



UNIVERSITÀ
DI SIENA
1240

DIPARTIMENTO DI ECONOMIA POLITICA E STATISTICA

Corso di Laurea Magistrale in
ECONOMICS

Estimating measurement error in India's
GDP

A synthetic control approach

Relatore:

Prof. Federico Crudu

Correlatore:

Prof. Giuliano Curatola

Candidato:

Advait Moharir

Anno Accademico 2021-2022

Acknowledgements

I want to thank a few people.

Table of Contents

Summary	1
Chapter 1: GDP Measurement in India and the World	3
1.1 National Accounting: A Brief History	4
1.2 GDP Measurement: Methodologies	6
1.2.1 Production/ Output Method	7
1.2.2 Income Method	7
1.2.3 Expenditure Method	8
1.3 The Indian GDP Debate	9
1.3.1 The Nagaraj - CSO Debate	9
1.3.2 Subramanian's Cross-Country Approach	12
1.4 Conclusion	13
Chapter 2: The Synthetic Control Method	15
2.1 Introduction	15
2.2 Formal Aspects	17
2.2.1 Setup	17
2.2.2 Estimation	18
2.2.3 Inference	20
2.3 Advantages and Limitations of SCM	23
2.4 Alternative Estimators	24
2.5 Generalized Synthetic Control	26
2.5.1 Framework	27
2.5.2 Estimation and Inference	27
2.6 Conclusion	29
Chapter 3: Estimating measurement error in India's GDP using SCM	31
3.1 Literature Review	32

3.2	Econometric Approach	33
3.3	Data	35
3.4	Results	36
3.4.1	Inference	42
3.5	Robustness Checks	44
3.6	Discussion and Conclusion	47
References		49

List of Tables

2.1	p-values for the Post-Pre Intervention MSPE Ratio for select states	23
2.2	Differences between the SCM and GSCM Estimators	30
3.1	Summary Statistics	35
3.2	Difference between growth rates (percentage) of actual and synthetic GDP (All specifications)	41
3.3	Fisher's exact p - values	42
3.4	Average ATT and p - values for GSCM specifications	43

List of Figures

1.1	Disaggregated GDP growth rates by sector (2012-13)	11
1.2	Disaggregated GDP growth rates by sector (2013-14)	11
2.1	Synthetic Control of California	17
2.2	Per-capita cigarette sales gaps in California and placebo gaps in 19 control states	21
2.3	Ratio of post and pre-treatment MSPE in California and donor states	22
3.1	Evolution of Log GDP for India and Synthetic India (SCM)	36
3.2	Log GDP gap between India and Synthetic India (SCM)	37
3.3	Variable weights and country weights for synthetic India	38
3.4	Evolution of Log GDP for India and Synthetic India (GSCM)	39
3.5	Log GDP gap between India and Synthetic India (GSCM)	39
3.6	Evolution of Log GDP for India and Synthetic India with economic covariates (GSCM)	40
3.7	Evolution of Log GDP for India and Synthetic India with all covariates (GSCM)	41
3.8	Ratio of post and pre-treatment MSPE for India and donor countries	43
3.9	Evolution of log GDP for actual and synthetic India with backdated treatment (SCM)	45
3.10	Evolution of log GDP for actual and synthetic India with backdated treatment (GSCM)	46
3.11	Evolution of log GDP for actual and synthetic India with quarterly data (GSCM)	47

Summary

This thesis consists of three chapters. The first chapter titled “GDP Measurement in India and the World” provides a broad overview of the historical evolution of GDP measurement. I first describe the origins of GDP as a key macroeconomic statistic and a proxy of welfare, tracing its evolution from its conception by William Petty in the early 17th century, to the globally accepted and harmonized definition as set out by the United Nations. I then turn my focus to the three methods of output measurement, before delving into the debate on measurement issues in Indian GDP, focussing on critiques of real GDP measurement post 2011.

The second chapter, titled “The Synthetic Control Method” explores the synthetic control method (SCM). I begin by providing a brief introduction to the method, followed by a technical overview of the SCM. I then comment on the various advantages and disadvantages of the SCM, and review the alternative methods which have emerged in recent times to address the lacunae of the SCM. I then describe the generalized synthetic control method (GSCM), a flexible and widely used alternative to the SCM, and conclude by listing the salient differences between the two.

