD) %*% tempD
RSSR2 <- 1 - tempN/tempD
############
# Specification #3: The First Half of the Pre-Treatment Periods
# Outcome Values.
specification <- 3
dataprep.out.spec3 <- dataprep(
  foo = data,
  predictors = c("V4"),
  predictors.op = "mean",
  time.predictors.prior = seq(from = 1, to = T0, by = 1),
```

```
  special.predictors = list(
    list("V3", seq(from = 1, to = 1, by = 1), "mean"),
    list("V3", seq(from = 2, to = 2, by = 1), "mean"),
    list("V3", seq(from = 3, to = 3, by = 1), "mean"),
    list("V3", seq(from = 4, to = 4, by = 1), "mean"),
    list("V3", seq(from = 5, to = 5, by = 1), "mean"),
    list("V3", seq(from = 6, to = 6, by = 1), "mean")),
  dependent = "V3",
  unit.variable = "V1",
  time.variable = "V2",
  treatment.identifier = j,
  controls.identifier = controlunits,
  time.optimize.ssr = seq(from = (3*T0/4 + 1), to = T0, by = 1),
  time.plot = seq(from = 1, to = Tf, by = 1))
synth.out.spec3 <- synth(data.prep.obj = dataprep.out.spec3,
                         method = "BFGS")
point.estimates.spec3 <-
  dataprep.out.spec3$Y0plot %*% synth.out.spec3$solution.w
gaps.spec3 <- dataprep.out.spec3$Y1plot - point.estimates.spec3
# Compute the RSMPE test statistic.
RMSPE.spec3 <- RMSPE(gaps.spec3, nullhypothesis, Tf, T0, specification)
# Compute a, b and Rsq
dataframe3 <- as.data.frame(
  cbind(dataprep.out.spec3$Y1plot[1:T0], point.estimates.spec3[1:T0]))
lm3 <- lm(V1 ~ V2, dataframe3)
a3 <- lm3$coefficients[1]
b3 <- lm3$coefficients[2]
rsq3 <- summary(lm3)$r.squared
# Compute RSSR
tempN <- as.matrix(dataframe3$V1 - dataframe3$V2)
tempD <- as.matrix(dataframe3$V1 - mean(dataframe3$V1))
tempN <- t(tempN) %*% tempN
tempD <- t(tempD) %*% tempD
RSSR3 <- 1 - tempN/tempD
############
# Specification #4: Odd Pre-Treatment Outcome Values.
specification <- 4
dataprep.out.spec4 <- dataprep(
  foo = data,
  predictors = c("V4"),
  predictors.op = "mean",
  time.predictors.prior = seq(from = 1, to = T0, by = 1),
```

```
  special.predictors = list(
    list("V3", seq(from = 1, to = 1, by = 1), "mean"),
    list("V3", seq(from = 3, to = 3, by = 1), "mean"),
    list("V3", seq(from = 5, to = 5, by = 1), "mean"),
    list("V3", seq(from = 7, to = 7, by = 1), "mean"),
    list("V3", seq(from = 9, to = 9, by = 1), "mean"),
    list("V3", seq(from = 11, to = 11, by = 1), "mean")),
  dependent = "V3",
  unit.variable = "V1",
  time.variable = "V2",
  treatment.identifier = j,
  controls.identifier = controlunits,
  time.optimize.ssr = seq(from = 2, to = T0, by = 2),
  time.plot = seq(from = 1, to = Tf, by = 1))
synth.out.spec4 <- synth(data.prep.obj = dataprep.out.spec4,
                         method = "BFGS")
point.estimates.spec4 <-
  dataprep.out.spec4$Y0plot %*% synth.out.spec4$solution.w
gaps.spec4 <- dataprep.out.spec4$Y1plot - point.estimates.spec4
# Compute the RSMPE test statistic.
RMSPE.spec4 <- RMSPE(gaps.spec4, nullhypothesis, Tf, T0, specification)
# Compute a, b and Rsq
dataframe4 <- as.data.frame(
  cbind(dataprep.out.spec4$Y1plot[1:T0], point.estimates.spec4[1:T0]))
lm4 <- lm(V1 ~ V2, dataframe4)
a4 <- lm4$coefficients[1]
b4 <- lm4$coefficients[2]
rsq4 <- summary(lm4)$r.squared
# Compute RSSR
tempN <- as.matrix(dataframe4$V1 - dataframe4$V2)
tempD <- as.matrix(dataframe4$V1 - mean(dataframe4$V1))
tempN <- t(tempN) %*% tempN
tempD <- t(tempD) %*% tempD
RSSR4 <- 1 - tempN/tempD
############
# Specification #5: Even Pre-Treatment Outcome Values.
specification <- 5
dataprep.out.spec5 <- dataprep(
  foo = data,
  predictors = c("V4"),
  predictors.op = "mean",
  time.predictors.prior = seq(from = 1, to = T0, by = 1),
```

