Synthetic Control Method with Missing Pre-treatment Outcomes

Naijia Liu* Sooahn Shin† Soichiro Yamauchi

August 14, 2024

Abstract

The Synthetic Control Method (SCM) is a powerful tool for causal inference in policy evaluation and comparative case studies. However, conventional SCM cannot accommodate observations with any missing pre-treatment outcomes, thereby resulting in their exclusion. The impact of such exclusion remains to be examined. To that end, we present in this paper a sensitivity analysis framework for SCM, which assesses the robustness of treatment effect estimates to variations in pre-treatment outcome data. Leveraging vertical regression as an estimation strategy, our framework offers new insights into decomposing bias that stems from omitted control units and missing pre-treatment outcomes. We then utilize the findings from bias decomposition to propose a sensitivity analysis for evaluating the effect of missing pre-treatment outcomes. We present an application of our proposed sensitivity analysis framework using an empirical data on Taiwan's expulsion from the International Monetary Fund (IMF).

*Assistant professor of Government, Harvard University
†PhD candidate in Government, Harvard University

1 Introduction

The Synthetic Control Method (SCM) has become one of the most widely used tools in political science for studying policy interventions with panel data. With SCM, researchers can estimate the treatment effect on a single unit using multiple control units with extensive pre-treatment period data. For example, SCM can be used to study how war affects post-war state-building trajectories (Schenoni, 2021), how compulsory voting influences vote-seeking behaviors (Singh, 2019), the impact of campaign finance regulations in the US (Gilens, Patterson and Haines, 2021), how the entry of a radical party affects polarization (Bischof and Wagner, 2019), as well as in a variety of other contexts (e.g., Eibl and Hertog, 2023; Marble et al., 2021; Blair, Grossman and Weinstein, 2022; Grumbach, 2023; Masterson and Yasenov, 2021; Tellez, 2022; De Kadt, Johnson-Kanu and Sands, 2020).

The main idea behind SCM is to construct a synthetic treated unit as a combination of control units, allowing researchers to approximate the counterfactual outcome that would have occurred if no treatment had been administered. The origins of SCM trace back to Abadie and Gardeazabal (2003), a study on the effects of terrorism on economic growth in the Basque Country. In this study, the authors used a combination of two Spanish regions to approximate the economic growth that the Basque Country would have experienced in the absence of terrorism. The intuition is that "a combination of units often provides a better comparison for the unit exposed to the intervention than any single unit alone (Abadie, Diamond and Hainmueller, 2010)." In line with this idea, the SCM was formalized for application in comparative case studies (Abadie, Diamond and Hainmueller, 2010, 2015), where a suitable comparison group is chosen by a data-driven procedure. This involves constructing a synthetic treated unit as a weighted average of control units, where the weights balance the pre-treatment outcomes.

The initial selection of comparable control units, from which a subset is chosen to construct a synthetic unit, is a fundamental step in conducting SCM. A natural follow-up question then is: What happens if the researcher starts with a constrained set of control units? This is especially relevant to practitioners, given that SCM requires panel data with sufficient pre-treatment periods, which are often subject to missingness. In such cases, researchers resort to excluding control units with any missing values. However, there has been limited examination of the potential implications of this practice, particularly concerning the robustness of results against the bias introduced by omitting control units. Does a 'good' pre-treatment fit guarantee a small bias? If not, under what conditions can we ensure that the bias remains small? How can researchers effectively report the sensitivity of their results to this bias? This paper aims to address these questions.

We address the missing data problem in SCM by decomposing the bias and proposing a sensitivity analysis. The main idea is to employ an alternative estimation technique for SCM: a vertical regression where the unit of analysis is the time period (Athey et al., 2021*a*). In this regression model, the control units become independent variables, and the treated unit becomes the dependent variable. The coefficients assigned to these control units can be interpreted as weights for constructing the synthetic unit. Based on this approach, we decompose the bias resulting from omitting the control unit into two parts: the weight of the missing unit on the synthetic unit and the imbalance between missing unit and other observed control units. We demonstrate the equivalence of such bias with the classical omitted variable bias, which enables researchers to apply existing tools that extend omitted variable bias framework for the sensitivity analysis (Cinelli and Hazlett, 2020).

In section 1.1, we motivate our method using a study on the economic impact of the 1990 German reunification (Abadie, Diamond and Hainmueller, 2015). The authors constructed a synthetic West Germany using data from 16 OECD member countries and assessed the robustness of their findings against different configuration of control units. However, such leave-one-out analysis for testing the sensitivity of the results to different choices of control units is only possible with complete observations. This motivates the development of methods to guide the assessment of the sensitivity of SCM estimates to missing values. The motivating example is followed by an introduction to the setup and the decomposition of bias. We then propose a sensitivity analysis based on this bias decomposition. Finally, we revisit the application study of Taiwan's expulsion from the IMF (Lipscy and Lee, 2019) and apply our proposed method.

1.1 Motivating Example

Abadie, Diamond and Hainmueller (2015) illustrates the fundamental idea of SCM and highlights its suitability for comparative case studies through an analysis of the economic repercussions of the 1990 German reunification on West Germany. The authors employed a panel dataset spanning from 1960 to 2003, where 16 OECD member countries constitute the "donor pool," that is, a reservoir of potential comparison units (Abadie, Diamond and Hainmueller, 2015). The original paper uses real per capita GDP, adjusted and measured in 2002 U.S. dollars (USD), as the outcome variable, along with other economic growth predictors such as the inflation rate and investment rate. In our illustration, we replicate the study using a vertical regression approach, which we further elaborate on in the following sections, and focus on the main outcome variable, real per capita GDP, without additional predictors.

The dataset used in this study does not contain any missing values, serving as a clear demonstration of our method. Figure 1 shows the synthetic control result following our vertical regression approach. Here, the per capita GDP of West Germany (black solid line) from 1960 to 2003 is compared to the synthetic West Germany (red dashed line), which is a weighted average of selected control units. The vertical line at 1990 indicates the year of the German reunification, when the treatment took place in West Germany. In this figure, we can observe that the pre-treatment fit of the synthetic West Germany is close to West Germany, indicating a successful construction of the synthetic control unit. The estimated treatment effect is -3206.6 (standard error: 1071.8) in 2003, suggesting a negative impact of the German reunification on West Germany's per capita GDP.

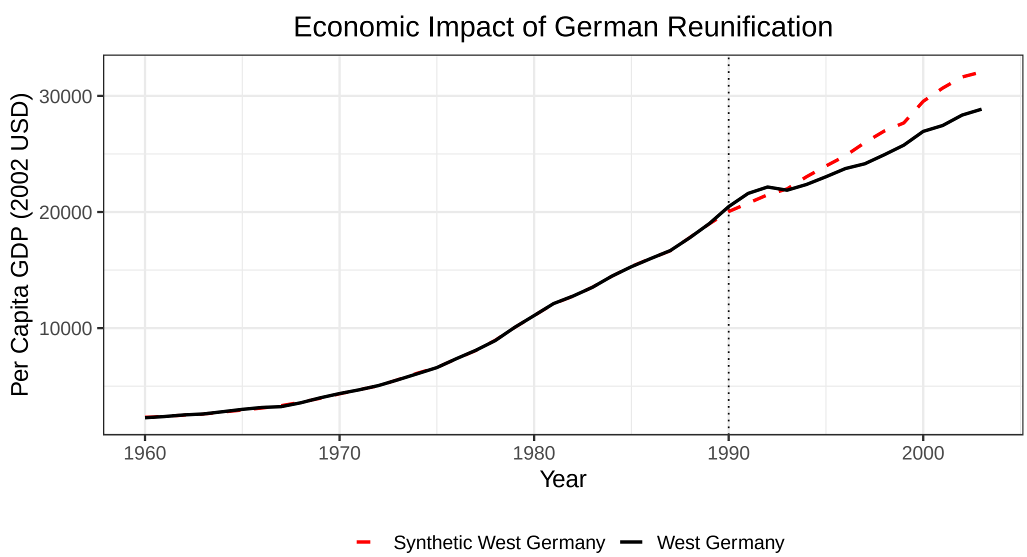

Figure 1: The SCM Estimate of the Impact of the 1990 German Reunification (Abadie, Diamond and Hainmueller, 2015) using Vertical Regression Estimator. The black solid line indicates the per capita GDP of West Germany, and the red dashed line represents that of synthetic West Germany.

Are these results robust to changes in the country weights? To answer this question, the authors conducted a leave-one-out analysis to assess the robustness of their results to changes in the donor pool. Specifically, they examined the impact of excluding each of the top five control units that received the highest weight in constructing the synthetic control unit on the estimated treatment effect. The pre-treatment fit of each leave-one-out analysis was close to the original synthetic control unit, as presented in the appendix Figure 9. However, the estimated treatment effect varied across different scenarios, as shown in Table 1. For example, excluding the Netherlands (USA) from the analysis yielded an estimated treatment effect of -2629.5 (-4829.4) in 2003. Although the substantive results remain consistent across these scenarios, the difference in the magnitude of the estimated treatment effect is non-negligible in these cases.

Omitted Country	1993	1998	2003
Netherlands	96 (135.9)	-1488.4 (433.3)	-2629.5 (1123.1)
Japan	-79.9 (162.5)	-1770.2 (430.4)	-2703.4 (987.3)
Switzerland	-146.1 (146.4)	-2120.9 (459.3)	-3098.7 (1032.3)
Austria	-111.6 (161.8)	-2248.4 (430.2)	-3752.2 (882.3)
USA	-27.5 (185.5)	-2407.8 (541.8)	-4829.4 (992.7)
None	-108.7 (162.8)	-2043 (488.3)	-3206.6 (1071.8)

Table 1: Leave-One-Out Analysis of the Impact of the 1990 German Reunification (Abadie, Diamond and Hainmueller, 2015) using Vertical Regression Estimator. The numbers in the table represent the estimated treatment effect in 1993, 1998, and 2003 (columns), respectively, when each of the top five control units (rows) is excluded from the analysis. The standard error is presented in parentheses.

This leave-one-out analysis highlights the importance of considering the impact of missing values on the study's results. In this specific case, we could directly observe the impact of omitting a control unit on the estimated treatment effect, as we have access to the full dataset. However, in practice, researchers often face missing values in their data, complicating the analysis. In such cases, researchers resort to excluding control units with missing values, but the impact of this exclusion cannot be directly assessed since a comparison of leave-one-out analysis with the full data is not feasible. To address this issue, we derive the bias due to missing values under SCM and propose a sensitivity analysis based on bias decomposition. Furthermore, we provide a method to utilize the partially observed data to estimate the sensitivity parameters and the bias. This allows for a systematic assessment of how missing data influence the estimated treatment effect, providing insights into the robustness of the findings.

2 Synthetic Control Method with Missing Pre-treatment Outcome

2.1 The Setup: SCM as Vertical Regression

Following the setup of the standard synthetic control method, we consider panel data with N units observed for T time periods, where $T_0 = \{1, \ldots, T-1\}$ is the pre-treatment period. We focus on a case where only one unit is treated and we observe a single post-treatment period T.

Let Y_t denote the outcome for the treated unit, and X_{it} denote the outcome of the control unit for $t = 1, \ldots, T$ and $i = 1, \ldots, N-2$. We also denote the observation of the last unit in the control group by Z_t.

We are interested in estimating the treatment effect on the treated unit at post-treatment period T:

$$\tau = Y_T(1) - Y_T(0)$$

where $Y_T(d)$ denote the potential outcome under the treatment condition $d \in \{0, 1\}$. Since the treated unit actually receives the treatment at time T, we have $Y_T(1) = Y_T^{obs}$. Therefore, the remaining task is to identify and estimate the counterfactual outcome $Y_T(0)$ using the control units.

Following Athey et al. (2021*b*), we can view SCM as a vertical regression, where we compute a set of weights for control units by regressing the outcome of the treated unit Y_t on the outcomes of the control units $(\mathbf{X}_t, Z_t)$ during the pre-treatment period. Note that here we relax the restriction of the original SCM that the weights are nonnegative and sum to one. In light of this setup, we structure the data as follows:

$$\left(\begin{array}{c|ccc|c} Y_1 & X_{11} & \cdots & X_{N-2,1} & Z_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Y_{T-1} & X_{1,T-1} & \cdots & X_{N-2,T-1} & Z_{T-1} \\ \hline Y_T & X_{1T} & \cdots & X_{N-2,T} & Z_T \end{array}\right) = \left(\begin{array}{c|c|c} \boldsymbol{Y}_{T_0} & \mathbf{X}_{T_0} & \boldsymbol{Z}_{T_0} \\ \hline Y_T & \boldsymbol{X}_T^\top & Z_T \end{array}\right)$$

Then, when we observe all outcomes we can estimate the weights by the vertical regression

$$(\widehat{\boldsymbol{\beta}}, \widehat{\gamma}) = \underset{\boldsymbol{\beta}, \gamma}{\operatorname{argmin}} \|\boldsymbol{Y}_{T_0} - \mathbf{X}_{T_0}\boldsymbol{\beta} - \boldsymbol{Z}_{T_0}\gamma\|_2^2 \tag{1}$$

and estimate the treatment effect by

$$\widehat{\tau} = Y_T - (\boldsymbol{X}_T^\top \widehat{\boldsymbol{\beta}} + Z_T \widehat{\gamma}). \tag{2}$$

Note that when $T < N$, there are infinitely many solution to the problem. In that case, we interpret the estimated weights as the norm minimizing solution (see Spiess, Imbens and Venugopal, 2024).

2.2 SCM with Missing Values

Now, consider a case where $\boldsymbol{Z}_{T_0}$ contains missing values and the unit is thus excluded from the analysis. We estimate weights by the following restricted problem

$$\widehat{\boldsymbol{\beta}}_{\text{res}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \|\boldsymbol{Y}_{T_0} - \mathbf{X}_{T_0}\boldsymbol{\beta}\|_2^2 \tag{3}$$

and estimate the treatment effect based on these weights:

$$\widehat{\tau}_{\text{res}} = Y_T - \boldsymbol{X}_T^\top \widehat{\boldsymbol{\beta}}_{\text{res}}. \tag{4}$$

Theorem 1 (Bias Formula)**.**

$$\widehat{\tau}_{\text{res}} - \widehat{\tau} = \underbrace{\widehat{\gamma}}_{\textit{weight}} \cdot \underbrace{\left(Z_T - \widehat{\boldsymbol{\eta}}^\top \boldsymbol{X}_T\right)}_{\widehat{\delta} \,\equiv\, \textit{imbalance}} - \underbrace{\{(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \widehat{\boldsymbol{\epsilon}}^\top\} \boldsymbol{X}_T}_{\textit{noise term}} \tag{5}$$

where $\widehat{\boldsymbol{\eta}} \equiv (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0}$ *is the coefficient obtained from regressing the missing unit on the observed control units in the pre-treatment period, and* $\widehat{\boldsymbol{\epsilon}} \equiv \mathbf{Y}_{T_0} - \mathbf{X}_{T_0}\widehat{\boldsymbol{\beta}} - \boldsymbol{Z}_{T_0}\widehat{\gamma}$ *is the residual from the vertical regression in Equation* (1)*.*

Based on the result of Theorem 1, we can decompose the bias into two terms except the noise: weight ($\widehat{\gamma}$) and imbalance ($\widehat{\delta}$). The weight term is the fitted coefficient of the missing unit on the treated unit in the full model. In the context of SCM, each coefficient corresponds to the weight of donor unit on a synthetic control unit, therefore this term can be viewed as the "weight" of the missing unit for constructing the synthetic control of the treated unit. This implies that the bias reduced to the noise term when Z does not contribute to the construction of the synthetic control unit (i.e., $\widehat{\gamma} = 0$).

The imbalance term consists of the imbalance between the missing units Z_T and the weighted observed control units in the post-treatment period $\boldsymbol{X}_T$, where the weights ($\widehat{\boldsymbol{\eta}}$) are obtained from the coefficient regressing the missing unit $\boldsymbol{Z}_{T_0}$ on the observed control units in the pre-treatment period $\mathbf{X}_{T_0}$. This implies that the bias will be small when the missing observation is well approximated by the weighted sum of the observed control units, where those weights are determined by the pre-treatment fit between the observed control units and the missing unit.

In our motivating example of German reunification, we hypothetically excluded the USA from the analysis. The weight term corresponds to the significance of the USA in constructing synthetic Germany. Intuitively, the bias resulting from omitting the USA will be small if the USA's weight is small. The imbalance term refers to the extent to which the USA's data can be approximated by data from other countries, such as Switzerland or Japan. If the USA exhibits a similar outcome trend to other observed control units, it can be substituted by those units, which may lead to small bias.

Illustration of the parameters. We further explain the weight and imbalance terms using a geometric illustration, as shown in Figure 2. In this example, we assume two control units, Z_t and X_t, and one treated unit, Y_t, for $t = 1, \ldots, T$. Additionally, let the true DGP follow $Y_t = \beta X_t + \gamma Z_t + \epsilon_t$ for $t = 1, \ldots, T$. Now, we consider a situation where the control unit $\boldsymbol{Z}$ contains missing values and is thus excluded from SCM. We present two toy examples, Case 1 (left plot) and Case 2 (right plot), which differ in terms of the relationship between the missing unit and the observed unit. For each case, this specific relationship holds consistently for both the pre- and post-treatment periods.

In Case 1 (left plot), we assume that $\boldsymbol{Z}$ and $\boldsymbol{X}$ are orthogonal to each other, meaning that the missing unit and the observed unit are substantially different. In this case, the residual from regressing $\boldsymbol{Z}$ on $\boldsymbol{X}$ will be large, resulting in an imbalance between two ($\widehat{\delta} \neq 0$). To construct the treated unit $\boldsymbol{Y}$ (red dot), we would need both direction vectors, which implies $\widehat{\gamma} \neq 0$.