The final chapter, titled “Estimating measurement error in India’s GDP using SCM” uses the SCM and GSCM to construct counterfactual GDP series, using data from 19 donor countries. I begin by reviewing the literature on GDP measurement error, and synthetic control in macroeconomics. I then specify the identification strategy, describe the data sources, and present the results. Finally, I do a number of robustness checks, and end with a discussion of the results and avenues for future research.

Chapter 1

GDP Measurement in India and the World

Economic growth and its determinants are a key area of macroeconomic research. The question of achieving sustained and inclusive growth is the cornerstone of economic policy for most developing countries. Accurate statistical measurement of GDP and related macroeconomic aggregates are crucial to identify the sectors of the economy that are performing well and those that are stagnating. Consequently, the measurement of GDP and related magnitudes are an object of scrutiny from academia, media and policymakers.

However, India's GDP numbers have been scrutinized even more than usual, especially over the last decade. Commentators have identified several issues in the data post-2011, the major ones listed as follows. One, the assumption that the informal and formal sectors grow at the same rates has led to an overestimation bias in the output of the manufacturing sector. Recent research has found a significant difference in manufacturing output as computed by formal and informal sector surveys ([Bhattacharya, 2019](#)). Two, inappropriate deflators have been applied to nominal GDP data, and applying the right deflators leads to significant differences in output estimates. Finally, the focal point of debate has centred on the use of a new database, called the MCA21 database. Critics have pointed out that along with other issues like firm misclassification and lack of state-level data, the assumption that non-reporting firms contribute positively to growth has led to a “blowing up” of manufacturing output.

The third issue has caused many to question the veracity of GDP estimates in

recent years. This followed by the lack of good quality data alternatives raise an important question - what is the extent of measurement error in India's GDP? In this thesis, I attempt to tackle this question, specifically ascertaining whether there is an overestimation of real output. This chapter sets the context for the empirical exercise implemented in Chapter - 3, by first providing a historical overview of GDP and national accounting (Section 1.1). The rest of the chapter is structured as follows. Section 1.2, discusses the three methods of GDP measurement - namely the production, income and expenditure methods. In Section 1.3, I present the Indian GDP debate, focussing on two major critiques: one, the exchange between R. Nagaraj and the Central Statistical Organization on the veracity of the real GDP series with the base year 2011, and two, the concerns raised by Arvind Subramanian, the former Chief Economic Advisor on the recent growth numbers, and their changing relationship with correlates of GDP. Section 1.4 concludes.

1.1 National Accounting: A Brief History

The Gross Domestic Product (GDP) is one of many figures published in national account statistics across the world¹. The System of National Accounts (SNA), defines it as

the sum of the gross value added of all the institutional units resident in a territory engaged in production (that is, gross value added at basic prices) plus any taxes, minus any subsidies, on products not included in the value of their outputs.

This is an almost universally accepted definition of GDP. However, such precise definitions of output aggregates are a recent phenomenon, with the first edition of the SNA only appearing in the mid -20th century. The origins of national accounting can be traced back to British official William Petty, who first produced estimates of income, expenditure and population in the 17th century to show that England had sufficient resources to raise taxes and take on its neighbours in the Anglo-Dutch war. In this period, warfare was the key motivation for estimating national output, with estimates prioritizing different things for prosperity, ranging from increasing trade to reducing domestic government debt (Coyle, 2014). An innovation in establishing a stable benchmark for what constitutes national income came in form of Adam Smith's *Wealth of Nations*, which stressed the difference between productive and unproductive

¹Others include Gross National, Net Domestic and Net National Product respectively.

work ([Smith, 2000](#)). Smith stressed that only labour that adds tangible value is productive, and this net the national debt was the national income. In this definition, services were unproductive and subtracted from the total output.

Modern national income accounting was born as a consequence of the Great Depression and the Second World War. Colin Clark in the UK and Simon Kuznets in the US worked on providing income estimates for their respective countries. Kuznets's estimates of output during the Great Depression period were crucial to justifying the subsequent recovery plan ([Kuznets, 1934](#)). During this time, there was a clash of ideas about whether government spending should be included in national income accounting, as according to Kuznets, it increased growth independent of individual welfare. This was abandoned in favour of the inclusion of government spending as many wartime economists and government officials argued for its necessity in expanding output during wartime ([Keynes, 2010](#)). The Second World War was hence crucial in cementing the notion of government spending as adding to rather than subtracting from production. More importantly, national output explicitly became an indicator of prosperity and not welfare.