```
  special.predictors = list(
    list("V3", seq(from = 2, to = 2, by = 1), "mean"),
    list("V3", seq(from = 4, to = 4, by = 1), "mean"),
    list("V3", seq(from = 6, to = 6, by = 1), "mean"),
    list("V3", seq(from = 8, to = 8, by = 1), "mean"),
    list("V3", seq(from = 10, to = 10, by = 1), "mean"),
    list("V3", seq(from = T0, to = T0, by = 1), "mean")),
  dependent = "V3",
  unit.variable = "V1",
  time.variable = "V2",
  treatment.identifier = j,
  controls.identifier = controlunits,
  time.optimize.ssr = seq(from = 1, to = (T0 - 1), by = 2),
  time.plot = seq(from = 1, to = Tf, by = 1))
synth.out.spec5 <- synth(data.prep.obj = dataprep.out.spec5,
                         method = "BFGS")
point.estimates.spec5 <-
  dataprep.out.spec5$Y0plot %*% synth.out.spec5$solution.w
gaps.spec5 <- dataprep.out.spec5$Y1plot - point.estimates.spec5
# Compute the RSMPE test statistic.
RMSPE.spec5 <- RMSPE(gaps.spec5, nullhypothesis, Tf, T0, specification)
# Compute a, b and Rsq
dataframe5 <- as.data.frame(
  cbind(dataprep.out.spec5$Y1plot[1:T0], point.estimates.spec5[1:T0]))
lm5 <- lm(V1 ~ V2, dataframe5)
a5 <- lm5$coefficients[1]
b5 <- lm5$coefficients[2]
rsq5 <- summary(lm5)$r.squared
# Compute RSSR
tempN <- as.matrix(dataframe5$V1 - dataframe5$V2)
tempD <- as.matrix(dataframe5$V1 - mean(dataframe5$V1))
tempN <- t(tempN) %*% tempN
tempD <- t(tempD) %*% tempD
RSSR5 <- 1 - tempN/tempD
############
# Specification #6: Mean of the Pre-Treatment Outcome Values.
specification <- 6
dataprep.out.spec6 <- dataprep(
  foo = data,
  predictors = c("V3", "V4"),
  predictors.op = "mean",
  time.predictors.prior = seq(from = 1, to = T0, by = 1),
```

```
  dependent = "V3",
  unit.variable = "V1",
  time.variable = "V2",
  treatment.identifier = j,
  controls.identifier = controlunits,
  time.optimize.ssr = seq(from = 1, to = T0, by = 1),
  time.plot = seq(from = 1, to = Tf, by = 1))
synth.out.spec6 <- synth(data.prep.obj = dataprep.out.spec6,
                         method = "BFGS")
point.estimates.spec6 <-
  dataprep.out.spec6$Y0plot %*% synth.out.spec6$solution.w
gaps.spec6 <- dataprep.out.spec6$Y1plot - point.estimates.spec6
# Compute the RSMPE test statistic.
RMSPE.spec6 <- RMSPE(gaps.spec6, nullhypothesis, Tf, T0, specification)
# Compute a, b and Rsq
dataframe6 <- as.data.frame(
  cbind(dataprep.out.spec6$Y1plot[1:T0], point.estimates.spec6[1:T0]))
lm6 <- lm(V1 ~ V2, dataframe6)
a6 <- lm6$coefficients[1]
b6 <- lm6$coefficients[2]
rsq6 <- summary(lm6)$r.squared
# Compute RSSR
tempN <- as.matrix(dataframe6$V1 - dataframe6$V2)
tempD <- as.matrix(dataframe6$V1 - mean(dataframe6$V1))
tempN <- t(tempN) %*% tempN
tempD <- t(tempD) %*% tempD
RSSR6 <- 1 - tempN/tempD
############
# Specification #7: First, Median and Last Pre-intervention period
specification <- 7
dataprep.out.spec7 <- dataprep(
  foo = data,
  predictors = c("V4"),
  predictors.op = "mean",
  time.predictors.prior = seq(from = 1, to = T0, by = 1),
  special.predictors = list(
    list("V3", seq(from = 1, to = 1, by = 1), "mean"),
    list("V3", seq(from = T0/2, to = T0/2, by = 1), "mean"),
    list("V3", seq(from = T0, to = T0, by = 1), "mean")),
  dependent = "V3",
  unit.variable = "V1",
  time.variable = "V2",
```