In Case 2 (right plot), we assume that $\boldsymbol{Z}$ and $\boldsymbol{X}$ are more correlated with each other compared to the previous example. Since $\boldsymbol{Z}$ and $\boldsymbol{X}$ are not in the same direction, the residual will be non-zero, resulting in an imbalance between the two ($\widehat{\delta} \neq 0$). Again, we would need both direction vectors to construct the treated unit, implying $\widehat{\gamma} \neq 0$. However, under this scenario, excluding $\boldsymbol{Z}$ potentially leads to less bias than for Case 1 due to the smaller magnitude of imbalance (more correlation).

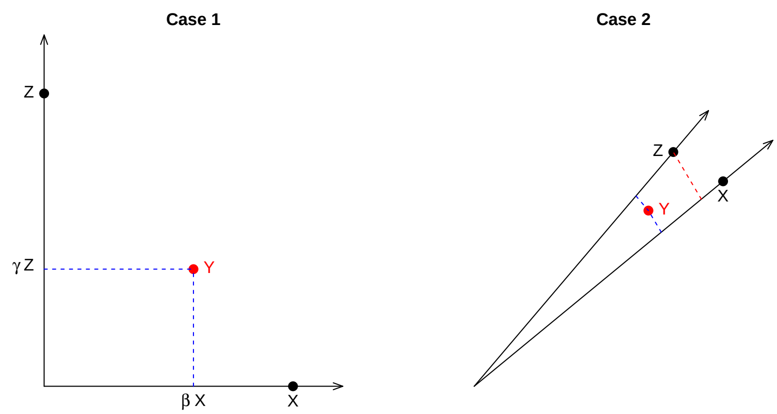

Figure 2: Two Toy Examples of Bias Decomposition. The two black direction vectors represent the outcomes of control units $\boldsymbol{Z}$ and $\boldsymbol{X}$. We assume that these direction hold consistently for both the pre- and post-treatment periods. The red dot represent the outcome of the treated unit, $\boldsymbol{Y}$. In Case 1 (left plot), we assume that $\boldsymbol{Z}$ and $\boldsymbol{X}$ are orthogonal. In Case 2 (right plot), those two are correalted.

In both Case 1 and Case 2, removing control unit Z will induce bias in the SCM result. Under the set up in Figure 2, $\boldsymbol{Z}$ and $\boldsymbol{X}$ would need to fall onto the same line to have a bias of zero. In other words, $\boldsymbol{Z}$ must not contain any additional information related to both $\boldsymbol{Y}$ and $\boldsymbol{X}$. In the following subsection, we further clarify the relationship between the two terms by utilizing two widely accepted DGPs of SCM.

2.3 Connection to Underlying DGPs of SCM

The bias decomposition suggests that when practitioners construct control units, they can conceptualize the bias resulting from omitting a control unit due to missingness in two components: weights (between treated and missing units) and imbalance (between observed control and missing units). It is important to note that the imbalance term is related to the imbalance of outcomes from the post-treatment period. By further expanding the bias terms based on the two underlying DGPs of the SCM — the autoregressive model and the linear factor model — we can address this dependence on post-treatment data and explore the relationship between the two bias terms and the conditions under which they are expected to be significant.

2.3.1 Autoregressive Model

We first consider the autoregressive model, one of the two underlying DGPs of SCM discussed in the original paper by Abadie, Diamond and Hainmueller (2010). It is well known that the SCM estimator is unbiased for ATT under this model, assuming a perfect pre-treatment fit of the synthetic control. Here, for simplicity, we omit the time-varying observed covariates from the model, but the discussion is generalizable to cases with covariates. Following Ben-Michael, Feller and Rothstein (2021), we consider a generalized model specification where an autoregressive model of order $p < T-1$ is a special case.

Assumption 1 (Autoregressive Model)**.** *The control potential outcomes at post-treatment period T is generated from the following model.*

$$\text{For treated unit,} \qquad Y_T(0) = \sum_{l=1}^{T-1} \omega_l Y_{t-l}(0) + \epsilon_T.$$

$$\text{For control units } i = 1, \dots, N-1, \qquad X_{iT}(0) = \sum_{l=1}^{T-1} \omega_l X_{i,t-l}(0) + \epsilon_{X,i,T}.$$

$$\text{For missing unit,} \qquad Z_T(0) = \sum_{l=1}^{T-1} \omega_l Z_{t-l}(0) + \epsilon_{Z,T}$$

where $\{\epsilon_T, \epsilon_{X,i,T}, \epsilon_{Z,T}\}$ satisfies the following strict exogeneity: $\mathbb{E}[\epsilon_T \mid \boldsymbol{Y}_{T_0}, \mathbf{X}_{T_0}, \boldsymbol{Z}_{T_0}] = 0$, $\mathbb{E}[\epsilon_{X,i,T} \mid \boldsymbol{Y}_{T_0}, \mathbf{X}_{T_0}, \boldsymbol{Z}_{T_0}] = 0 \ \forall i = 1, \dots, N-1$, and $\mathbb{E}[\epsilon_{Z,T} \mid \boldsymbol{Y}_{T_0}, \mathbf{X}_{T_0}, \boldsymbol{Z}_{T_0}] = 0$.

Recall that $X_{iT}(0) = X_{iT}$ for all $i = 1, \dots, N-1$ and $Z_T(0) = Z_T$ since control and missing units did not receive the treatment in the post-treatment period. In matrix form, the model for control and missing units can be written as follows.

$$\boldsymbol{X}_T(0) = \mathbf{X}_{T_0}^\top \boldsymbol{\omega} + \boldsymbol{\epsilon}_{X,T}, \tag{6}$$

$$Z_T(0) = \boldsymbol{Z}_{T_0}^\top \boldsymbol{\omega} + \epsilon_{Z,T} \tag{7}$$

where $\boldsymbol{\omega} = (\omega_1 \ \dots \ \omega_{T-1})^\top$ and $\boldsymbol{\epsilon}_{X,T} = (\epsilon_{X,1,T} \ \dots \ \epsilon_{X,N-1,T})^\top$.

Intuitively speaking, under the autoregressive model, we can represent the outcome in the post-treatment period as a linear combination of outcomes from the pre-treatment periods. The contribution of each pre-treatment period to the post-treatment outcome is captured by the autoregressive coefficient $\boldsymbol{\omega}$. Thus, based on this DGP, we can expand the post-treatment data in the expression of imbalance term from Theorem 1 using Equations (6) and (7). This expansion reveals a direct connection between the weight and imbalance terms.

Corollary 1 (Bias Formula under Autoregressive Model). *Under Assumption 1, the weight* $\widehat{\gamma}$ *and imbalance* $\widehat{\delta}$ *are given by*

$$\widehat{\gamma} = (\widehat{\boldsymbol{e}}\widehat{\boldsymbol{e}}^\top)^{-1}(\widehat{\boldsymbol{e}}\boldsymbol{Y}_{T_0})$$
$$\widehat{\delta} = \widehat{\boldsymbol{e}}\boldsymbol{\omega} + \underbrace{\epsilon_{Z,T} - \widehat{\boldsymbol{\eta}}^\top \boldsymbol{\epsilon}_{X,T}}_{\text{noise term}}$$

where $\widehat{\boldsymbol{e}} \equiv \boldsymbol{Z}_{T_0}^\top\{\mathbf{I} - \mathbf{X}_{T_0}(\mathbf{X}_{T_0}^\top\mathbf{X}_{T_0})^{-1}\mathbf{X}_{T_0}^\top\}$ *is a row vector of residuals from a vertical regression of the missing unit on the observed control units in the pre-treatment period.*

Corollary 1 demonstrates that each of the weight and imbalance terms can be expressed in terms of a multiplication of a residual vector $\widehat{\boldsymbol{e}}$ with $\boldsymbol{Y}_{T_0}$ or $\widehat{\boldsymbol{e}}$ with $\boldsymbol{\omega}$, respectively. Specifically, $\widehat{e}_t$ quantifies the extent to which the observed control units explain the missing unit in the pre-treatment period t. This can be viewed as proto-imbalance, which refers to the imbalance that does not account for the contribution of each pre-treatment period to the post-treatment outcome. For instance, in our motivating example, $\widehat{\boldsymbol{e}}$ represents the extent to which the GDP per capita of the USA can be explained by a fitted model of the GDP per capita of other control units and the regression coefficient $\widehat{\boldsymbol{\eta}}$ in each pre-treatment period. Thus, $\widehat{\boldsymbol{e}}\boldsymbol{Y}_{T_0}$ in the weight term $\widehat{\gamma}$ reflects how the treated unit's outcome aligns with the residual $\widehat{\boldsymbol{e}}$ over the pre-treatment period. Similarly, $\widehat{\boldsymbol{e}}\boldsymbol{\omega}$ in the imbalance term $\widehat{\delta}$ represents the weighted sum of the residual $\widehat{\boldsymbol{e}}$, where the weights ($\boldsymbol{\omega}$) represent the contribution of each pre-treatment period to the post-treatment outcome.

We further illustrate when the bias is expected to be large or small using toy examples. Suppose we have two pre-treatment periods and one observed control unit. Figure 3 visually illustrates two scenarios: (a) where the weight and the imbalance terms are large, and (b) where those are small. In each figure, x- and y-axes correspond to pre-treatment period 1 and 2 respectively, and thus the arrows correspond to the pre-treatment outcomes of the units (e.g. $\boldsymbol{X}_{T_0} = (5,1)^\top$). If the two arrows are parallel, this indicates that the two units share similar pre-treatment outcomes; if they are orthogonal, it indicates that the outcomes for the two units differ significantly. The two scenarios, (a) and (b), assume the same pre-treatment outcomes for the observed control unit (red arrow) and the missing unit (blue arrow). Note that the green arrow represents the residual $\widehat{\boldsymbol{e}}$ from projecting the missing unit onto the observed control unit, where the magnitude of $\widehat{e}_2$

is larger than that of $\widehat{e}_1$. That is, the missing unit's outcome is less well approximated by the observed control unit in period 2 compared to period 1.

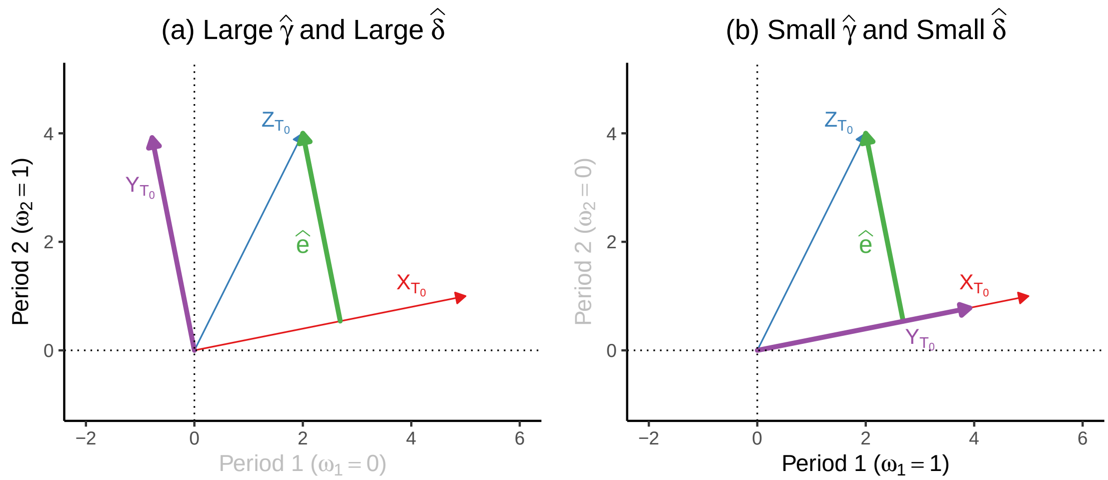

Figure 3: Toy Examples of Bias Decomposition under Autoregressive Model: Each panel represents a toy example where x- and y-axes correspond to pre-treatment period 1 and 2 respectively. The purple arrow represents the pre-treatment outcome of the treated unit ($\boldsymbol{Y}_{T_0}$), the red arrow represents that of the observed control unit ($\boldsymbol{X}_{T_0}$), and the blue arrow represents that of the missing unit ($\boldsymbol{Z}_{T_0}$). The green arrow represents $\widehat{\boldsymbol{e}}$, the residual vector from projecting the missing unit onto the observed control unit.

The two scenarios, however, differ in the pre-treatment outcome of the treated unit (purple arrow) and the autoregressive coefficient $\boldsymbol{\omega}$ (weight for period 1 and 2). This difference leads to variations in the weight and imbalance terms. In scenario (a), the treated unit and the observed control unit display significantly different pre-treatment outcomes, whereas the treated unit and the missing unit show comparatively similar pre-treatment outcomes. In fact, the treated unit is parallel to the residual $\widehat{\boldsymbol{e}}$ over the pre-treatment period, resulting in a large weight term $\widehat{\gamma}$. Also, since the weight of period 2 is large, the imbalance denoted by $\widehat{\boldsymbol{e}}$ is even more exaggerated in $\widehat{\delta}$. In contrast, in scenario (b), the treated unit and the observed control unit exhibit largely similar pre-treatment outcomes. Here, the treated unit is orthogonal to the residual $\widehat{\boldsymbol{e}}$ over the pre-treatment period, resulting in a small weight term. Additionally, since the weight of period 2 is small, the imbalance in period 2 is less pronounced.

In line with the geometric interpretation of this bias decomposition under autoregressive model, we can derive an upper bound on the bias (Proposition 1). In simpler terms, the bias is at its largest when $\widehat{\boldsymbol{e}}$ and $\boldsymbol{\omega}$ are parallel, indicating that the residual of missing unit, which cannot be explained by the observed donor unit, is large for those pre-treatment periods that contribute most to the post-treatment outcome; and if $\boldsymbol{Y}_{T_0}$ and $\widehat{\boldsymbol{e}}$ are parallel, meaning that the pre-treatment outcome of treated unit can be perfectly explained by the residual of the missing unit that the observed donor unit cannot account for. In practice, we do not know the true $\boldsymbol{\omega}$ and thus cannot directly observe the upper bound of the bias; however this can be estimated using the outcome variables of the observed control unit (i.e. $\widehat{\boldsymbol{\omega}}$).

Proposition 1 (Bound on the Bias under Autoregressive Model)**.** *Under Assumption 1, by Cauchy-Schwarz inequality we have,*

$$\mathbb{E}\left[\widehat{\delta\gamma}\right] \leq \|\boldsymbol{\omega}\|_2 \|\boldsymbol{Y}_{T_0}\|_2$$

where $\|\cdot\|_2$ *denotes* ℓ_2 *norm and the equality holds if* $\{\widehat{\boldsymbol{e}}, \boldsymbol{\omega}\}$ *is linearly dependent, and so does* $\{\widehat{\boldsymbol{e}}, \boldsymbol{Y}_{T_0}\}$.

2.3.2 Linear Factor Model

Next, we consider the linear factor model, under which the bias of the SCM estimator can be bounded by a function that converges to zero as the number of pre-treatment periods increases (Abadie, Diamond and Hainmueller, 2010). Again, we omit the time-varying observed covariates from the model, and follow the model specification from Ben-Michael, Feller and Rothstein (2021).

Assumption 2 (Linear Factor Model)**.** *Let* $\{\mu_{kt}\}_{k=1}^K$ *denote* K *latent time-varying factors at time* t *and* $\{\phi_k, \phi_{X,i,k}, \phi_{Z,k}\}_{k=1}^K$ *denote latent factor loadings. The control potential outcomes at time* $t = 1, \ldots, T$ *is generated from the following model.*

For treated unit, $$Y_t(0) = \sum_{k=1}^K \phi_k \mu_{kt} + \epsilon_t.$$

For control units $i = 1, \ldots, N-1$, $$X_{it}(0) = \sum_{k=1}^K \phi_{X,i,k} \mu_{kt} + \epsilon_{X,i,t}.$$

For missing unit, $$Z_t(0) = \sum_{k=1}^K \phi_{Z,k} \mu_{kt} + \epsilon_{Z,t}$$

where $\{\epsilon_t, \epsilon_{X,i,t}, \epsilon_{Z,t}\}$ *satisfies the following strict exogeneity:* $\mathbb{E}[\epsilon_t \mid \{\mu_{kt}, \phi_k, \phi_{X,i,k}, \phi_{Z,k}\}] = 0$, $\mathbb{E}[\epsilon_{X,i,t} \mid \{\mu_{kt}, \phi_k, \phi_{X,i,k}, \phi_{Z,k}\}] = 0 \ \forall i = 1, \ldots, N-1,$ *and* $\mathbb{E}[\epsilon_{Z,t} \mid \{\mu_{kt}, \phi_k, \phi_{X,i,k}, \phi_{Z,k}\}] = 0.$

In addition to the model specification, we impose constraints on the structure of loadings and factors for identification. Let $\boldsymbol{\mu}_t = (\mu_{1t} \ \dots \ \mu_{Kt})^\top$, and $\boldsymbol{\mu}$ denote a $T-1$ by K matrix of which t-th row is $\boldsymbol{\mu}_t^\top$, the factors of the pre-treatment period T_0. Similarly, let $\boldsymbol{\phi}_{X,i} = (\phi_{X,i,1} \ \dots \ \phi_{X,i,K})^\top$, $\boldsymbol{\phi}_Z = (\phi_{Z,1} \ \dots \ \phi_{Z,K})^\top$ and $\boldsymbol{\phi}_X$ denote a K by $N-1$ matrix, of which i-th column is $\boldsymbol{\phi}_{X,i}$. Following Bai (2009) and Xu (2017), we impose two sets of constraints for the identification of the model without loss of generality: (1) all factor $\boldsymbol{\mu}_t$ are normalized, and they are orthogonal to each other (i.e. $\boldsymbol{\mu}^\top \boldsymbol{\mu}/(T-1) = \mathbf{I}$); (2) $[\boldsymbol{\phi}_X \ \boldsymbol{\phi}_Z][\boldsymbol{\phi}_X \ \boldsymbol{\phi}_Z]^\top$ is a diagonal matrix (implying that the rows of $[\boldsymbol{\phi}_X \ \boldsymbol{\phi}_Z]$ are orthogonal, or in other words factor loadings of control units are orthogonal across k).