The first two decades of the post-war era, characterized as the Golden Age was characterized by rapid growth and low inflation and unemployment in the United States and Europe. National account statistics began to be standardized, and the United Nations established the SNA in 1953. The ideas of demand management and the importance of public spending in ensuring full employment outlined in Keynes' General Theory centralized the notion that government can control output. This period also saw the Cold War emerge between the US and the Soviet Union, and led to the development of exchange rate and purchasing power parity (PPP) adjusted GDP to make international comparisons. The notion of GDP as an end in itself was challenged by the challenges of the 1970s, like stagflation and the critiques offered by movements like environmentalism. This led to the development of alternative measures of development like the Human Development Index (HDI).

The next two decades were marked by a shift in the understanding of the causes of growth: fiscal policy receded into the background, while monetary policy and price stability took centre - stage. Growth theory endogenized technology, and human capital, arguing that innovation and R&D create a virtuous cycle([Romer, 1986, 1990](#)). The key challenge during this period was the measurement of services: with the GDP largely originating from the need to estimate the material constraints

of the economy, measuring the value of services, especially public services proved to be difficult as inputs are difficult to quantify. With explosion and innovation and variety, it became difficult for GDP to account for rapid changes in prices and improvements in quality, leading the Boskin Commission in the US to conclude that inflation was overstated and growth understated by an average of 1.3% points (Oulton, 1998). Methodologically, this led to the usage of hedonic price indices, which uses qualitative characteristics of products to explain their prices. Secondly, software purchases were now not seen as intermediate goods, but as investments due to their role in enhancing productivity. While controversial, both these methodological changes shaped GDP measurement significantly in this period.

The 2008 financial crisis led to questions about the role of financial services in GDP, as the output share of the sector remained high despite large-scale bankruptcies during this period. Contributions of this sector are measured in a fuzzy manner - in proportion to risks that banks and institutions take relative to their outstanding balances, leading studies to conclude that the sectoral output may be overstated from anywhere between one-half to one-fifth (Haldane & Madouros, 2011). Another emerging concern was the limitations of GDP in incorporating the role of the informal or so-called shadow economy, due to the arbitrary nature of what is measured and what is not. This has changed in recent times, with the use of time-use surveys and other calibrations to incorporate this aspect into GDP calculations.

Finally, studies on the link between prosperity and happiness, as well as challenges like climate change have led to the emergence of movements pushing for expansion of GDP to incorporate sustainability and other aspects of well-being (Easterlin, McVey, Switek, Sawangfa, & Zweig, 2010). While the debate is shifting in many advanced countries, the primacy of growth in lifting people out of poverty in developing countries, as well as helping reduce inequality (Milanovic, 2013) means that GDP will continue to remain a centrepiece of economic discourse in the years to come.

1.2 GDP Measurement: Methodologies

In this section, I briefly describe the three common approaches used to estimate the GDP of a nation, namely the Production or output method, the Expenditure method, and the Income method. Since these are accounting frameworks, the estimates generated by all three methods are equal by definition, notwithstanding

differences due to discrepancies. However, data requirements vary for each method, as the next few paragraphs will make clear.

1.2.1 Production/ Output Method

The production output, following its name, simply adds up the value of all goods and services produced in the economy. This requires estimating gross sales that have taken place over a given period (say a quarter, or a year), and then subtracting the value of the goods used to produce these, also called “intermediate goods,” to avoid the problem of double counting.

$$\text{GDP(O)} = \text{Value of output (less change in inventories)} - \text{Intermediate goods} \quad (1.1)$$

The tracking of intermediate and final goods is often tricky; the development of the input-output method by economist Wassily Leontief formalized the production method systematically, allowing the tracking of the goods going into production and the value of the final output. This method became attractive to estimate GDP in planning-based economies and was widely adopted in the USSR ([Tretyakova & Birman, 1976](#)).

1.2.2 Income Method

The income method looks at the economy from a different point of view and measures the GDP as the sum of incomes earned. All output in the economy yields income from the various factors of production. The first component is the compensation of employees (COE), consisting of wages earned in cash and kind by workers employed, as well as any kind of benefits provided by employers, like social security or a provident fund. The second component includes income earned on property, and income earned from entrepreneurship. The former consists of rent and interest, while the latter consists of profits and dividends. Together, these constitute the Operating Surplus (OS). Finally, there is income earned by “own-account” production, consisting of self-employed workers and unincorporated enterprises, whose incomes do not fit into any of the neat categories defined above. This is called Mixed Income (MI). Adding these up, we get:

$$\text{GDP(I)} = \text{COE} + \text{OS} + \text{MI} \quad (1.2)$$

Statistics for GDP using the income method are usually available only for advanced economies with well-developed statistical departments like the US. This is because estimating incomes requires regular and meticulous data from household surveys, which are often not regularly conducted in developing countries. The income approach serves as a valuable alternative source to validate estimates from other approaches, as detailed further in Chapter - 3.