```
          treatment.identifier = j,
          controls.identifier = controlunits,
          time.optimize.ssr = seq(from = 1, to = T0, by = 1),
          time.plot = seq(from = 1, to = Tf, by = 1))
        synth.out.spec7 <- synth(data.prep.obj = dataprep.out.spec7,
                                 method = "BFGS")
        point.estimates.spec7 <-
          dataprep.out.spec7$Y0plot %*% synth.out.spec7$solution.w
        gaps.spec7 <- dataprep.out.spec7$Y1plot - point.estimates.spec7
        # Compute the RSMPE test statistic.
        RMSPE.spec7 <- RMSPE(gaps.spec7, nullhypothesis, Tf, T0, specification)
        # Compute a, b and Rsq
        dataframe7 <- as.data.frame(
          cbind(dataprep.out.spec7$Y1plot[1:T0], point.estimates.spec7[1:T0]))
        lm7 <- lm(V1 ~ V2, dataframe7)
        a7 <- lm7$coefficients[1]
        b7 <- lm7$coefficients[2]
        rsq7 <- summary(lm7)$r.squared
        # Compute RSSR
        tempN <- as.matrix(dataframe7$V1 - dataframe7$V2)
        tempD <- as.matrix(dataframe7$V1 - mean(dataframe7$V1))
        tempN <- t(tempN) %*% tempN
        tempD <- t(tempD) %*% tempD
        RSSR7 <- 1 - tempN/tempD
        ############
        # Save the useful information in a matrix
        cluster <- as.numeric(paste(seed, counter, sep = ""))
        return(c(synth.out.spec1$solution.v, synth.out.spec2$solution.v, synth.out.spec3$sol
))      }
    # Test whether we estimated a synthetic control unit for each placebo unit.
    if (dim(synth_results)[1] == J) {
      # Update the counter
      counter <- counter + 1
      # Save the weights for each synthetic control unit associated with each
      # placebo unit and each specification

      weights1 <- as.matrix(as.vector(t(synth_results[, 1:13])), J*(J-1), 1)
      weights2 <- as.matrix(as.vector(t(synth_results[, 14:23])), J*(J-1), 1)
      weights3 <- as.matrix(as.vector(t(synth_results[, 24:30])), J*(J-1), 1)
      weights4 <- as.matrix(as.vector(t(synth_results[, 31:37])), J*(J-1), 1)
      weights5 <- as.matrix(as.vector(t(synth_results[, 38:44])), J*(J-1), 1)
      weights6 <- as.matrix(as.vector(t(synth_results[, 45:46])), J*(J-1), 1)
```

```
weights7 <- as.matrix(as.vector(t(synth_results[, 47:50])), J*(J-1), 1)

# Row fractions for weights1
weights1 <- as.data.frame(weights1)
weights1 <- sapply(weights1, as.numeric)
row_sums_1 <- rowSums(weights1)
row_fraction_1 <- row_sums_1 / sum(row_sums_1)
print(row_fraction_1)
row_fraction_1_total = row_fraction_1_total + row_fraction_1


# Row fractions for weights2
weights2 <- as.data.frame(weights2)
weights2 <- sapply(weights2, as.numeric)
row_sums_2 <- rowSums(weights2)
row_fraction_2 <- row_sums_2 / sum(row_sums_2)
print(row_fraction_2)
row_fraction_2_total = row_fraction_2_total + row_fraction_2

# Row fractions for weights3
weights3 <- as.data.frame(weights3)
weights3 <- sapply(weights3, as.numeric)
row_sums_3 <- rowSums(weights3)
row_fraction_3 <- row_sums_3 / sum(row_sums_3)
print(row_fraction_3)
row_fraction_3_total = row_fraction_3_total + row_fraction_3

# Row fractions for weights4
weights4 <- as.data.frame(weights4)
weights4 <- sapply(weights4, as.numeric)
row_sums_4 <- rowSums(weights4)
row_fraction_4 <- row_sums_4 / sum(row_sums_4)
print(row_fraction_4)
row_fraction_4_total = row_fraction_4_total + row_fraction_4