Our goal here is to further expand the bias formula using the model to clarify the relationship between two bias terms, as we have shown with the autoregressive model. Under linear factor model, we have two sets of model parameters: latent time-varying factors $\boldsymbol{\mu}_t$ and factor loadings — $\boldsymbol{\phi}$, $\boldsymbol{\phi}_{X,i}$, and $\boldsymbol{\phi}_{Z,i}$, respectively for treated, observed control, and missing units. As with the previous model, we can express the post-treatment data $\boldsymbol{X}_T$ and $\boldsymbol{Z}_T$ in our bias formula using these two sets of model parameters. Furthermore, using the assumed model and the constraint $\boldsymbol{\mu}^\top \boldsymbol{\mu}/(T-1) = \mathbf{I}$, we can represent the factor loadings in terms of pre-treatment period data, which enables us to further clarify the connection between the weight and imbalance terms:

$$
\begin{aligned}
\boldsymbol{\phi}_X &= \frac{1}{T-1}\boldsymbol{\mu}^\top(\mathbf{X}_{T_0} - \boldsymbol{\epsilon}_X) \\
\boldsymbol{\phi}_Z &= \frac{1}{T-1}\boldsymbol{\mu}^\top(\boldsymbol{Z}_{T_0} - \boldsymbol{\epsilon}_Z)
\end{aligned}
$$

where $\boldsymbol{\epsilon}_X$ is a $T-1$ by $N-1$ matrix of which t-th row is $\boldsymbol{\epsilon}_{X,t} = (\epsilon_{X,1,t} \ \dots \ \epsilon_{X,N-1,t})^\top$, and $\boldsymbol{\epsilon}_Z = (\epsilon_{Z,t} \ \dots \ \epsilon_{Z,t})^\top$.

Corollary 2 (Bias Formula under Linear Factor Model)**.** *Under Assumption 2, the weight $\widehat{\gamma}$ and imbalance $\widehat{\delta}$ are given by*

$$
\begin{aligned}
\widehat{\gamma} &= (\widehat{\boldsymbol{e}}\widehat{\boldsymbol{e}}^\top)^{-1}(\widehat{\boldsymbol{e}}\boldsymbol{Y}_{T_0}) \\
\widehat{\delta} &= \frac{1}{T-1}\widehat{\boldsymbol{e}}\boldsymbol{\mu}\boldsymbol{\mu}_T - \underbrace{\frac{1}{T-1}(\boldsymbol{\epsilon}_Z - \boldsymbol{\epsilon}_X\widehat{\boldsymbol{\eta}})^\top\boldsymbol{\mu}\boldsymbol{\mu}_T + \epsilon_{Z,T} - \widehat{\boldsymbol{\eta}}^\top\boldsymbol{\epsilon}_{X,T}}_{\text{noise term}}
\end{aligned}
$$

where $\widehat{\boldsymbol{e}} \equiv \boldsymbol{Z}_{T_0}^\top\{\mathbf{I} - \mathbf{X}_{T_0}(\mathbf{X}_{T_0}^\top\mathbf{X}_{T_0})^{-1}\mathbf{X}_{T_0}^\top\}$ is a row vector of residuals from a vertical regression of the missing unit on the observed control units in the pre-treatment period.

Corollary 2 presents a similar result to that of the autoregressive model. Under the linear factor model, the weight term remains as before: a multiplication of a residual vector $\widehat{\boldsymbol{e}}$ with $\boldsymbol{Y}_{T_0}$, normalized by the length of $\widehat{\boldsymbol{e}}$. The imbalance term is now the product of the residual vector $\widehat{\boldsymbol{e}}$ with model parameters $\boldsymbol{\mu}\boldsymbol{\mu}_T$, normalized by $\boldsymbol{\mu}^\top\boldsymbol{\mu}$. Here, $\boldsymbol{\mu}\boldsymbol{\mu}_T$ can be interpreted as a weighted sum of latent time-varying factors in the pre-treatment period ($\boldsymbol{\mu}$), where the weights are determined by how much k-th factor contributes to the post-treatment outcome at time T ($\boldsymbol{\mu}_T$). Note that imbalance term can also be expressed as

$$\widehat{\delta} = (\boldsymbol{\phi}_Z - \boldsymbol{\phi}_X\widehat{\boldsymbol{\eta}})^\top\boldsymbol{\mu}_T + \underbrace{\epsilon_{Z,T} - \widehat{\boldsymbol{\eta}}^\top\boldsymbol{\epsilon}_{X,T}}_{\text{noise term}}$$

where $\boldsymbol{\phi}_Z - \boldsymbol{\phi}_X\widehat{\boldsymbol{\eta}}^\top$ signifies imbalance between the factor loadings of the missing unit and the observed donor units, weighted by $\widehat{\boldsymbol{\eta}}$.

As a result, similar implications can be derived regarding when the bias is expected to be large or small, akin to what is shown in Figure 3. The only change required is substituting the model parameter $\boldsymbol{\omega}$ with $\boldsymbol{\mu}\boldsymbol{\mu}_T$. Analogously, we can derive the upper bound on the bias.

Proposition 2 (Bound on the Bias under Linear Factor Model)**.** *Under Assumption 2, by Cauchy-Schwarz inequality we have,*

$$\mathbb{E}\big[|\widehat{\delta}\widehat{\gamma}|\big] \leq \frac{1}{T-1}\|\boldsymbol{\mu}\boldsymbol{\mu}_T\|_2\|\boldsymbol{Y}_{T_0}\|_2$$

where the equality holds if $\{\widehat{\boldsymbol{e}}, \boldsymbol{\mu}\boldsymbol{\mu}_T\}$ *is linearly dependent, and so does* $\{\widehat{\boldsymbol{e}}, \boldsymbol{Y}_{T_0}\}$.

The main takeaway from the bias derivation under two DGPs is that a good pre-treatment fit without a missing unit — indicated by a small weight term $\widehat{\gamma} = (\widehat{\boldsymbol{e}}\widehat{\boldsymbol{e}}^\top)^{-1}(\widehat{\boldsymbol{e}}\boldsymbol{Y}_{T_0})$ — does not guarantee a negligible amount of bias in the SCM estimate. Even with a small weight term, a significant imbalance between the observed control unit and missing unit at time t (large $\widehat{e}_t$), and aggregating those over t that notably contributes to the post-treatment outcome under such DGP (large ω_t or $\boldsymbol{\mu}_t\boldsymbol{\mu}_T$), can result in a non-negligible amount of bias in the SCM estimate.

Furthermore, imputing the missing values and estimating the treatment effect may not guarantee small bias in cases where there is an imbalance in factor loadings. In Appendix B, we conduct simulation studies to evaluate bias from vertical regression with missing outcomes using simulated data generated with the linear factor model. In light of these results, we propose a sensitivity analysis based on the bias derivation in the following section.

3 Sensitivity Analysis

3.1 Equivalence between Two Vertical Regression Approaches

In this section, we demonstrate that the bias from the restricted vertical regression in Equation (3) is equivalent to the omitted variable bias in the regression coefficient in the linear regression. This equivalence result allows us to connect the bias analysis in SCM with the existing tools that extends the omitted variable bias framework for the sensitivity analysis for unobserved confounders in the causal inference literature (Cinelli and Hazlett, 2020). As emphasized by Cinelli and Hazlett (2020), the familiar omitted variable bias framework offers a simple yet intuitive tools for the routine reporting of how robust a result is against the potential for unobserved confounding. We show that practitioners can directly utilize these tools in a vertical regression format to report how robust their SCM estimate is against the bias resulting from the omission of the missing unit.

To begin with, we consider the following alternative vertical regression with additional treatment indicator variable D_t:

$$\left(\widehat{\tau}^*, \widehat{\boldsymbol{\beta}}^*, \widehat{\gamma}^*\right) = \underset{\tau,\boldsymbol{\beta},\gamma}{\operatorname{argmin}} \|\boldsymbol{Y} - \boldsymbol{D}\tau - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{Z}\gamma\|_2^2 \tag{8}$$

where we have augmented $\boldsymbol{D}$ into the data

$$\left(\begin{array}{c|c|ccc|c} Y_1 & 0 & X_{11} & \cdots & X_{N-1,1} & Z_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ Y_{T-1} & 0 & X_{1,T-1} & \cdots & X_{N-1,T-1} & Z_{T-1} \\ Y_T & 1 & X_{1T} & \cdots & X_{N-1,T} & Z_T \end{array}\right) = \begin{bmatrix} \boldsymbol{Y} & \boldsymbol{D} & \mathbf{X} & \boldsymbol{Z} \end{bmatrix}.$$

Here, $D_T = 1$ (post-treatment period), and $D_t = 0$ for $t = 1, \ldots, T-1$ (pre-treatment period). Note that the alternative vertical regression includes observations from the post-treatment period, whereas previously in Equation (1) we only include observations from the pre-treatment periods. In Proposition 3, we first formally establish that the regression presented here with augmented $\boldsymbol{D}$ and the vertical regression are algebraically equivalent, in that the two approaches produce numerically identical fitted coefficients and treatment effect estimates.

Proposition 3 (Equivalence of Two Vertical Regressions)**.** *The fitted coefficient of the treatment indicator $\boldsymbol{D}$ in Equation (8) is algebraically equivalent to the treatment effect estimate of the* *vertical regression in Equation* (1)

$$\widehat{\tau}^* = \widehat{\tau}$$

where $\widehat{\tau} = Y_T - (\boldsymbol{X}_T^\top \widehat{\boldsymbol{\beta}} + Z_T \widehat{\gamma})$ *is given in Equation* (2)*. Additionally, the fitted coefficients of* $\mathbf{X}$ *and* $\boldsymbol{Z}$ *in Equation* (8) *is algebraically equivalent to those of the vertical regression in Equation* (1)*.*

$$\widehat{\boldsymbol{\beta}}^* = \widehat{\boldsymbol{\beta}}, \quad \text{and} \quad \widehat{\gamma}^* = \widehat{\gamma}.$$

Similarly, we consider the restricted version of the alternative vertical regression, where $\boldsymbol{Z}$ is not included in the regression due to missing data. We then show the equivalence of the fitted coefficient with that of the restricted version of the previous vertical regression model.

$$\left(\widehat{\tau}^*_{\text{res}}, \widehat{\boldsymbol{\beta}}^*_{\text{res}}\right) = \underset{\tau,\ \boldsymbol{\beta}}{\operatorname{argmin}} \|\boldsymbol{Y} - \boldsymbol{D}\tau - \mathbf{X}\boldsymbol{\beta}\|_2^2 \tag{9}$$

Corollary 3 (Equivalence of Two Restricted Vertical Regressions)**.** *The fitted coefficient of the treatment indicator* $\boldsymbol{D}$ *in Equation* (9) *is algebraically equivalent to the treatment effect estimate of the vertical regression in Equation* (3)

$$\widehat{\tau}^*_{res} = \widehat{\tau}_{res}$$

where $\widehat{\tau}_{res} = Y_T - \boldsymbol{X}_T^\top \widehat{\boldsymbol{\beta}}_{res}$ *is given in Equation* (4)*.*

The bias in $\widehat{\tau}^*_{\text{res}}$ in Equation (9) with respect to $\widehat{\tau}^*$ is known as the omitted variable bias, as it originates from omitting $\boldsymbol{Z}$ from the regression. In the next section, we apply the classical omitted variable bias solution to the bias in $\widehat{\tau}^*_{\text{res}}$ following Cinelli and Hazlett (2020), upon which we derive the basic formula for a simple sensitivity analysis with two parameters. Subsequently, in light of the results from Proposition 3 and Corollary 3, we demonstrate the equivalence of the two sensitivity parameters with the weight and imbalance terms discussed in the previous section. This allows us to apply the same tools used in sensitivity analysis for unobserved confounders to this alternative missing data problem in SCM.

3.2 Sensitivity Parameters

In this section, following Cinelli and Hazlett (2020), we apply the classical omitted variable bias formula to expand the bias in $\widehat{\tau}^*_{\text{res}}$ with respect to $\widehat{\tau}^*$, and illustrate the sensitivity analysis using two parameters based on this bias formula. As a direct result of the equivalence between two vertical regression approaches, we show that the two sensitivity parameters derived from omitted variable bias are equivalent to the weight and imbalance terms discussed in the previous section. Building on this, we illustrate how the existing tools such as bivariate contour plot designed for the sensitivity analysis of unobserved confounders, can also be utilized in our framework.

We start by applying Frisch-Waugh-Lovell theorem to derive the bias in $\widehat{\tau}^*_{\text{res}}$:

$$\begin{aligned}\widehat{\tau}^*_{\text{res}} &= \frac{\text{cov}\left(\boldsymbol{D}^{\perp \mathbf{X}}, \boldsymbol{Y}^{\perp \mathbf{X}}\right)}{\text{var}\left(\boldsymbol{D}^{\perp \mathbf{X}}\right)} \\ &= \widehat{\tau}^* + \widehat{\gamma}^* \frac{\text{cov}\left(\boldsymbol{D}^{\perp \mathbf{X}}, \boldsymbol{Z}^{\perp \mathbf{X}}\right)}{\text{var}\left(\boldsymbol{D}^{\perp \mathbf{X}}\right)} \\ &= \widehat{\tau}^* + \widehat{\gamma}^* \widehat{\delta}^* \end{aligned} \tag{10}$$

where $\text{cov}(\cdot)$ and $\text{var}(\cdot)$ denote the sample covariance and variance, and $\boldsymbol{D}^{\perp \mathbf{X}}$ is the residualized variable after partialing out the components linearly explained by observed control units $\mathbf{X}$, similarly for other quantities. Here, we define $\widehat{\delta}^* \equiv \text{cov}\left(\boldsymbol{D}^{\perp \mathbf{X}}, \boldsymbol{Z}^{\perp \mathbf{X}}\right)/\text{var}\left(\boldsymbol{D}^{\perp \mathbf{X}}\right)$.

Theorem 2 (Equivalence of Bias Decomposition)**.** *The omitted variable bias in Equation* (10) *is algebraically equivalent to bias decomposition in Theorem 1*

$$\widehat{\delta}^* = \widehat{\delta}, \quad \textit{and} \quad \widehat{\gamma}^* = \widehat{\gamma}.$$

where $\widehat{\gamma} = Z_T - \widehat{\boldsymbol{\eta}}^\top \boldsymbol{X}_T$ *is given in Equation* (5)*.*

By Theorem 2, the classical omitted variable bias formula can be seen as an alternative way of deriving the bias decomposition in Theorem 1. As stated earlier, one strength of connecting the bias to omitted variable bias is that we can use the same tools designed for sensitivity analysis of unobserved confounders as outlined in Cinelli and Hazlett (2020), using two sensitivity parameters: the weight term ($\widehat{\gamma}$) and the imbalance term ($\widehat{\delta}$). We illustrate how to utilize sensitivity contour plot, partial R^2 reparameterization, and robustness value in addressing this missing data problem with SCM, using the motivating example of West German reunification.

Sensitivity contour plot. In the motivating example, we are interested in the bias specification of the SCM estimate regarding the effect of the 1990 German reunification on West Germany.

The vertical regression model, augmented with $\boldsymbol{D}$ in this setup, is as follows:

$$\begin{aligned}\text{W Germany}_t \sim \widehat{\tau}^*_{\text{res}} D_t &+ \widehat{\beta}^*_{\text{res},1} \cdot \text{Australia}_t + \widehat{\beta}^*_{\text{res},2} \cdot \text{Austria}_t + \widehat{\beta}^*_{\text{res},3} \cdot \text{Belgium}_t \\ &+ \widehat{\beta}^*_{\text{res},4} \cdot \text{Denmark}_t + \widehat{\beta}^*_{\text{res},5} \cdot \text{France}_t + \widehat{\beta}^*_{\text{res},6} \cdot \text{Greece}_t \\ &+ \widehat{\beta}^*_{\text{res},7} \cdot \text{Italy}_t + \widehat{\beta}^*_{\text{res},8} \cdot \text{Japan}_t + \widehat{\beta}^*_{\text{res},9} \cdot \text{Netherlands}_t \\ &+ \widehat{\beta}^*_{\text{res},10} \cdot \text{New Jealand}_t + \widehat{\beta}^*_{\text{res},11} \cdot \text{Norway}_t + \widehat{\beta}^*_{\text{res},12} \cdot \text{Portugal}_t \\ &+ \widehat{\beta}^*_{\text{res},13} \cdot \text{Spain}_t + \widehat{\beta}^*_{\text{res},14} \cdot \text{Switzerland}_t + \widehat{\beta}^*_{\text{res},15} \cdot \text{UK}_t \\ &+ \widehat{\beta}^*_{\text{res},16} \cdot \text{USA}_t\end{aligned}$$

for $t = 1, \ldots, T$ where W Germany$_t$ denotes the West German GDP per capita at period t, and similarly for other variables. Our ultimate goal is to quantify how robust our estimate of $\widehat{\tau}$ is against the bias that results from omitting the control unit, which can be equated to omitting a variable in this vertical regression context.