1.2.3 Expenditure Method

The expenditure approach looks at the economy from the exact opposite vantage point as the production method. While the latter looks backward from a point of view of finished output, the former looks at all the input costs going into the economy, or the sum of expenditures. There are four components estimated in this approach. The first is Consumption (C), consisting of all expenditure by private consumers on goods and services. The second is Investment (I), which is the sum of all capital expenditure undertaken domestically by the private sector. The third component is Government (G), which consists of both consumption expenditure and investment undertaken by the government, while the last component is net exports (X-M), which consists of the difference between money spent by foreigners on domestic goods, and domestic consumers on foreign goods and services. Together, we have

$$\text{GDP (E)} = C + I + G + (X - M) \quad (1.3)$$

This approach is one of the most commonly used to estimate GDP. Increases and shortfalls in spending serve as a clear and precise way of communicating fluctuations in the economy, leading to this framework used often to explain the impact of policy on output. For example, the effect of fiscal stimulus (increase in G) on GDP during recessions is an oft-debated topic.

In India, GDP is estimated using the production and expenditure method. The former estimates gross value added (GVA) by eight sectors of the economy: Agriculture and allied sectors, Mining, Manufacturing, Electricity and utilities, Construction, Trade, restaurants and transport, Finance and Real Estate, and Public Administration and Defense. The expenditure method uses the format described previously. However, there is little to no data estimated using the income method, due to the lack of regular survey data on household incomes.

With this brief overview of the history and measurement of GDP in mind, I now

present the debate on the measurement of India's real GDP in the next section.

1.3 The Indian GDP Debate

India's Central Statistics Office (CSO) is the agency responsible for publishing estimates of GDP since 1950. Following best practices, the CSO also undertakes revisions of the base year used to measure GDP with constant prices. Updating the base year is a routine practice and typically does not affect the sectoral as well as the aggregate growth trend. However, the set of national account statistics published after changing the base year from 2004 to 2011 has been subject to much scrutiny ([Nair, 2019](#); [Sapre & Sengupta, 2017](#)). In this section, I discuss two major critiques: one about the methodology of measuring manufacturing output, while the other tackling the shift in the relationship between the new GDP numbers and some correlates.

1.3.1 The Nagaraj - CSO Debate

prompting an exchange between R Nagaraj, an economist at the Indira Gandhi Institute of Development Research (IGIDIR) and the CSO, in the pages of the *Economic and Political Weekly*, the details of which I summarize below.

Nagaraj begins by pointing out that there is a significant difference in growth rates for the real GDP as estimated using the 2011 and 2004 series, at 5.5% and 4.7% respectively ([Nagaraj, 2015a](#)). The primary difference between the old and new series is driven by changes in measures of gross value added, savings and investment due to a change in the database used by the ministry. The new database called MCA21 has been argued to be better as it contains firm returns filed with the Ministry of Corporate Affairs and has a much larger sample size. The veracity of the new series is questionable, however, due to differences in estimates reported by the sub-committee in its 2014 provisional report and the 2015 final report. PCS savings, investment and GVA shot up by 257%, 34% and 108% respectively between the old and news reports. This combined with the fact that the MCA21 is not publicly available, and that the estimates were not vetted independently was a cause of great concern. Nagaraj stresses that caution must be exercised as the previous database used to measure manufacturing GVA suffered from poor response rates and insufficient filing data. This is supplemented by using a "blow-up" factor to compute the aggregates in proportion

to the paid-up capital. Citing previous work, Nagaraj argues that relative to alternative sources, there has been a significant overestimation in the GVA numbers.

In its reply ([CSO, 2015](#)), the CSO argues that concerns regarding the veracity of PCS output are misplaced. It begins by pointing out that the draft report cited by Nagaraj was a work in progress, and incomplete in many ways. The output in the draft report was restricted to the total revenue reported by the filing companies. However, in the final report the CSO felt that using data from individual components determining the revenue shows the complete picture better; hence the re-estimation and difference in final numbers. It also argues that the scaling-up factor was justified, as there is a deadline for companies to file returns with the database, and incomplete returns were accounted for using this method. Furthermore, paid-up capital for companies who were active but not filed was used to scale up, as no other parameter was available. The other reasons pertained to changes in definitions, usage of individual components of revenue instead of total revenue, and usage of better quality data. Finally, some reclassification and adjustments done to comply with SNA 2008 contributed to the gap between the two figures. Hence, the CSO argues that there is no reason to doubt the veracity of estimates post - 2011.