# Row fractions for weights5
weights5 <- as.data.frame(weights5)
weights5 <- sapply(weights5, as.numeric)
row_sums_5 <- rowSums(weights5)
row_fraction_5 <- row_sums_5 / sum(row_sums_5)
print(row_fraction_5)
row_fraction_5_total = row_fraction_5_total + row_fraction_5
```

```
      # Row fractions for weights6
      weights6 <- as.data.frame(weights6)
      weights6 <- sapply(weights6, as.numeric)
      row_sums_6 <- rowSums(weights6)
      row_fraction_6 <- row_sums_6 / sum(row_sums_6)
      print(row_fraction_6)
      row_fraction_6_total = row_fraction_6_total + row_fraction_6

      #Row fractions for weights6
      weights7 <- as.data.frame(weights7)
      weights7 <- sapply(weights7, as.numeric)
      row_sums_7 <- rowSums(weights7)
      row_fraction_7 <- row_sums_7 / (sum(row_sums_7))
      print(row_fraction_7)
      row_fraction_7_total = row_fraction_7_total + row_fraction_7



      # Compute the Sum of Weights given to the wrong comparison units for each
      # specification.
#      weights1.sum <- SWWU(weights1, g)
     # MCmatrix[, 27:33] <- weights.sum
      # Compute our measure of variability of weights
      #vow <- VoW(weights)
      #MCmatrix[, 34:37] <- vow
      # Compare weights of each specification against specification 1
      #wa1 <- WA1(weights)
      #MCmatrix[, 38:43] <- wa1
      # Save the gaps of each synthetic control unit associated with each
      # placebo unit and each specification
      #gaps <- matrix(NA, J*10, 7)
      #gaps[, 1] <- as.matrix(as.vector(t(synth_results[, 6:15])), J*10, 1)
      #gaps[, 2] <- as.matrix(as.vector(t(synth_results[, 42:51])), J*10, 1)
      #gaps[, 3] <- as.matrix(as.vector(t(synth_results[, 78:87])), J*10, 1)
      #gaps[, 4] <- as.matrix(as.vector(t(synth_results[, 114:123])), J*10, 1)
      #gaps[, 5] <- as.matrix(as.vector(t(synth_results[, 150:159])), J*10, 1)
      #gaps[, 6] <- as.matrix(as.vector(t(synth_results[, 186:195])), J*10, 1)
      #gaps[, 7] <- as.matrix(as.vector(t(synth_results[, 222:231])), J*10, 1)
      # Compute Measure 1 of variability of estimated gaps. This measure is
      # similar in spirit to VoW.
      #vogm1 <- VoGM1(gaps)
```

```
    #MCmatrix[, 44:47] <- vogm1
    # Compare the estimated gaps of each specification against specification
    # 1. This measure is similar in spirit to WA1.
    #ga1 <- GA1(gaps)
    #MCmatrix[, 48:53] <- ga1
    # Compute Measure 2 of variability of estimated gaps. This measure
    # computes the mean gap across time for each specification and, then,
    # takes the difference between the largest and the smallest mean.
    #vogm2 <- VoGM2(gaps)
    #MCmatrix[, 54:57] <- vogm2
    # Compute Measure 3 of variability of estimated gaps. This measures
    # computes the mean gap across time for each specification and, then,
    # computes the standard deviation across means.
    #vogm3 <- VoGM3(gaps)
    #MCmatrix[, 58:61] <- vogm3
    # Save the estimated gaps.
    #MCmatrix[, 62:71] <- synth_results[, 6:15] # Specification 1
    #MCmatrix[, 72:81] <- synth_results[, 42:51] # Specification 2
    #MCmatrix[, 82:91] <- synth_results[, 78:87] # Specification 3
    #MCmatrix[, 92:101] <- synth_results[, 114:123] # Specification 4
    #MCmatrix[, 102:111] <- synth_results[, 150:159] # Specification 5
    #MCmatrix[, 112:121] <- synth_results[, 186:195] # Specification 6
    #MCmatrix[, 122:131] <- synth_results[, 222:231] # Specification 7
    # Store the R^2
    #MCmatrix[, 132:138] <- synth_results[, c(38, 74, 110, 146, 182, 218, 254)]
    # Store the a coefficients
    #MCmatrix[, 139:145] <- synth_results[, c(39, 75, 111, 147, 183, 219, 255)]
    # Store the b coefficients
    #MCmatrix[, 146:152] <- synth_results[, c(40, 76, 112, 148, 184, 220, 256)]
    # Store the RSSR
    #MCmatrix[, 153:159] <- synth_results[, c(41, 77, 113, 149, 185, 221, 257)]
    # Combine all the Monte Carlo repetition results in one matrix.
    #MCresults[(1 + (counter-1)*J):(20 + (counter-1)*J), 1:159] <- MCmatrix
  }
}