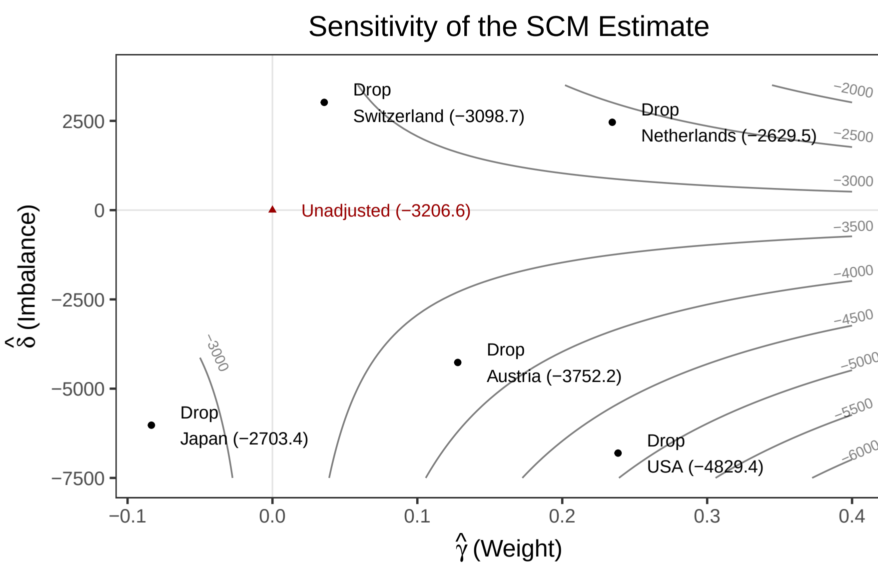

Figure 4: Sensitivity Contour Plot of the West German Reunification Example Using Vertical Regression. In this figure, the x-axis shows hypothetical value of weight sensitivity parameter ($\widehat{\gamma}$) and the x-axis shows that of imbalance ($\widehat{\delta}$). The contour line illustrates how the adjusted SCM estimate would be under those sensitivity parameters, i.e. $\widehat{\tau}^*_{\text{res}} - \widehat{\gamma\delta}$. The red triangle represents the unadjusted treatment estimate $\widehat{\tau}^*_{\text{res}}$, i.e. the estimate from a model including all the control units and assuming that $(\widehat{\gamma}, \widehat{\delta}) = (0, 0)$. Each circle presents the weight and imbalance terms under the case where we omit one of the observed control units from our analysis. The numbers inside the parentheses indicate the adjusted treatment estimate with those sensitivity parameters (the same quantity as shown in the contour line).

In this example, the treatment effect estimate ($\widehat{\tau}^*_{\text{res}}$) in 2003 using the vertical regression with the entire control units is -3206.6 (per capita GDP, 2002 USD). Figure 4 investigates the sensitivity of this treatment effect estimate to hypothetical values of sensitivity parameters on the x- and y-axes. Specifically, this figure shows the hypothetical "true" treatment effect estimate (contour line) after adjusting for bias, i.e. $\widehat{\tau}^*_{\text{res}} - \widehat{\gamma}\widehat{\delta}$ where $\widehat{\gamma}$ is the value on the x-axis and $\widehat{\delta}$ on the y-axis. Here, the red triangle indicates the case with unadjusted treatment effect, where we assume zero weight and imbalance terms.

To facilitate the interpretation of hypothetical sensitivity parameters, we also include reference points in Figure 4 where we omit one of the observed control units from the analysis, and estimate the weight and imbalance terms resulting from such an omission. For example, if we omit the USA from our vertical regression model (orange circle), the weight of the USA in constructing the synthetic West Germany is 0.24 and the imbalance between the USA and the other control units is -6804.6, leading to a total bias of -1622.8 (2002 USD). Thus, if there were a missing unit similar to the USA in terms of weight and imbalance parameters, the adjusted estimate, after accounting for the omission of such a unit will be -4829.4. In the appendix Section C.2, we present the results for an additional post-treatment period, from 1990 to 2002.

Partial R^2 parameterization. Cinelli and Hazlett (2020) consider a reparameterization of the omitted variable bias in terms of partial R^2 values. The benefit of this reparameterization is that it is scale free and allows us to construct some useful analyses including sensitivity of t-values.

$$|\widehat{\tau}^*_{\text{res}} - \widehat{\tau}^*| = \text{se}\,(\widehat{\tau}^*_{\text{res}}) \,/\, \left(\frac{R^2_{\boldsymbol{Y}\sim\boldsymbol{Z}|\boldsymbol{D},\mathbf{X}} R^2_{\boldsymbol{Z}\sim\boldsymbol{D}|\mathbf{X}}}{1 - R^2_{\boldsymbol{Z}\sim\boldsymbol{D}|\mathbf{X}}} \text{df} \right)$$

where df denote the degrees of freedom of the restricted regression. Note that we used $R^2_{\boldsymbol{Z}\sim\boldsymbol{D}|\mathbf{X}}$ instead of $R^2_{\boldsymbol{D}\sim\boldsymbol{Z}|\mathbf{X}}$ using the symmetry of the partial R^2 values to avoid the specification of the treatment model. Following Cinelli and Hazlett (2020), we can also derive relative change in variance as $\text{df}(1 - R^2_{\boldsymbol{Y}\sim\boldsymbol{Z}|\boldsymbol{D},\mathbf{X}})/\{(\text{df}-1)(1 - R^2_{\boldsymbol{Z}\sim\boldsymbol{D}|\mathbf{X}})\}$, which can be used to create a sensitivity contour plot of the t-statistic. An example of the sensitivity contour plot using partial R^2 values is presented in Appendix Figure 12.

4 Empirical Illustration

Lipscy and Lee (2019) claim that International Monetary Fund (IMF) creates moral hazard asymmetrically, thereby reducing the expected costs of risky lending and policies for states that hold political influence over the institution. To substantiate the causal assertions, authors applied SCM to the case of Taiwan, which was expelled from the IMF in 1980. Original findings suggest that Taiwan's expulsion prompted a notable increase in precautionary international reserves and the adoption of exceptionally conservative financial policies (Lipscy and Lee, 2019). The replication of the result using the vertical regression approach, with international reserves as a share of GDP, shows the same findings, with a treatment effect of 0.38 in 1990, as presented in Figure 5.

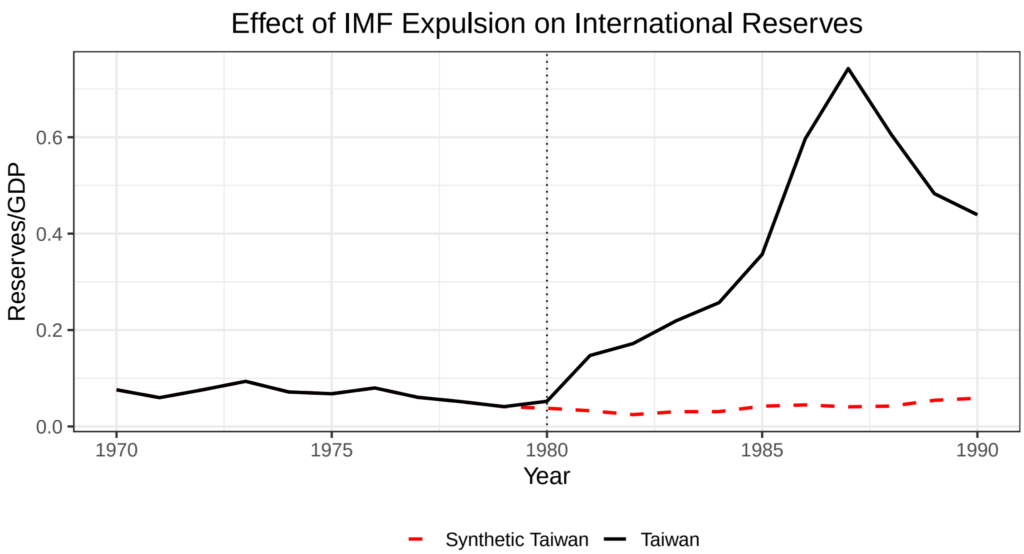

Figure 5: The SCM Estimate of the Impact of Taiwan's Expulsion from the IMF on International Reserves (Lipscy and Lee, 2019) using Vertical Regression Estimator. The black solid line indicates Taiwan's international reserves as a share of GDP, and the red dashed line represents that of synthetic Taiwan. In this example, we have one treated unit, 54 control units, and 10 pre-treatment periods, and we use the pseudo-inverse to compute the ordinary least square estimate.

The original analysis utilizes panel data spanning from 1970 to 1990, encompassing total 193 countries, with the treatment period for Taiwan being 1980. However, the dataset contains a substantial number of missing entries, as illustrated in Figure 6. In the left plot of this figure, we present the histogram of the missingness rate of outcomes across pre-treatment periods for each country available in the dataset. Among the 193 countries, 75 have completely missing pre-treatment outcomes, making it impossible to include them in the donor pool. Ninety-eight countries have completely observed pre-treatment outcomes, and 55 of these were included in the original study as they have no missing values in other predictors used in the SCM analysis. The remaining 20 countries including China have partially missing pre-treatment outcomes, and their missingness patterns are visualized in the right plot of Figure 6. Using our proposed method, we can further utilize these partially observed countries, which would otherwise be discarded, to estimate the bias driven by omitting each country. This allows researchers to use each case as a reference point in the sensitivity analysis.

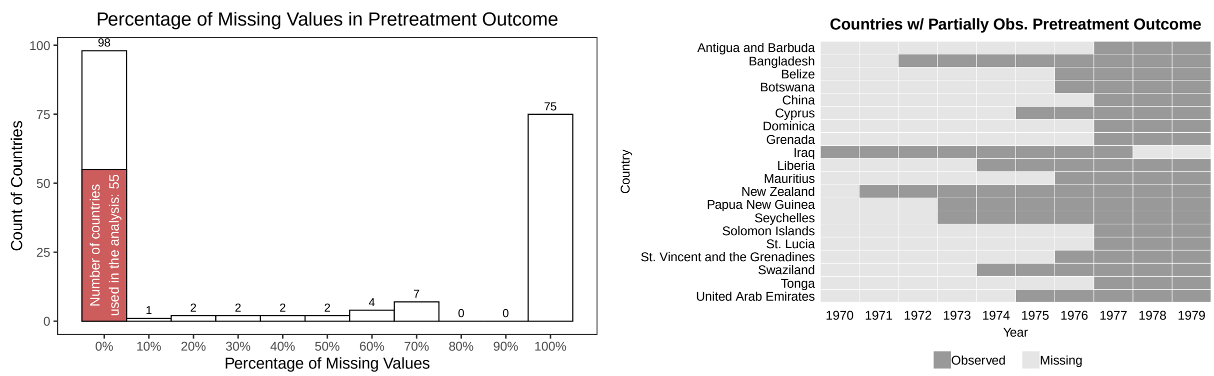

Figure 6: Missing Data in Pre-treatment Outcomes of the Taiwan-IMF Study. The left plot shows the histogram of the missingness rate of outcomes across pre-treatment periods for each country. The right plot shows the missingness pattern of 20 countries with partially missing pre-treatment outcomes. The analysis conducted by Lipscy and Lee (2019) opted to discard 138 countries with any missing values in predictor variables during the pre- to post-treatment periods, resulting in a reduced sample size of 55 countries.

Missing pre-treatment outcomes might not occur completely at random. For instance, less developed countries may be more susceptible to missing data compared to developed countries due to structural reasons or a lack of resources for data collection. Consequently, it is imperative for researchers to conduct a sensitivity analysis to demonstrate how the results could potentially change. In Figure 7, we present the sensitivity analysis of the treatment effect in 1990 with respect to the bias resulting from the missing unit. Here, the x-axis represents the hypothetical value of the weight of the missing country for constructing a synthetic Taiwan, and the y-axis represents the imbalance between the missing country and other observed control countries. The contour line illustrates the adjusted estimate of the impact of Taiwan's expulsion from the IMF on its precautionary international reserves accounting for such bias. The contour value at the red triangle denotes the unadjusted estimate, which is equivalent to the replicated result using complete cases and vertical regression. To ease the interpretation of sensitivity parameters, we provide reference points by estimating, with their partially observed data, the expected bias resulting from the omission of dropped countries. For instance, if we include China in the analysis, which shows a marginal estimated weight and imbalance, the adjusted treatment effect remains close to the unadjusted estimate, demonstrating the robustness of the SCM estimate against the bias from missing China. The robustness of the results to missing units is consistent across other post-treatment periods, as shown in Appendix D.

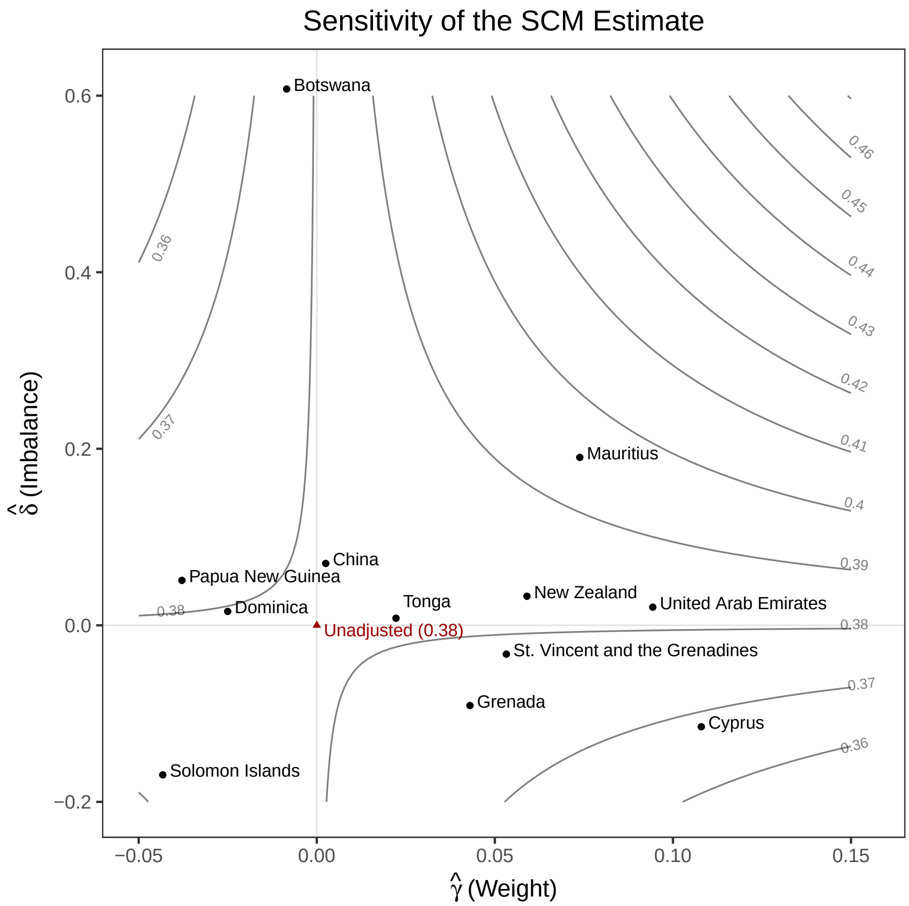

Figure 7: Sensitivity Contour Plot of the Taiwan-IMF Study Using Vertical Regression. In this figure, the x-axis shows hypothetical value of weight sensitivity parameter ($\widehat{\gamma}$) and the x-axis shows that of imbalance ($\widehat{\delta}$). The contour line illustrates how the adjusted SCM estimate would be under those sensitivity parameters, i.e. $\widehat{\tau}^*_{\text{res}} - \widehat{\gamma}\widehat{\delta}$. The red triangle represents the unadjusted treatment estimate $\widehat{\tau}^*_{\text{res}}$, i.e. the estimate from a model including all the control units and assuming that $(\widehat{\gamma}, \widehat{\delta}) = (0, 0)$. Each circle presents the estimated weight and imbalance terms for each country with partially observed pre-treatment data.

5 Conclusion

By examining synthetic control method through the lens of vertical regression, we present a comprehensive decomposition of bias, shedding light on the conditions under which the treatment effect estimates are robust to different specification of control units. This decomposition allows for a nuanced understanding of the sources of potential bias inherent in SCM estimation due to the missingness, thereby facilitating a simple and intuitive sensitivity analysis. Following the theoretical framework outlined above, we proceed to conduct an empirical study aimed at illustrating the potential bias stemming from the omission of control units in research.

Our sensitivity analysis offers valuable insights for scholars using the SCM method, providing a concise overview of how missing pre-treatment outcomes may influence study results. Unless one strongly believes that missing values follow a completely random process, conducting sensitivity analysis is always recommended for SCM studies. Expanding upon the bias decomposition and sensitivity analysis proposed in this paper, future research could delve into exploring the connection between bias and missingness mechanisms, as well as comparing different methods for addressing missingness in SCM analysis, such as dropping the time period instead of the unit or imputation with zeros.

References

Abadie, Alberto, Alexis Diamond and Jens Hainmueller. 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, Alberto, Alexis Diamond and Jens Hainmueller. 2015. "Comparative politics and the synthetic control method." *American Journal of Political Science* 59(2):495–510.

Abadie, Alberto and Javier Gardeazabal. 2003. "The economic costs of conflict: A case study of the Basque Country." *American economic review* 93(1):113–132.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens and Khashayar Khosravi. 2021*a*. "Matrix completion methods for causal panel data models." *Journal of the American Statistical Association* 116(536):1716–1730.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens and Khashayar Khosravi. 2021*b*. "Matrix completion methods for causal panel data models." *Journal of the American Statistical Association* 116(536):1716–1730.