In his rejoinder ([Nagaraj, 2015b](#)), questions the usage of dis-aggregated revenue data instead of aggregate data, and argues that the drastic difference in estimates indicates a lack of consistency in the MCA21 database. The main critique centres on the usage of the ‘scaling up’ factor: as the number of companies sampled was lower in 2013-14 than the previous year, the scaling-up factor is larger, causing overestimation. He further argues that the problem of PCS estimation precedes this procedure, as very few firms with large output are legally registered, already causing the PCS to skew upwards.

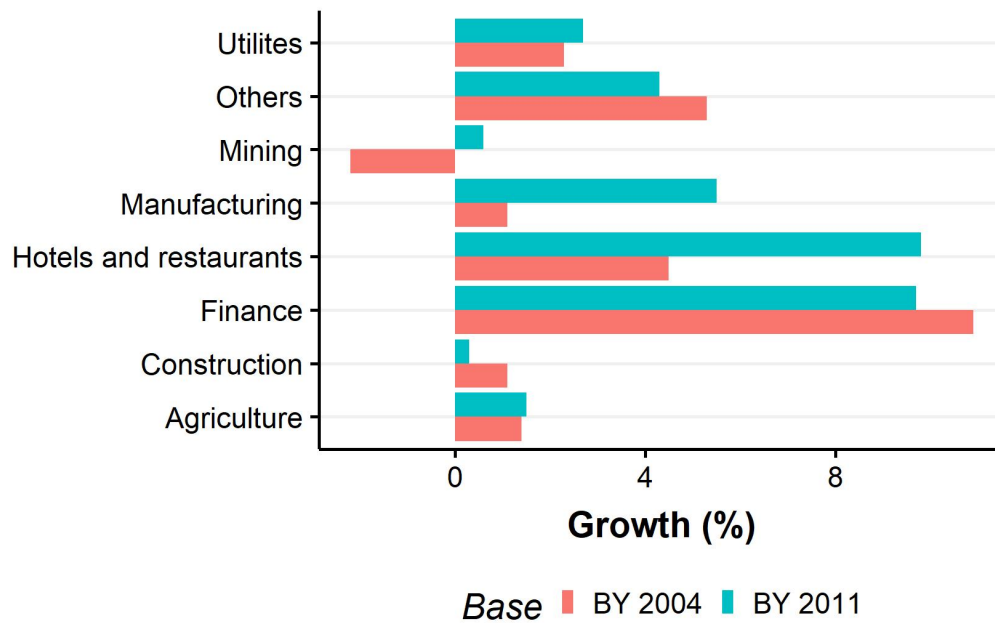


Figure 1.1: Disaggregated GDP growth rates by sector (2012-13)

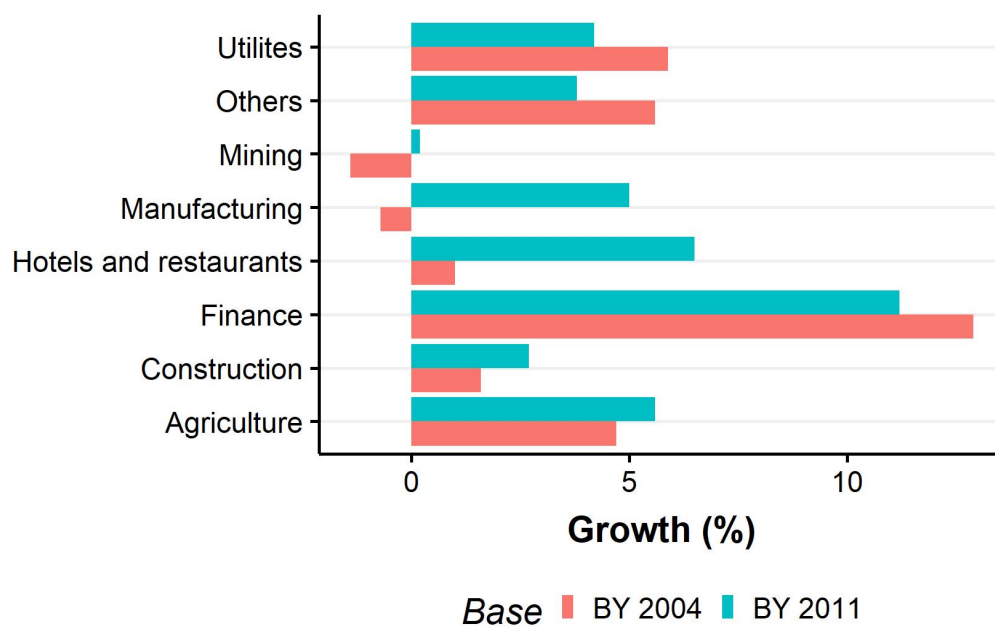


Figure 1.2: Disaggregated GDP growth rates by sector (2013-14)