V_1_proportion <- row_fraction_1_total[1] / sum(row_fraction_1_total)
V_2_proportion <- row_fraction_2_total[1] / sum(row_fraction_2_total)
V_3_proportion <- row_fraction_3_total[1] / sum(row_fraction_3_total)
V_4_proportion <- row_fraction_4_total[1] / sum(row_fraction_4_total)
V_5_proportion <- row_fraction_5_total[1] / sum(row_fraction_5_total)
V_6_proportion <- row_fraction_6_total[1] / sum(row_fraction_6_total)
```

```
  V_7_proportion <- row_fraction_7_total[1] / sum(row_fraction_7_total)
#  MCresults[1, 160] <- attempt
  ##############################################################################
  # Rename the columns of MCresults.
#  colnames(MCresults) <- c(
 #   "cluster", "r", "T0", "Tf", "j",
  #  "RMSPE_v_1", "RMSPE_v_2", "RMSPE_v_3", "RMSPE_v_4", "RMSPE_v_5", "RMSPE_v_6", "RMSPE_v_
  #  "RMSPE_a_1", "RMSPE_a_2", "RMSPE_a_3", "RMSPE_a_4", "RMSPE_a_5", "RMSPE_a_6", "RMSPE_a_
  #  "post_1", "post_2", "post_3", "post_4", "post_5", "post_6", "post_7",
  #  "MW_1", "MW_2", "MW_3", "MW_4", "MW_5", "MW_6", "MW_7",
  #  "VoW_all", "VoW_exc7", "VoW_ex6", "VoW_ex6_7",
  #  "WA1_2", "WA1_3", "WA1_4", "WA1_5", "WA1_6", "WA1_7",
  #  "VoGM1_all", "VoGM1_exc7", "VoGM1_ex6", "VoGM1_ex6_7",
  #  "GA1_2", "GA1_3", "GA1_4", "GA1_5", "GA1_6", "GA1_7",
  #  "VoGM2_all", "VoGM2_exc7", "VoGM2_ex6", "VoGM2_ex6_7",
  #  "VoGM3_all", "VoGM3_exc7", "VoGM3_ex6", "VoGM3_ex6_7",
  #  "gap_t1_s1", "gap_t2_s1", "gap_t3_s1", "gap_t4_s1", "gap_t5_s1", "gap_t6_s1", "gap_t7_s
  #  "gap_t1_s2", "gap_t2_s2", "gap_t3_s2", "gap_t4_s2", "gap_t5_s2", "gap_t6_s2", "gap_t7_s
  #  "gap_t1_s3", "gap_t2_s3", "gap_t3_s3", "gap_t4_s3", "gap_t5_s3", "gap_t6_s3", "gap_t7_s
  #  "gap_t1_s4", "gap_t2_s4", "gap_t3_s4", "gap_t4_s4", "gap_t5_s4", "gap_t6_s4", "gap_t7_s
  #  "gap_t1_s5", "gap_t2_s5", "gap_t3_s5", "gap_t4_s5", "gap_t5_s5", "gap_t6_s5", "gap_t7_s
  #  "gap_t1_s6", "gap_t2_s6", "gap_t3_s6", "gap_t4_s6", "gap_t5_s6", "gap_t6_s6", "gap_t7_s
  #  "gap_t1_s7", "gap_t2_s7", "gap_t3_s7", "gap_t4_s7", "gap_t5_s7", "gap_t6_s7", "gap_t7_s
  #  "Rsq_1", "Rsq_2", "Rsq_3", "Rsq_4", "Rsq_5", "Rsq_6",  "Rsq_7",
  #  "a_1", "a_2", "a_3", "a_4", "a_5", "a_6", "a_7",
  #  "b_1", "b_2", "b_3", "b_4", "b_5", "b_6", "b_7",
  #  "Rtilde_1", "Rtilde_2", "Rtilde_3", "Rtilde_4", "Rtilde_5", "Rtilde_6", "Rtilde_7",
  #  "attempts")
  # Export the results for a .csv file.
#  file.partial <- paste(
 #   "partial_results/MCresults_T0-", T0, "_g-", g, "_r-", r, ".csv", sep = "")
#  write.csv(MCresults, file.partial)
}
# Stop parallel backend
stopCluster(cl)
# Clear the memory.
rm(list=ls(all = TRUE))
gc(verbose = TRUE)
```