Bai, Jushan. 2009. "Panel data models with interactive fixed effects." *Econometrica* 77(4):1229–1279.

Ben-Michael, Eli, Avi Feller and Jesse Rothstein. 2021. "The augmented synthetic control method." *Journal of the American Statistical Association* 116(536):1789–1803.

Bischof, Daniel and Markus Wagner. 2019. "Do voters polarize when radical parties enter parliament?" *American Journal of Political Science* 63(4):888–904.

Blair, Christopher W, Guy Grossman and Jeremy M Weinstein. 2022. "Liberal displacement policies attract forced migrants in the global south." *American Political Science Review* 116(1):351–358.

Cinelli, Carlos and Chad Hazlett. 2020. "Making sense of sensitivity: Extending omitted variable bias." *Journal of the Royal Statistical Society Series B-Statistical Methodology* 82(1):39–67.

De Kadt, Daniel, Ada Johnson-Kanu and Melissa L Sands. 2020. "State violence, party formation, and electoral accountability: The political legacy of the Marikana massacre." *American Political Science Review* pp. 1–21.

Eibl, Ferdinand and Steffen Hertog. 2023. "From Rents to Welfare: Why Are Some Oil-Rich States Generous to Their People?" *American Political Science Review* pp. 1–20.

Gilens, Martin, Shawn Patterson and Pavielle Haines. 2021. "Campaign finance regulations and public policy." *American Political Science Review* 115(3):1074–1081.

Grumbach, Jacob M. 2023. "Laboratories of democratic backsliding." *American Political Science Review* 117(3):967–984.

Lipscy, Phillip Y. and Haillie Na-Kyung Lee. 2019. "The IMF as a biased global insurance mechanism: Asymmetrical moral hazard, reserve accumulation, and financial crises." *International Organization* 73(1):35–64.

Marble, William, Salma Mousa, Alexandra A Siegel et al. 2021. "Can exposure to celebrities reduce prejudice? The effect of Mohamed Salah on Islamophobic behaviors and attitudes." *American Political Science Review* 115(4):1111–1128.

Masterson, Daniel and Vasil Yasenov. 2021. "Does halting refugee resettlement reduce crime? Evidence from the US refugee ban." *American Political Science Review* 115(3):1066–1073.

Schenoni, Luis L. 2021. "Bringing war back in: Victory and state formation in Latin America." *American Journal of Political Science* 65(2):405–421.

Singh, Shane P. 2019. "Compulsory voting and parties' vote-seeking strategies." *American Journal of Political Science* 63(1):37–52.

Spiess, Jann, Guido Imbens and Amar Venugopal. 2024. "Double and single descent in causal inference with an application to high-dimensional synthetic control." *Advances in Neural Information Processing Systems* 36.

Tellez, Juan Fernando. 2022. "Land, opportunism, and displacement in civil wars: evidence from Colombia." *American Political Science Review* 116(2):403–418.

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

Xu, Yiqing and Licheng Liu. 2021. *gsynth: Generalized Synthetic Control Method*. R package version 1.2.1.
URL: *https://CRAN.R-project.org/package=gsynth*

A Proofs

A.1 Proofs of Section 2

A.1.1 Proof of Theorem 1

Proof. From the specified model, we have

$$\begin{aligned}\widehat{\tau} &= Y_T - (\boldsymbol{X}_T^\top \widehat{\boldsymbol{\beta}} + Z_T \widehat{\gamma}) \\ \widehat{\tau}_{\text{res}} &= Y_T - \boldsymbol{X}_T^\top \widehat{\boldsymbol{\beta}}_{\text{res}}.\end{aligned}$$

Observe that

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\text{res}} - \widehat{\boldsymbol{\beta}} &= (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \boldsymbol{Y}_{T_0} - \widehat{\boldsymbol{\beta}} \\ &= (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top (\mathbf{X}_{T_0} \widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_{T_0} \widehat{\gamma} + \widehat{\boldsymbol{\epsilon}}) - \widehat{\boldsymbol{\beta}} \\ &= (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} \widehat{\gamma} + (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \widehat{\boldsymbol{\epsilon}}\end{aligned}$$

With some algebra,

$$\begin{aligned}\widehat{\tau}_{\text{res}} - \widehat{\tau} &= Y_T - \widehat{\boldsymbol{\beta}}_{\text{res}}^\top \boldsymbol{X}_T - (Y_T - \widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_T - Z_T \widehat{\gamma}) \\ &= \widehat{\gamma} Z_T - (\widehat{\boldsymbol{\beta}}_{\text{res}} - \widehat{\boldsymbol{\beta}})^\top X_T \\ &= \widehat{\gamma} Z_T - \{(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} \widehat{\gamma} + (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \widehat{\boldsymbol{\epsilon}}\}^\top X_T \\ &= \underbrace{\widehat{\gamma}}_{\text{weight}} \cdot \underbrace{(Z_T - \{(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0}\}^\top \boldsymbol{X}_T)}_{\text{imbalance}} - \underbrace{\{(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \widehat{\boldsymbol{\epsilon}}^\top\} \boldsymbol{X}_T}_{\text{noise term}}\end{aligned}$$

□

A.1.2 Proof of Corollary 1

Proof. Under Assumption 1, the model for control and missing units can be written as follows.

$$\begin{aligned}\boldsymbol{X}_T &= \mathbf{X}_{T_0}^\top \boldsymbol{\omega} + \boldsymbol{\epsilon}_{X,T}, \\ Z_T &= \boldsymbol{Z}_{T_0}^\top \boldsymbol{\omega} + \epsilon_{Z,T}\end{aligned}$$

We will use the following two notations throughout the proof:

$$\widehat{\boldsymbol{\eta}} = (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0}$$
$$\widehat{\boldsymbol{e}} = \boldsymbol{Z}_{T_0}^\top \mathbf{M}_{X_{T_0}}$$
$$\mathbf{M}_{X_{T_0}} = \mathbf{I} - \mathbf{X}_{T_0}(\boldsymbol{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top$$

We start with the imbalance term:

$$\begin{aligned}
\widehat{\delta} &= Z_T - \widehat{\boldsymbol{\eta}}^\top \boldsymbol{X}_T \\
&= \boldsymbol{Z}_{T_0}^\top \boldsymbol{\omega} + \epsilon_{Z,T} - \widehat{\boldsymbol{\eta}}^\top (\mathbf{X}_{T_0}^\top \boldsymbol{\omega} + \boldsymbol{\epsilon}_{X,T}) \\
&= \boldsymbol{Z}_{T_0}^\top \boldsymbol{\omega} - \widehat{\boldsymbol{\eta}}^\top \mathbf{X}_{T_0}^\top \boldsymbol{\omega} + \epsilon_{Z,T} - \widehat{\boldsymbol{\eta}}^\top \boldsymbol{\epsilon}_{X,T} \\
&= \boldsymbol{Z}_{T_0}^\top \{\mathbf{I} - \mathbf{X}_{T_0}(\boldsymbol{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top\} \boldsymbol{\omega} + \epsilon_{Z,T} - \widehat{\boldsymbol{\eta}}^\top \boldsymbol{\epsilon}_{X,T} \\
&= \widehat{\boldsymbol{e}} \boldsymbol{\omega} + \underbrace{\epsilon_{Z,T} - \widehat{\boldsymbol{\eta}}^\top \boldsymbol{\epsilon}_{X,T}}_{\text{noise term}}
\end{aligned}$$

Next, by Frisch–Waugh–Lovell theorem, the weight term can be written as

$$\begin{aligned}
\widehat{\gamma} &= (\boldsymbol{Z}_{T_0}^\top \mathbf{M}_{X_{T_0}} \boldsymbol{Z}_{T_0})^{-1} (\boldsymbol{Z}_{T_0}^\top \mathbf{M}_{X_{T_0}} \boldsymbol{Y}_{T_0}) \\
&= (\boldsymbol{Z}_{T_0}^\top \mathbf{M}_{X_{T_0}} \mathbf{M}_{X_{T_0}}^\top \boldsymbol{Z}_{T_0})^{-1} (\boldsymbol{Z}_{T_0}^\top \mathbf{M}_{X_{T_0}} \boldsymbol{Y}_{T_0}) \\
&= (\widehat{\boldsymbol{e}} \widehat{\boldsymbol{e}}^\top)^{-1} (\widehat{\boldsymbol{e}} \boldsymbol{Y}_{T_0})
\end{aligned}$$

□

A.1.3 Proof of Proposition 1

Proof.

$$\begin{aligned}
\mathbb{E}[\widehat{\delta}\widehat{\gamma}] &= \mathbb{E}[(\widehat{\boldsymbol{e}}\boldsymbol{\omega} + \epsilon_{Z,T} - \widehat{\boldsymbol{\eta}}^\top \boldsymbol{\epsilon}_{X,T})(\widehat{\boldsymbol{e}}\widehat{\boldsymbol{e}}^\top)^{-1}(\widehat{\boldsymbol{e}}\boldsymbol{Y}_{T_0})] \\
&= \mathbb{E}[\widehat{\boldsymbol{e}}\boldsymbol{\omega}(\widehat{\boldsymbol{e}}\widehat{\boldsymbol{e}}^\top)^{-1}(\widehat{\boldsymbol{e}}\boldsymbol{Y}_{T_0})] + \mathbb{E}[(\epsilon_{Z,T} - \widehat{\boldsymbol{\eta}}^\top \boldsymbol{\epsilon}_{X,T})(\widehat{\boldsymbol{e}}\widehat{\boldsymbol{e}}^\top)^{-1}(\widehat{\boldsymbol{e}}\boldsymbol{Y}_{T_0})] \\
&= \widehat{\boldsymbol{e}}\boldsymbol{\omega}(\widehat{\boldsymbol{e}}\widehat{\boldsymbol{e}}^\top)^{-1}(\widehat{\boldsymbol{e}}\boldsymbol{Y}_{T_0})
\end{aligned}$$

where the last line holds by strict exogeneity assumption. By the Cauchy–Schwarz inequality,

$$
\begin{aligned}
|\widehat{\boldsymbol{e}}\boldsymbol{\omega}| &\leq \|\widehat{\boldsymbol{e}}\|\|\boldsymbol{\omega}\| \\
\frac{|\widehat{\boldsymbol{e}}\boldsymbol{Y}_{T_0}|}{\|\widehat{\boldsymbol{e}}\|^2} &\leq \frac{\|\widehat{\boldsymbol{e}}\|\|\boldsymbol{Y}_{T_0}\|}{\|\widehat{\boldsymbol{e}}\|^2} = \frac{\|\boldsymbol{Y}_{T_0}\|}{\|\widehat{\boldsymbol{e}}\|}.
\end{aligned}
$$

Combining these two inequalities:

$$
\widehat{\boldsymbol{e}}\boldsymbol{\omega}(\widehat{\boldsymbol{e}}\widehat{\boldsymbol{e}}^\top)^{-1}(\widehat{\boldsymbol{e}}\boldsymbol{Y}_{T_0}) \leq |\widehat{\boldsymbol{e}}\boldsymbol{\omega}(\widehat{\boldsymbol{e}}\widehat{\boldsymbol{e}}^\top)^{-1}(\widehat{\boldsymbol{e}}\boldsymbol{Y}_{T_0})| \leq \|\boldsymbol{\omega}\|\|\boldsymbol{Y}_{T_0}\|
$$

where the equality holds if $\{\widehat{\boldsymbol{e}}, \boldsymbol{\omega}\}$ is linearly dependent, and so does $\{\widehat{\boldsymbol{e}}, \boldsymbol{Y}_{T_0}\}$. □

A.1.4 Proof of Corollary 2

Proof. The proof for the weight term is same as Corollary 1. Note that we have

$$
\begin{aligned}
\boldsymbol{\phi}_X &= \frac{1}{T-1}\boldsymbol{\mu}^\top(\mathbf{X}_{T_0} - \boldsymbol{\epsilon}_X) \\
\boldsymbol{\phi}_Z &= \frac{1}{T-1}\boldsymbol{\mu}^\top(\boldsymbol{Z}_{T_0} - \boldsymbol{\epsilon}_Z)
\end{aligned}
$$

For the imbalance term,

$$
\begin{aligned}
\widehat{\delta} &= Z_T - \widehat{\boldsymbol{\eta}}^\top \boldsymbol{X}_T \\
&= (\boldsymbol{\phi}_Z^\top \boldsymbol{\mu}_T + \epsilon_{Z,T}) - \widehat{\boldsymbol{\eta}}^\top(\boldsymbol{\phi}_X^\top \boldsymbol{\mu}_T + \boldsymbol{\epsilon}_{X,T}) \\
&= (\boldsymbol{\phi}_Z - \boldsymbol{\phi}_X \widehat{\boldsymbol{\eta}})^\top \boldsymbol{\mu}_T + (\epsilon_{Z,T} - \widehat{\boldsymbol{\eta}}^\top \boldsymbol{\epsilon}_{X,T})
\end{aligned}
$$

Using the results above

$$
= \frac{1}{T-1}\left(\boldsymbol{Z}_{T_0} - \boldsymbol{\epsilon}_Z - \mathbf{X}_{T_0}\widehat{\boldsymbol{\eta}} + \boldsymbol{\epsilon}_X \widehat{\boldsymbol{\eta}}\right)^\top \boldsymbol{\mu}\boldsymbol{\mu}_T + (\epsilon_{Z,T} - \widehat{\boldsymbol{\eta}}^\top \boldsymbol{\epsilon}_{X,T})
$$

Plugging in $\widehat{\boldsymbol{\eta}} = \{(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}\mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0}\}$

$$
\begin{aligned}
&= \frac{1}{T-1}\boldsymbol{Z}_{T_0}^\top\{\mathbf{I} - \mathbf{X}_{T_0}(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}\mathbf{X}_{T_0}^\top\}\boldsymbol{\mu}\boldsymbol{\mu}_T - \frac{1}{T-1}(\boldsymbol{\epsilon}_Z - \boldsymbol{\epsilon}_X \widehat{\boldsymbol{\eta}})^\top \boldsymbol{\mu}\boldsymbol{\mu}_T + (\epsilon_{Z,T} - \widehat{\boldsymbol{\eta}}^\top \boldsymbol{\epsilon}_{X,T}) \\
&= \frac{1}{T-1}\widehat{\boldsymbol{e}}\boldsymbol{\mu}\boldsymbol{\mu}_T - \frac{1}{T-1}(\boldsymbol{\epsilon}_Z - \boldsymbol{\epsilon}_X \widehat{\boldsymbol{\eta}})^\top \boldsymbol{\mu}\boldsymbol{\mu}_T + (\epsilon_{Z,T} - \widehat{\boldsymbol{\eta}}^\top \boldsymbol{\epsilon}_{X,T})
\end{aligned}
$$

□

A.1.5 Proof of Proposition 2

Proof. Proof is analogous to Proposition 1. □

A.2 Proofs of Section 3

A.2.1 Lemmas

Lemma 1 (Sherman-Morrison Formula)**.** *If A and $A+B$ are invertible and B is rank 1, then*

$$(A+B)^{-1} = A^{-1} - \frac{1}{1+g}A^{-1}BA^{-1}$$

where $g = \mathit{Trace}(BA^{-1})$ and $g \neq -1$.