Figures 1.1 and 1.2 show the disaggregated growth rates for gross value added

by various sectors of the economy, using the 2004 and 2011 base years. Both figures are taken from Nagaraj’s rejoinder. As mentioned before, while the change in base year leads to changes in levels, growth trends remain largely smooth. This principle holds broadly for most sectors in both figures, except for Hotels and restaurants and Manufacturing. The growth numbers as estimated using the 2011 base are significantly larger than those with the old base. For the former, this could be explained due to the reclassification of many sectors to the Trade category, of which Hotels and Restaurants are a subset. However, for the latter, the gap is a cause of concern according to Nagaraj, as the only methodological change pertains to the usage of a new database, and the continued usage of the blowing up factor to account for incomplete data. Especially, for the years 2013-14, the direction of the trends are in opposite directions, with the old base estimating negative growth. Nagaraj concludes that “seeds of doubt” regarding the accuracy of the GDP figures, especially the growth rates of PCS GVA remain.

In a recent summary([Nagaraj, 2021](#)), Nagaraj reiterates these concerns and argues that PCS output remains overestimated. The Nagaraj- CSO debate shows that measurement of manufacturing area output is an area of concern. Other technical issues with the new series are detailed in [Nagaraj & Srinivasan \(2016\)](#).

1.3.2 Subramanian’s Cross-Country Approach

Arvind Subramanian, India’s former Chief Economic Advisor makes the case that India’s GDP is overestimated using a different methodology ([Subramanian, 2019](#)). As a ‘smell test,’ he compiles a list of 17 indicators strongly correlated with GDP growth (electricity consumption, vehicle sales, real credit etc.), and finds that pre-2011, 16 of these are positively correlated with GDP growth, but post - 2011, 11 indicators correlate *negatively*. Given that average growth rates in both periods are similar, this is a concerning finding.

To further test this proposition, Subramanian runs two sets of regressions: first, a cross-sectional specification which regresses GDP growth on electricity, credit, export and import growth, along with their interactions with a time dummy T , which accounts for pre-and-post - 2011 time periods, as well as an India dummy and its interaction with T . The second is a panel specification, regressing the same variables, but with log levels, along with unit and time fixed effects. The coefficient of interest was that of India time dummy

interaction which tells if India's GDP was overestimated post-2011 relative to pre - 2011. The specification in essence follows a difference-in-difference structure, using a sample of more than 80 low and middle-income countries.

The idea is to identify a sample of countries which have a robust correlation between growth and some select indicators, and check if India's growth numbers fall into this broad pattern or not. Subramanian finds that the India time dummy interaction coefficient is statistically insignificant pre-2011 and significant post-2011 at 1% levels, indicating that India is an outlier and that there is an overestimation of output relative to the pre - 2011 period. Using these results, Subramanian argues that there is a 2.5% overestimation in India's GDP annually from 2011-2016. Subramanian's explanation for these results, much like Nagaraj's centres on changes in manufacturing output measurement, particularly the shift from volume-based to value-based measurement. He concludes by calling for a re-examination of National Income Accounting, and transparency in government statistics.

1.4 Conclusion

In this chapter, I provided a broad overview of the history of national income accounting worldwide, as well as different methods used to measure GDP. Following this, I explored two critiques of real GDP measurement in India in recent times. These studies are a cause for concern as accurate measurement of GDP is crucial to economic policymaking in India. Maximizing growth continues to be a cornerstone of developmental policy, with India and China emerging as two of the fastest-growing economies in the 21st century, even as advanced economies are experiencing stable but stagnant growth.

Estimating measurement error precisely requires good quality data from alternate sources, which I detail further in Chapter - 3. The lack of data from the income - side makes this difficult to estimate for countries like India. Hence, I take a different approach, using the synthetic control method (SCM) to estimate a counterfactual GDP using data from countries similar to India, which did not implement a methodological change in 2011. Chapter - 2 goes into the details of the SCM, while Chapter - 3 uses the SCM to estimate measurement error in Indian GDP.

Chapter 2

The Synthetic Control Method

The synthetic control method is a relatively recent innovation in causal inference. It was first introduced by Alberto Abadie and Javier Gardeazabal in their 2003 paper investigating estimating the impact of conflict on GDP in the Basque county ([Abadie & Gardeazabal, 2003](#)). Since then, the method has been used widely to study the impact of policy interventions involving large economic units like cities, states or countries ([Abadie, Diamond, & Hainmueller, 2010, 2015](#)). Nobel laureate Guido Imbens and Susan Athey call it “arguably the most important innovation in the policy evaluation literature in the last 15 years” in a recent review of applied econometrics ([Athey & Imbens, 2017](#)).

In this chapter, I present an overview of the synthetic control method (SCM), beginning with a non-technical introduction. In Section 2.2, I examine the theory behind the method, focussing on the setup, estimation and inference. In Section 2.3, I present the advantages and limitations of the SCM. I review the advances in the literature on alternative estimation methods in Section 2.4. In Section 2.5 I delve deeper into the mechanics of the generalized synthetic control methods (GSCM) and conclude in the last section.

2.1 Introduction

The SCM emerged in the context of comparative case studies. Case studies often study the effect of a policy or intervention on a particular outcome, by comparing it to other determinants of the outcome. For example, a medicine case study focuses on an ailment, while a political case study examines the impact of an electoral strategy. Comparative case studies compare one or more

units exposed to the event or intervention of interest to one or more unexposed units. For example, [Abadie, Diamond, & Hainmueller \(2010\)](#) examines the impact of Proposition 99, a tobacco control program implemented in California and compares its impact on cigarette sales relative to other states. Only some units must be exposed to the intervention, while other comparable units are not.

Comparative case studies have been used in economics for a long time. However, studies typically relied on the comparison between one unit where a policy was implemented (also called ‘treatment’ unit) to a similar unit, where the policy did not take place (also called ‘control’ unit). A classic example is the study by [Card \(1990\)](#). Card examines the impact of the Mariel boatlift, which brought Cuban workers to Miami on the wages of low-skilled workers. He considers various single ‘control’ cities like Houston or Philadelphia, where no such event took place to estimate the difference between wages. The key assumption here is that wages in Miami and the control city would be the same, had the boatlift not taken place. This method is the classic difference-in-difference approach, which has been since used widely in economics([Card & Krueger, 1994](#); [Dube, Lester, & Reich, 2010](#)), with interesting updates to the original methodology ([Callaway & Sant’Anna, 2021](#); [Goodman-Bacon, 2021](#)).

Synthetic control builds upon difference-in-differences, with the key difference being that instead of considering a single control, a weighted average of a group of controls is considered. When the control group is small in number, their weighted average provides a better unit of comparison than any single entity([Abadie, 2021](#)). Specifically, the estimator chooses the weights such that this combination of control units reflects the treated unit closely. This is the approach followed in [Abadie, Diamond, & Hainmueller \(2010\)](#). The authors use a weighted average of cigarette sales, composed of 38 states, to study the impact of Proposition 99 in California.