Lemma 2. *We let*

$$\mathbf{X}_{(T\times N)} = \begin{bmatrix} \mathbf{X}_{T_0} \\ \boldsymbol{X}_T^\top \end{bmatrix}, \quad \mathbf{X}_{T_0} = \begin{bmatrix} \boldsymbol{X}_1^\top \\ \vdots \\ \boldsymbol{X}_{T-1}^\top \end{bmatrix} \tag{11}$$

where $\mathbf{X}_{T_0}$ is a $(T-1)\times N$ matrix, and $\boldsymbol{X}_t$ is a N-length column vector. Then,

$$(\mathbf{X}^\top\mathbf{X})^{-1} = (\mathbf{X}_{T_0}^\top\mathbf{X}_{T_0} + \boldsymbol{X}_T\boldsymbol{X}_T^\top)^{-1} \tag{12}$$

$$= (\mathbf{X}_{T_0}^\top\mathbf{X}_{T_0})^{-1} - \mathbf{P} \tag{13}$$

where $\mathbf{P}$ is given by

$$\mathbf{P} = \frac{1}{1+\boldsymbol{X}_T^\top(\mathbf{X}_{T_0}^\top\mathbf{X}_{T_0})^{-1}\boldsymbol{X}_T}(\mathbf{X}_{T_0}^\top\mathbf{X}_{T_0})^{-1}\boldsymbol{X}_T\boldsymbol{X}_T^\top(\mathbf{X}_{T_0}^\top\mathbf{X}_{T_0})^{-1}. \tag{14}$$

Proof of Lemma 2. From the definition of $\mathbf{X}$ we have

$$\mathbf{X}^\top\mathbf{X} = \begin{bmatrix} \mathbf{X}_{T_0}^\top & \boldsymbol{X}_T \end{bmatrix} \begin{bmatrix} \mathbf{X}_{T_0} \\ \boldsymbol{X}_T^\top \end{bmatrix}$$

$$= \mathbf{X}_{T_0}^\top\mathbf{X}_{T_0} + \boldsymbol{X}_T\boldsymbol{X}_T^\top$$

Now, from Lemma 1, take $\mathbf{A} = \mathbf{X}_{T_0}^\top\mathbf{X}_{T_0}$ and $\mathbf{B} = \boldsymbol{X}_T\boldsymbol{X}_T^\top$ and realize that $\mathbf{B}$ is rank 1. Then, we apply the result of the lemma and obtain

$$(\mathbf{X}^\top \mathbf{X})^{-1} = (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0} + \boldsymbol{X}_T \boldsymbol{X}_T^\top)^{-1} \tag{15}$$

$$= (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} - \mathbf{P}. \tag{16}$$

The explicit form of $\mathbf{P}$ is given by

$$\begin{aligned}\mathbf{P} &= \frac{1}{1 + \mathrm{Tr}(\boldsymbol{X}_T \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1})} (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \\ &= \frac{1}{1 + \mathrm{Tr}(\boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T)} (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \\ &= \frac{1}{1 + \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T} (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}.\end{aligned}$$

□

Lemma 3.

$$\begin{aligned}\boldsymbol{X}_T^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{X}_T &= \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T - \boldsymbol{X}_T^\top \mathbf{P} \boldsymbol{X}_T \\ &= \frac{Q}{1+Q}\end{aligned}$$

where Q is given by

$$Q = \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T.$$

Proof. From Lemma 2, $\mathbf{P}$ is given by

$$\mathbf{P} = \frac{1}{1 + \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T} (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}.$$

Accordingly,

$$\begin{aligned}\boldsymbol{X}_T^\top \mathbf{P} \boldsymbol{X}_T &= \frac{1}{1 + \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T} \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T \\ &= \frac{Q^2}{1+Q}.\end{aligned}$$

By Lemma 2, we can show that

$$\begin{aligned}\boldsymbol{X}_T^\top(\mathbf{X}^\top\mathbf{X})^{-1}\boldsymbol{X}_T &= \boldsymbol{X}_T^\top(\mathbf{X}_{T_0}^\top\mathbf{X}_{T_0})^{-1}\boldsymbol{X}_T - \boldsymbol{X}_T^\top\mathbf{P}\boldsymbol{X}_T \\ &= Q - \frac{Q^2}{1+Q} \\ &= \frac{Q}{1+Q}\end{aligned}$$

□

Lemma 4 (Matrix Inversion in Block Form)**.** *Let a $(m+1)\times(m+1)$ matrix $\mathbf{M}$ be partitioned into a block form:*

$$\mathbf{M} = \begin{bmatrix}\mathbf{A} & \boldsymbol{b} \\ \boldsymbol{b}^\top & c\end{bmatrix}$$

where $\mathbf{A}$ is a $m\times m$ matrix, $\boldsymbol{b}$ is a m-length column vector, and c is a scalar. If $\mathbf{A}$ is invertible, then the inverse of $\mathbf{M}$ is given by

$$\mathbf{M}^{-1} = \begin{bmatrix}(\mathbf{A} - \frac{1}{c}\boldsymbol{b}\boldsymbol{b}^\top)^{-1} & -\frac{1}{k}\mathbf{A}^{-1}b \\ -\frac{1}{k}b^\top\mathbf{A}^{-1} & \frac{1}{k}\end{bmatrix}$$

where $k = c - \boldsymbol{b}^\top\mathbf{A}^{-1}\boldsymbol{b}$.

Lemma 5 (Coefficients on Restricted Vertical Regression)**.** *Let $Q = \boldsymbol{X}_T^\top(\mathbf{X}_{T_0}^\top\mathbf{X}_{T_0})^{-1}\boldsymbol{X}_T$, then the coefficients on the restricted vertical regression with augmented $\boldsymbol{D}$ in Equation (9) is given by*

$$\begin{pmatrix}\widehat{\boldsymbol{\beta}}^*_{res} \\ \widehat{\tau}^*_{res}\end{pmatrix} = \begin{bmatrix}(\mathbf{X}_{T_0}^\top\mathbf{X}_{T_0})^{-1}(\mathbf{X}_{T_0}^\top\mathbf{Y}_{T_0} + \boldsymbol{X}_T Y_T) - (1+Q)(\mathbf{X}^\top\mathbf{X})^{-1}\boldsymbol{X}_T Y_T \\ -(1+Q)\boldsymbol{X}_T^\top(\mathbf{X}^\top\mathbf{X})^{-1}(\mathbf{X}_{T_0}^\top\mathbf{Y}_{T_0} + \boldsymbol{X}_T Y_T) + (1+Q)Y_T\end{bmatrix}.$$

Proof of Lemma 5. Let $\widetilde{\mathbf{X}}$ be the $T\times(N+1)$ design matrix of the augmented vertical regression:

$$\begin{aligned}\widetilde{\mathbf{X}} &= \begin{bmatrix}\mathbf{X} & \boldsymbol{D}\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{X}_{T_0} & \mathbf{0} \\ \boldsymbol{X}_T^\top & 1\end{bmatrix}.\end{aligned}$$

Then, with some algebra we have

$$\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_{T_0}^\top & \boldsymbol{X}_T \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_{T_0} & \mathbf{0} \\ \boldsymbol{X}_T^\top & 1 \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{X}_{T_0}^\top \mathbf{X}_{T_0} + \boldsymbol{X}_T \boldsymbol{X}_T^\top & \boldsymbol{X}_T \\ \boldsymbol{X}_T^\top & 1 \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{X}^\top \mathbf{X} & \boldsymbol{X}_T \\ \boldsymbol{X}_T^\top & 1 \end{bmatrix}$$

and

$$\widetilde{\mathbf{X}}^\top \boldsymbol{Y} = \begin{bmatrix} \mathbf{X}_{T_0}^\top & \boldsymbol{X}_T \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{Y}_{T_0} \\ Y_T \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{X}_{T_0}^\top \boldsymbol{Y}_{T_0} + \boldsymbol{X}_T Y_T \\ Y_T \end{bmatrix}.$$

To derive the inverse of $\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}}$, we apply Lemma 4 where we set $\mathbf{M} = \widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}}$, $\mathbf{A} = \mathbf{X}^\top \mathbf{X}$, $\boldsymbol{b} = X_T$, and $c = 1$. Then, we have

$$\begin{aligned} k &= c - \boldsymbol{b}^\top \mathbf{A}^{-1} \boldsymbol{b} \\ &= 1 - \boldsymbol{X}_T^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{X}_T \\ &= 1 - \frac{Q}{1+Q} \\ &= \frac{1}{1+Q}. \end{aligned}$$

where $Q = \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T$ by Lemma 3. Following Lemma 4, we have

$$(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1} = \begin{bmatrix} (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} & -(1+Q)(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{X}_T \\ -(1+Q) \boldsymbol{X}_T^\top (\mathbf{X}^\top \mathbf{X})^{-1} & 1+Q \end{bmatrix}.$$

Thus, the ordinary least squares estimator of the restricted regression coefficients is given by

$$\begin{aligned}
\begin{pmatrix} \widehat{\boldsymbol{\beta}}^*_{\text{res}} \\ \widehat{\tau}^*_{\text{res}} \end{pmatrix} &= (\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^\top \boldsymbol{Y} \\
&= \begin{bmatrix} (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} & -(1+Q)(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{X}_T \\ -(1+Q)\boldsymbol{X}_T^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} & 1+Q \end{bmatrix} \begin{bmatrix} \mathbf{X}_{T_0}^\top \boldsymbol{Y}_{T_0} + \boldsymbol{X}_T Y_T \\ Y_T \end{bmatrix} \\
&= \begin{bmatrix} (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} (\mathbf{X}_{T_0}^\top \boldsymbol{Y}_{T_0} + \boldsymbol{X}_T Y_T) - (1+Q)(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{X}_T Y_T \\ -(1+Q)\boldsymbol{X}_T^\top (\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}_{T_0}^\top \boldsymbol{Y}_{T_0} + \boldsymbol{X}_T Y_T) + (1+Q) Y_T \end{bmatrix}
\end{aligned}$$

where the final line provides the desired result. □

A.2.2 Proof of Proposition 3 and Corollary 3

Proof. We first provide proof for Corollary 3. Proof for Proposition 3 is analogous.

In the first part of the proof, we show that $\widehat{\boldsymbol{\beta}}^*_{\text{res}} = \widehat{\boldsymbol{\beta}}_{\text{res}}$, and in the second part we show that $\widehat{\tau}^*_{\text{res}} = \widehat{\tau}_{\text{res}}$.

Part 1. Regression Coefficients From Lemma 5, we have

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^*_{\text{res}} &= (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} (\mathbf{X}_{T_0}^\top \boldsymbol{Y}_{T_0} + \boldsymbol{X}_T Y_T) - (1+Q)(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{X}_T Y_T \\
&= (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \boldsymbol{Y}_{T_0} + \{(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T - (1+Q)(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{X}_T\} Y_T
\end{aligned}$$

To prove that $\widehat{\boldsymbol{\beta}}^*_{\text{res}} = (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \boldsymbol{Y}_{T_0} = \widehat{\boldsymbol{\beta}}_{\text{res}}$, it is sufficient to show that

$$(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T - (1+Q)(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{X}_T = 0.$$

By Lemma 2, we have

$$\begin{aligned}
&(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T - (1+Q)(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{X}_T \\
&= (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T - (1+Q)\{(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} - \mathbf{P}\} \boldsymbol{X}_T \\
&= -(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T Q + (1+Q)\mathbf{P} \boldsymbol{X}_T \\
&= -(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T + (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T \\
&= 0
\end{aligned}$$

where the second to the last equality follows by plugging $\mathbf{P} = (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}/(1+Q)$ and $Q = \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T$ in the expression. Then, it follows that $\widehat{\boldsymbol{\beta}}_{\text{res}}^* = \widehat{\boldsymbol{\beta}}_{\text{res}}$.

Part 2. Treatment Effect From Lemma 2 and 5, we have

$$\begin{aligned}
\widehat{\tau}_{\text{res}}^* &= -(1+Q)\boldsymbol{X}_T^\top (\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}_{T_0}^\top \boldsymbol{Y}_{T_0} + \boldsymbol{X}_T Y_T) + (1+Q) Y_T \\
&= Y_T - \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \boldsymbol{Y}_{T_0} + \boldsymbol{X}_T^\top \mathbf{P} (\mathbf{X}_{T_0}^\top \boldsymbol{Y}_{T_0} + \boldsymbol{X}_T Y_T) \\
&\quad - Q \boldsymbol{X}_T^\top \{(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} - \mathbf{P}\} (\mathbf{X}_{T_0}^\top \boldsymbol{Y}_{T_0} + \boldsymbol{X}_T Y_T) \\
&= Y_T - \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \boldsymbol{Y}_{T_0} \\
&\quad + \{(1+Q)\boldsymbol{X}_T^\top \mathbf{P} - Q \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}\} (\mathbf{X}_{T_0}^\top \boldsymbol{Y}_{T_0} + \boldsymbol{X}_T Y_T)
\end{aligned}$$

By plugging $P = (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}/(1+Q)$ and $Q = \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T$ in the expression, we have

$$\begin{aligned}
&(1+Q)\boldsymbol{X}_T^\top \mathbf{P} - Q \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \\
&\quad = \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} - \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \boldsymbol{X}_T \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \\
&\quad = 0.
\end{aligned}$$

Thus, we conclude that

$$\begin{aligned}
\widehat{\tau}_{\text{res}}^* &= Y_T - \boldsymbol{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \boldsymbol{Y}_{T_0} \\
&= Y_T - \boldsymbol{X}_T^\top \widehat{\boldsymbol{\beta}}_{\text{res}} \\
&= \widehat{\tau}_{\text{res}}
\end{aligned}$$

which completes the proof. □

A.2.3 Proof of Theorem 2

Proof. Recall that, by Proposition 3 and Corollary 3, we have equivalence results of $\widehat{\tau}^* = \widehat{\tau}$, $\widehat{\tau}_{\text{res}}^* = \widehat{\tau}_{\text{res}}$, and $\widehat{\gamma}^* = \widehat{\gamma}$. Thus, it is sufficient to show that $\widehat{\delta}^* = \widehat{\delta}$.

Let $\mathbf{M}_{\mathbf{X}} = \mathbf{I} - \mathbf{X}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}$. In matrix format, we want to show:

$$\begin{aligned}(\boldsymbol{D}^\top \mathbf{M}_X \boldsymbol{D})^{-1}\mathbf{D}^\top \mathbf{M}_X \boldsymbol{Z} &= Z_T - \boldsymbol{X}_T^\top(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}\mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} \\ &= Z_T - \widehat{\boldsymbol{\eta}}^\top \boldsymbol{Z}_{T_0} \\ &= \widehat{\delta}.\end{aligned}$$

Focusing on the first term of left hand side, by Lemma 3, we have

$$\begin{aligned}\boldsymbol{D}^\top \mathbf{M}_{\mathbf{X}} \boldsymbol{D} &= 1 - \mathbf{X}_T^\top(\mathbf{X}^\top\mathbf{X})^{-1}\boldsymbol{X}_T \\ &= 1 - \frac{Q}{1+Q} \\ &= \frac{1}{1+Q}\end{aligned}$$

where $Q = \boldsymbol{X}_T^\top(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}\boldsymbol{X}_T$.

The second term on the left hand side is given by

$$\begin{aligned}\boldsymbol{D}^\top \mathbf{M}_{\mathbf{X}} \boldsymbol{Z} &= Z_T - \boldsymbol{X}_T^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \boldsymbol{Z} \\ &= Z_T - \boldsymbol{X}_T^\top(\mathbf{X}^\top\mathbf{X})^{-1}(\mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} + \boldsymbol{X}_T Z_T) \\ &= Z_T - \boldsymbol{X}_T^\top\{(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} - \mathbf{P}\}\mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} - \frac{Q}{1+Q} Z_T \\ &= \widehat{\delta} + \boldsymbol{X}_T^\top \mathbf{P} \mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} - \frac{Q}{1+Q} Z_T.\end{aligned}$$

where we use Lemma 2 in the second to the last equality.

Combining the two, we have

$$\left(\frac{1}{1+Q}\right)^{-1}\left(\widehat{\delta} + \boldsymbol{X}_T^\top \mathbf{P} \mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} - \frac{Q}{1+Q} Z_T\right) = (1+Q)\widehat{\delta} + (1+Q)\boldsymbol{X}_T^\top \mathbf{P} \mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} - Q Z_T$$

It remains to show that

$$(1+Q)\widehat{\delta} + (1+Q)\boldsymbol{X}_T^\top \mathbf{P} \mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} - Q Z_T = \widehat{\delta}.$$

Now,

$$
\begin{aligned}
&Q\widehat{\delta} + (1+Q)\mathbf{X}_T^\top \mathbf{P}\mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} - QZ_T \\
&= (1+Q)\mathbf{X}_T^\top \mathbf{P}\mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} - Q\mathbf{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}\mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} \\
&= (1+Q)\mathbf{X}_T^\top \frac{1}{1+\mathbf{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}\mathbf{X}_T} (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}\mathbf{X}_T\mathbf{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}\mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} \\
&\qquad - Q\mathbf{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}\mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} \\
&= (1+Q)\frac{Q}{1+Q}\mathbf{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}\mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} - Q\mathbf{X}_T^\top (\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1}\mathbf{X}_{T_0}^\top \boldsymbol{Z}_{T_0} \\
&= 0
\end{aligned}
$$

which completes the proof. □

B Simulation Study

In this section, we conduct simulation studies to evaluate bias from vertical regression with missing outcomes using simulated data generated with the linear factor model and West German reunification data.

B.1 Setup

To begin with, we use the generalized SCM (Xu, 2017) to estimate an linear factor model with two latent factors based on GDP per capita data from the West German reunification example. Using the estimated factors, we augmented time factors to expand the pre-treatment period data. Specifically, we fit a spline regression with 5 degrees of freedom for each factor and randomly generated fitted values preceding and following the observed time periods. See Figure 8 for more details. We selected the estimated factor loadings of three control units (Austria, USA, and Japan), which were reported in the original study as having the three largest weights, and those of one additional control unit (the United Kingdom). Finally, using these augmented factors and a subset of the factor loadings, we generated synthetic outcome data for four control units and a single treated unit, covering 99 pre-treatment periods and a single post-treatment period for each simulation.

For each simulated dataset, we analyze a hypothetical scenario where half of the outcome data for a single control unit is missing. Specifically, we estimate the causal effect using the vertical regression estimator in Equation (4), excluding one of the four control units from the analysis. We also consider possible remedies, such as mean/median imputation and bias correction using partially observed data. In the latter case, we estimate the bias by applying the proposed bias decomposition to the observed half of the outcome data for the excluded unit.

B.2 Results

With a total of $1,000$ simulated datasets, we compare the root mean squared error of causal estimates based on the vertical regression estimator, using the following specifications:

1. Complete Observations: Using all available outcomes.

2. Complete `gsynth`: Interacted fixed effects model estimation by Xu (2017) using all available outcomes. We use the `R` package `gsynth` by Xu and Liu (2021) for the implementation.

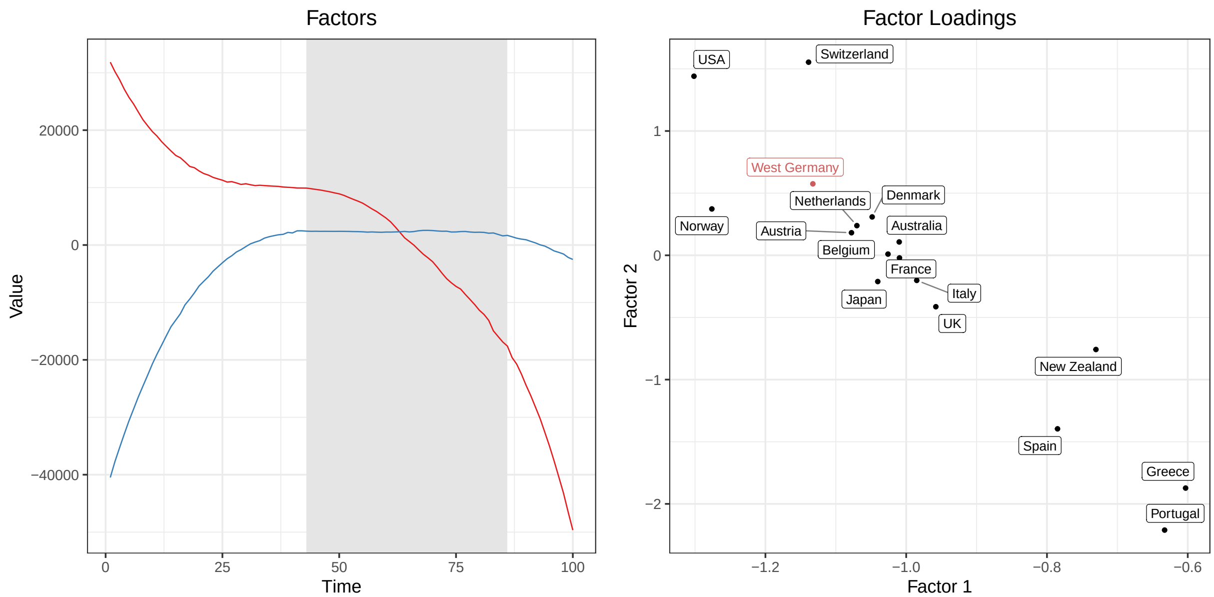

Figure 8: Factor and Factor Loadings for Simulation Study. Left panel visualizes factors used in generating synthetic data. The red line represents the value of the first factor, and the blue line represents the second factor. Note that for $t \in \{43, \ldots, 87\}$ (grey area), we use the factors from the linear factor model fitted with West German reunification data. We then augmented the data using a spline regression for $t \in \{1, \ldots, 42\}$ and $\{88, \ldots, 100\}$. Right panel visualizes factor loadings of the fitted model using West German reunification data. A subset of these factor loadings is used to generate synthetic data.

Cross-validation is used to select the number of factors within the interval of $[0, 2]$.

3. Unit-wise Deletion: Excluding one control unit that has missing values.

4. Mean Imputation: Imputing missing values using the sample mean of the partially observed outcomes for the excluded unit.

5. Median Imputation: Imputing missing values using the sample median of the partially observed outcomes for the excluded unit.

6. Bias Correction: Estimating bias by applying the proposed bias decomposition to the observed half of the outcome data.

Table 2 shows the bias of these different approaches using the root mean squared error (RMSE) from Monte Carlo simulations. Note that the vertical regression estimate with complete observations yields a bias of 1.24, which approaches zero as the number of pre-treatment periods increases (Abadie, Diamond and Hainmueller, 2010). As a comparison, the estimate of `gsynth` using complete observations yields a bias of 3.438. Most importantly, we can see that unit-wise deletion

Estimator	RMSE
Complete Observations	1.240
Complete `gsynth`	3.438
Unit-wise Deletion	1.528
Mean Imputation	1.530
Median Imputation	1.530
Bias Correction	1.254

Table 2: Root Mean Squared Error (RMSE) of Monte Carlo Simulation Studies: The true treatment effect is zero. The results are averaged over four missing scenarios where missing values occur in one of the four control units.

results in a greater bias of 1.528 compared to complete observations. Bias correction using the proposed bias decomposition improves the RMSE to 1.254, unlike the results observed with mean and median imputation. Note that this result reflects an average over four missing scenarios, where missing values occur in one of the four control units. Table 3 shows the RMSE of each hypothetical scenario.

Estimator	Austria	USA	Japan	UK
Unit-wise Deletion	1.263	1.242	2.149	1.257
Mean Imputation	1.266	1.247	2.147	1.262
Median Imputation	1.266	1.246	2.147	1.262
Bias Correction	1.243	1.240	1.290	1.243

Table 3: Root Mean Squared Error (RMSE) of Monte Carlo Simulation Studies for Each Scenario: The true treatment effect is zero. For example, dropping Austria from the model results in an RMSE of 1.263.

Using the same simulated dataset, we also analyze a hypothetical scenario where half of the outcome data for two control units are missing. Here, since there are multiple missing units, we follow the bias correction procedure outlined below, which is modified from the previous single-unit case. Let Z_1^{partial} and Z_2^{partial} be the control units with missing values, and $X_1, \ldots, X_{N-1}$ be the control units with fully observed data. The main idea is to create a new variable, $\widehat{Z}^{\text{partial}}$, that is a linear combination of Z_1^{partial} and Z_2^{partial}, where the weights come from the fitted coefficient of each unit. Then the weight of $\widehat{Z}^{\text{partial}}$ (denoted by $\widehat{\gamma}^{\text{partial}}$) becomes 1 and the imbalance (denoted by $\widehat{\delta}^{\text{partial}}$) can be estimated with a vertical regression. The detailed procedure is as follows:

1. Fit a vertical regression model $Y \sim Z_1^{\text{partial}} + Z_2^{\text{partial}} + X_1 + \ldots + X_{N-1}$. Let $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ denote the fitted coefficient of Z_1^{partial} and Z_2^{partial}, respectively.

2. Create a new variable: $\widehat{Z}^{\text{partial}} \equiv \widehat{\gamma}_1 Z_1^{\text{partial}} + \widehat{\gamma}_2 Z_2^{\text{partial}}$.

3. Fit a vertical regression model $\widehat{Z}^{\text{partial}} \sim X_1 + \ldots + X_{N-1}$.

4. Estimate the imbalance: $\widehat{\delta}^{\text{partial}} = (\widehat{Z}_T^{\text{partial}} - \{(\mathbf{X}_{T_0}^\top \mathbf{X}_{T_0})^{-1} \mathbf{X}_{T_0}^\top \widehat{\boldsymbol{Z}}_{T_0}^{\text{partial}}\}^\top \boldsymbol{X}_T)$.

5. Compute the estimated bias: $\widehat{\gamma}^{\text{partial}} \widehat{\delta}^{\text{partial}} = \widehat{\delta}^{\text{partial}}$.

Tables 4 and 5 show the results, which are similar to the single missing unit case except for a larger bias in imputation methods.

Estimator	RMSE
Complete Observations	1.240
Complete gsynth	3.438
Unit-wise Deletion	3.000
Mean Imputation	3.058
Median Imputation	3.025
Bias Correction	1.323

Table 4: Root Mean Squared Error (RMSE) of Monte Carlo Simulation Studies: The true treatment effect is zero. The results are averaged over six missing scenarios where missing values occur in two of the four control units.

Estimator	Aust., Japan	Aust., UK	Aust., USA	Japan, UK	USA, Japan	USA, UK
Unit-wise Deletion	1.297	6.111	1.311	2.620	1.334	2.148
Mean Imputation	1.310	6.251	1.319	2.657	1.345	2.168
Median Imputation	1.302	6.172	1.318	2.630	1.338	2.157

Table 5: Root Mean Squared Error (RMSE) of Monte Carlo Simulation Studies for Each Scenario: The true treatment effect is zero. For example, dropping Austria and Japan from the model results in an RMSE of 1.297.

C Additional Results from the West German Reunification Study

C.1 Leave-One-Out Analysis of SCM Estimate in 1990 – 2003

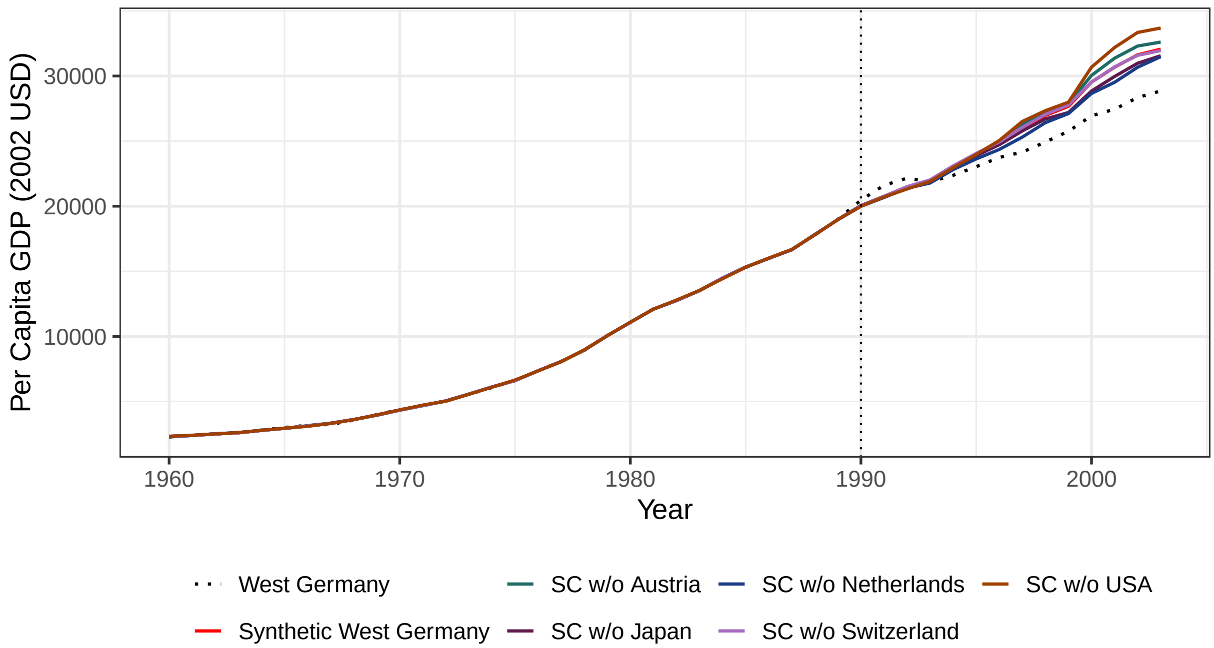

Figure 9: Leave-One-Out Analysis of the Impact of the 1990 German Reunification (Abadie, Diamond and Hainmueller, 2015) using Vertical Regression Estimator. Each solid line represents the synthetic control unit with one control unit omitted at a time. The black dotted line represents the observed time trend of West Germany, and the red solid line represents the synthetic control unit with the full data.

Year	Austria	Japan	Netherlands	Switzerland	USA	*None*
1990	429 (112)	418 (114)	472 (120)	444 (106)	469 (130)	424 (113)
1991	842 (153)	887 (151)	933 (160)	851 (150)	897 (177)	842 (154)
1992	684 (203)	761 (192)	777 (216)	651 (197)	832 (227)	673 (205)
1993	-112 (162)	-80 (162)	96 (136)	-146 (146)	-28 (185)	-109 (163)
1994	-675 (197)	-627 (198)	-444 (178)	-727 (162)	-586 (228)	-664 (198)
1995	-984 (258)	-849 (261)	-597 (226)	-1019 (211)	-932 (311)	-927 (267)
1996	-1218 (335)	-981 (348)	-618 (285)	-1195 (316)	-1306 (408)	-1102 (360)
1997	-2089 (442)	-1628 (509)	-1153 (444)	-1899 (505)	-2356 (567)	-1824 (532)
1998	-2248 (430)	-1770 (430)	-1488 (433)	-2121 (459)	-2408 (542)	-2043 (488)
1999	-2127 (636)	-1438 (544)	-1359 (674)	-2012 (645)	-2233 (778)	-1914 (681)
2000	-3101 (920)	-1904 (912)	-1717 (1077)	-2591 (1060)	-3738 (1141)	-2589 (1084)
2001	-3919 (1001)	-2515 (1114)	-2068 (1220)	-3251 (1237)	-4739 (1292)	-3224 (1265)
2002	-3956 (960)	-2633 (1092)	-2326 (1219)	-3241 (1191)	-4999 (1173)	-3279 (1220)
2003	-3752 (882)	-2703 (987)	-2629 (1123)	-3099 (1032)	-4829 (993)	-3207 (1072)

Table 6: Leave-One-Out Analysis of the Impact of the 1990 German Reunification (Abadie, Diamond and Hainmueller, 2015) using Vertical Regression Estimator. The columns represent the control unit that is omitted from the analysis, with the last column (*None*) showing the results with the full data. The rows represent the post-treatment period. The numbers represent the estimated treatment effect using the vertical regression approach, with the standard error provided in parentheses.

C.2 Sensitivity Contour Plot of SCM Estimate in 1991 – 2002

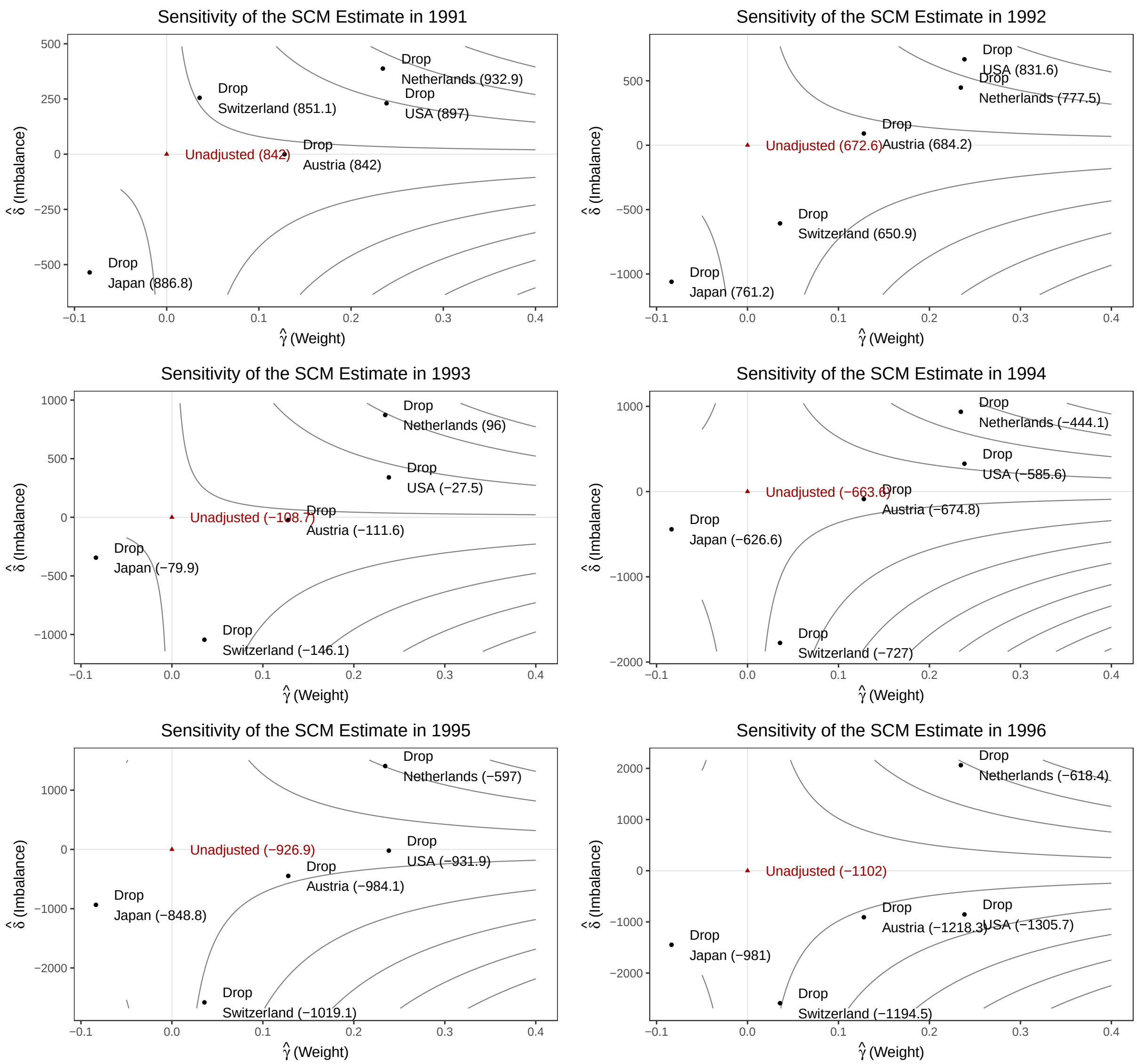

Figure 10: Sensitivity Contour Plot of the West German Reunification Example Using Vertical Regression (1991 – 1996). In this figure, the x-axis shows hypothetical value of weight sensitivity parameter ($\widehat{\gamma}$) and the x-axis shows that of imbalance ($\widehat{\delta}$). The contour line illustrates how the adjusted SCM estimate would be under those sensitivity parameters, i.e. $\widehat{\tau}^*_{\text{res}} - \widehat{\gamma}\widehat{\delta}$. The red triangle represents the unadjusted treatment estimate $\widehat{\tau}^*_{\text{res}}$, i.e. the estimate from a model including all the control units and assuming that $(\widehat{\gamma}, \widehat{\delta}) = (0, 0)$. Each circle presents the weight and imbalance terms under the case where we omit one of the observed control units from our analysis. The numbers inside the parentheses indicate the adjusted treatment estimate with those sensitivity parameters (the same quantity as shown in the contour line).

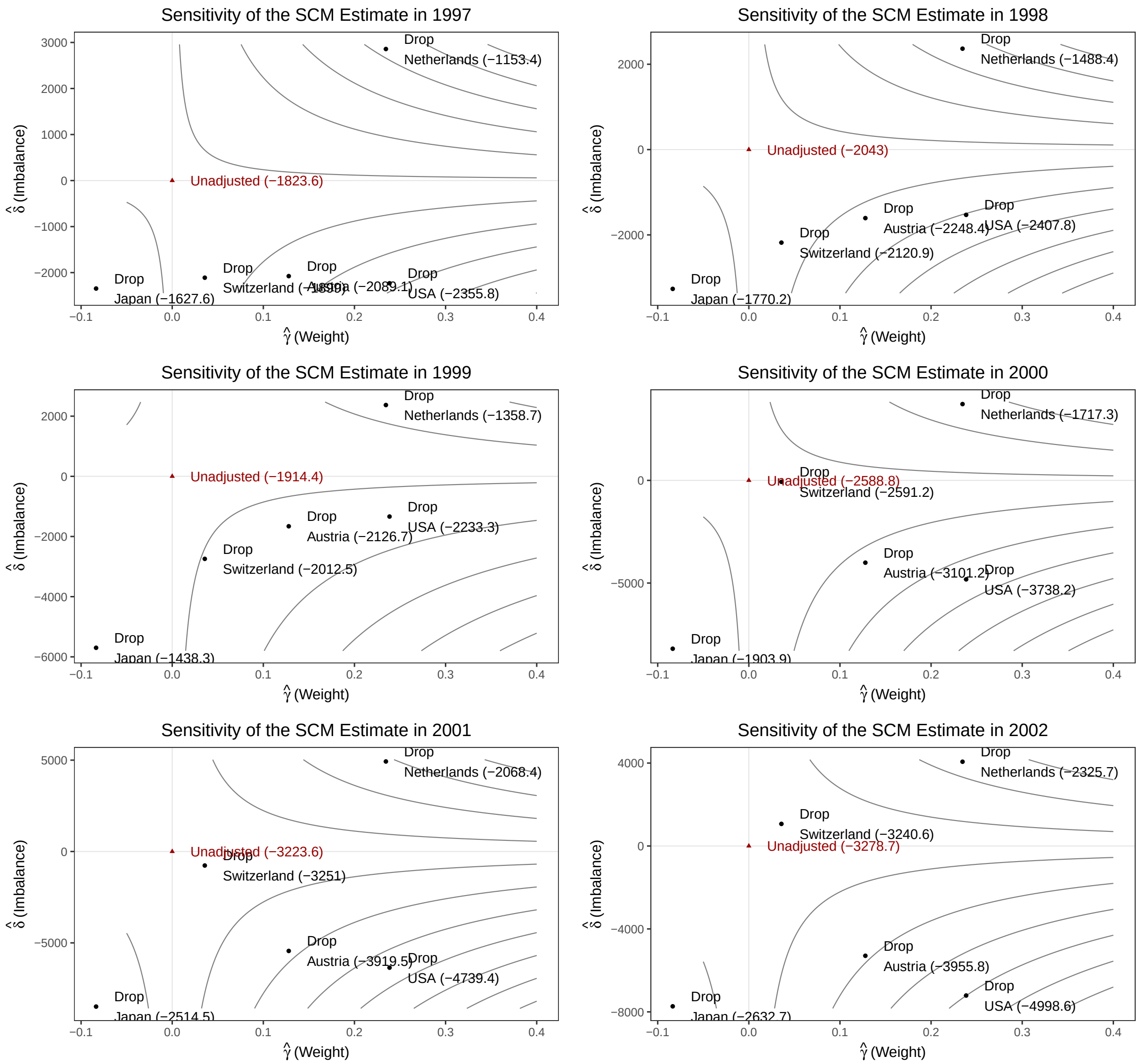

Figure 11: Sensitivity Contour Plot of the West German Reunification Example Using Vertical Regression (1997 – 2002). In this figure, the x-axis shows hypothetical value of weight sensitivity parameter ($\widehat{\gamma}$) and the x-axis shows that of imbalance ($\widehat{\delta}$). The contour line illustrates how the adjusted SCM estimate would be under those sensitivity parameters, i.e. $\widehat{\tau}^*_{\text{res}} - \widehat{\gamma}\widehat{\delta}$. The red triangle represents the unadjusted treatment estimate $\widehat{\tau}^*_{\text{res}}$, i.e. the estimate from a model including all the control units and assuming that $(\widehat{\gamma}, \widehat{\delta}) = (0, 0)$. Each circle presents the weight and imbalance terms under the case where we omit one of the observed control units from our analysis. The numbers inside the parentheses indicate the adjusted treatment estimate with those sensitivity parameters (the same quantity as shown in the contour line).

C.3 Sensitivity Contour Plot with Partial R^2

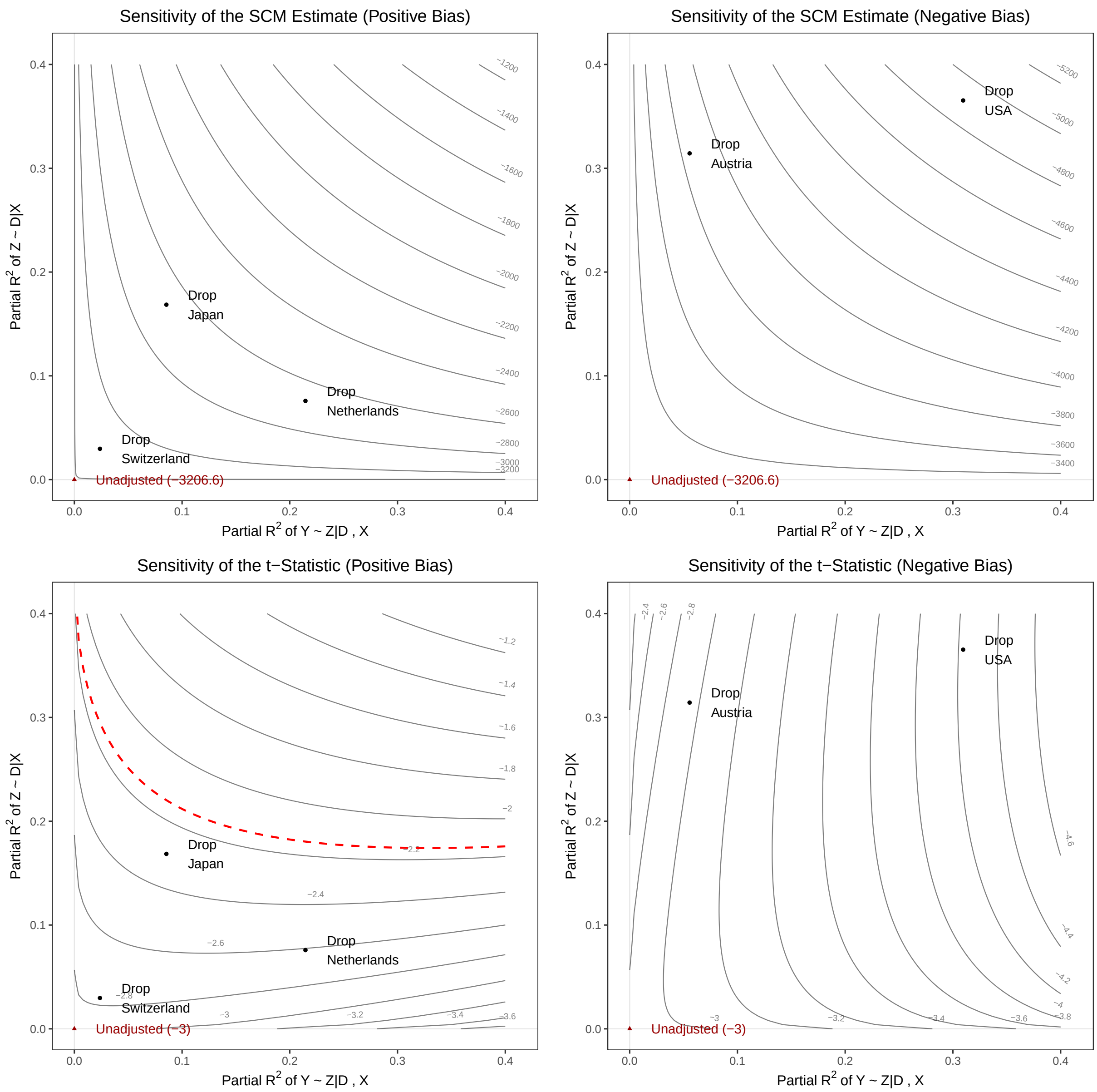

Figure 12: Sensitivity Contour Plot with Partial R^2 of the West German Reunification Example Using Vertical Regression (2003). The left two plots assume a positive bias, such as that resulting from omitting Japan, the Netherlands, or Switzerland. The right two plots assume a negative bias, such as that resulting from omitting Austria or the USA. The top two plots show the sensitivity of the SCM estimate, and the bottom two plots show the sensitivity of the t-statistic. The red dashed line in the bottom left plot shows the contour line for a critical value of the t-statistic at the 0.05 significance level.

D Additional Results from the Taiwan-IMF Study

D.1 Sensitivity Contour Plot of SCM Estimate in 1980 – 1989

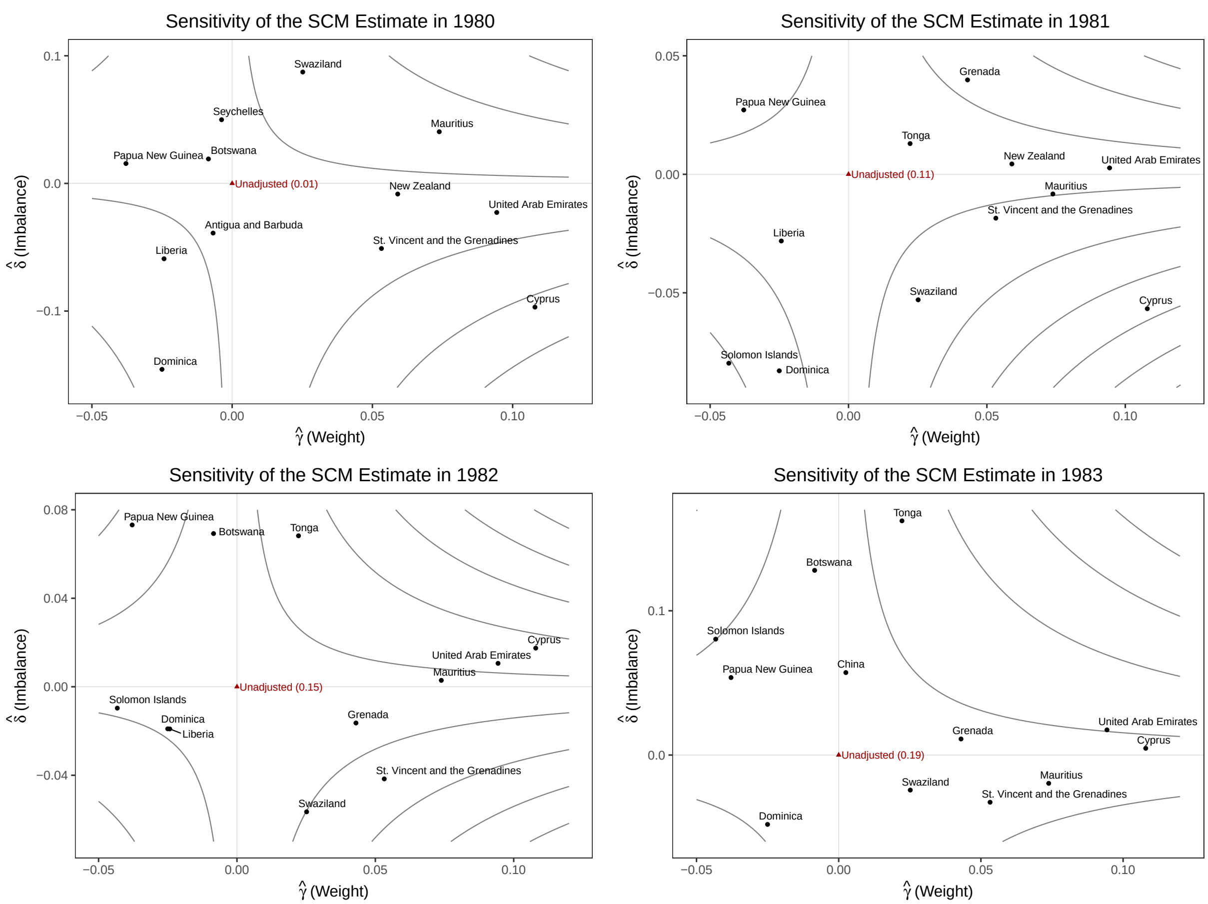

Figure 13: Sensitivity Contour Plot of the Taiwan-IMF Study Using Vertical Regression (1980 – 1983). In this figure, the x-axis shows hypothetical value of weight sensitivity parameter $(\widehat{\gamma})$ and the x-axis shows that of imbalance $(\widehat{\delta})$. The contour line illustrates how the adjusted SCM estimate would be under those sensitivity parameters, i.e. $\widehat{\tau}^*_{\text{res}} - \widehat{\gamma}\widehat{\delta}$. The red triangle represents the unadjusted treatment estimate $\widehat{\tau}^*_{\text{res}}$, i.e. the estimate from a model including all the control units and assuming that $(\widehat{\gamma}, \widehat{\delta}) = (0, 0)$. Each circle presents the estimated weight and imbalance terms for each country with partially observed pre-treatment data.

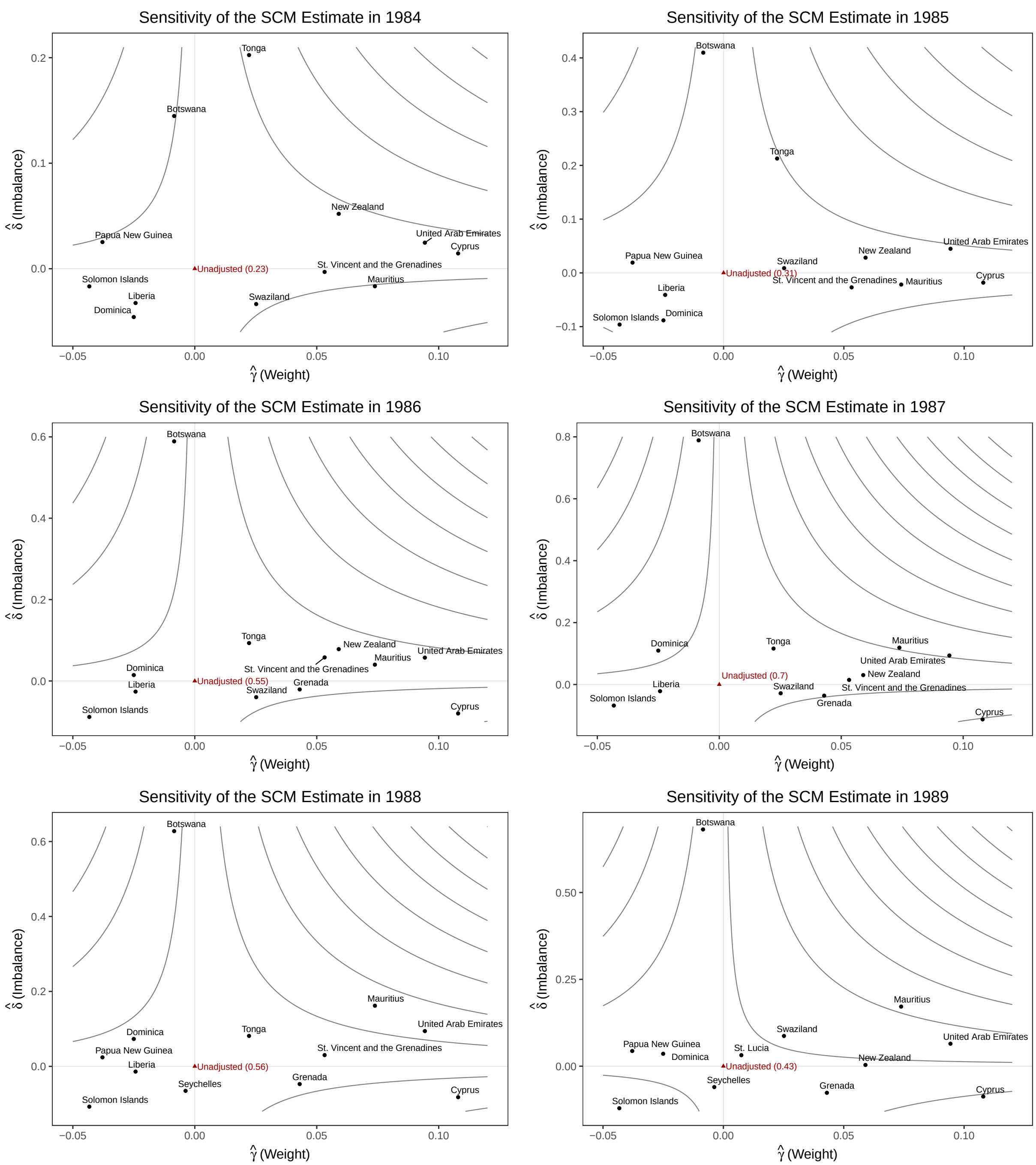

Figure 14: Sensitivity Contour Plot of the Taiwan-IMF Study Using Vertical Regression (1984 – 1989). In this figure, the x-axis shows hypothetical value of weight sensitivity parameter $(\widehat{\gamma})$ and the x-axis shows that of imbalance $(\widehat{\delta})$. The contour line illustrates how the adjusted SCM estimate would be under those sensitivity parameters, i.e. $\widehat{\tau}^*_{\text{res}} - \widehat{\gamma}\widehat{\delta}$. The red triangle represents the unadjusted treatment estimate $\widehat{\tau}^*_{\text{res}}$, i.e. the estimate from a model including all the control units and assuming that $(\widehat{\gamma}, \widehat{\delta}) = (0, 0)$. Each circle presents the estimated weight and imbalance terms for each country with partially observed pre-treatment data.