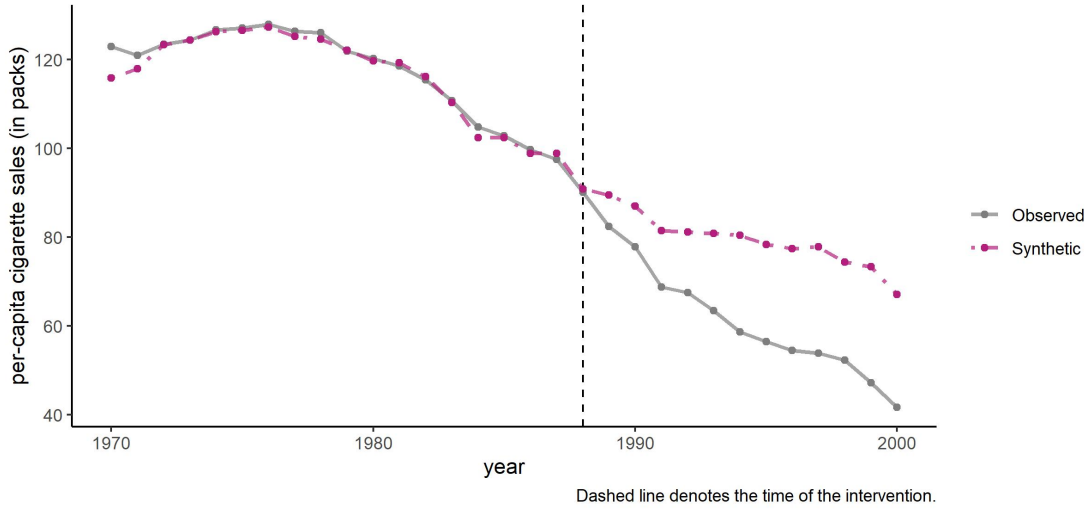


Figure 2.1: Synthetic Control of California

Figure 2.1 shows the per capita cigarette sales in California and synthetic California from 1970-2000. According to the SCM model, synthetic California is an estimation of what cigarette sales in California would have been, had Proposition 99 not existed. The gap between the two trends post the treatment year (1988), shows that cigarette sales would have been significantly higher in the counterfactual, indicating that the law was successful in reducing tobacco consumption in California. Thus, the SCM offers a very clean and intuitive way to estimate the causal effect of a policy intervention with a small number of large units, like cities, counties or countries. In the next section, I introduce the formal model and elaborate on how estimation and inference take place.

2.2 Formal Aspects

To maintain consistency of notation, I stick to the model setup as seen in [Abadie \(2021\)](#).

2.2.1 Setup

We begin by observing data for $J + 1$ units, where $j = 1, 2, \dots, J + 1$. Assume that $j = 1$ is the first unit is the unit where the intervention occurred or the treated unit. The rest of the units, from $j = 2, 3 \dots J + 1$, form the set of units unaffected by the treatment and are referred to as the donor pool. Let the data be observed for T periods, where the first T_0 periods are those before the intervention. The outcome

of interest observed for unit j at time t is referred to as Y_{jt} . We also observe k predictors of the outcome j , X_{1j}, \dots, X_{kj} . The vectors $\mathbf{X}_1, \dots, \mathbf{X}_{J+1}$ contain the values of the predictors for units $j = 1, \dots, J + 1$, respectively. The $k \times J$ matrix, $\mathbf{X}_0 = [\mathbf{X}_2, \dots, \mathbf{X}_{J+1}]$, collects the values of the predictors for the J untreated units.

For unit j in time t , the potential response *without* the intervention is defined as Y_{jt}^N , while for the unit affected by the intervention $j = 1$ at time $t > T_0$, the potential response is defined as Y_{1t}^I . Thus the effect of the intervention at $t > T_0$ is given by

$$\alpha_{1t} = Y_{1t}^I - Y_{jt}^N$$

Since the unit of interest is observed after T_0 , we have $Y_{1t} = Y_{1t}^I$. The unknown response is the counterfactual, which is the evolution of the outcome of interest in absence of the intervention for $t > T_0$. Now, let D_{it} take the value 1 if unit i is exposed to the intervention and 0 otherwise. Then the observed outcome Y_{it} is given by

$$Y_{it} = Y_{it}^N + \alpha_{it}D_{it} \quad (2.1)$$

As only the first unit is exposed to the treatment and only after T_0 , we have

$$D_{it} = \begin{cases} 1 & , i = 1, t > T_0 \\ 0 & \text{otherwise} \end{cases} \quad (2.2)$$

The effect of intervention on unit 1 at time $t > T_0$, given by α_{1t} , is

$$\alpha_{1t} = Y_{1t}^I - Y_{1t}^N = Y_{1t} - Y_{1t}^N \quad (2.3)$$

Since Y_{1t} is observed after the intervention, we simply need to estimate Y_{1t}^N to estimate α_{1t} .

2.2.2 Estimation

Assume that Y_{1t}^N is given by a linear factor model. Initially used as a model of psychological evaluations ([Spearman, 1904](#)), linear factor models are used to model unobserved heterogeneity in economics. In macroeconomics, they are used often to model the unobserved common factors driving the dynamics of time series variables ([Forni, Hallin, Lippi, & Reichlin, 2000](#); [Stock & Watson, 2016](#)). Typically, the idiosyncratic error term has a defined structure, consisting of a limited number of

unobserved common factors, and their parameters, known as factor loadings. The model used to estimate SC is given by

$$Y_{1t}^N = \delta_t + \theta_t \mathbf{Z}_t + \lambda_t \mu_t + \epsilon_{it} \quad (2.4)$$

where δ_t is the time trend, \mathbf{Z}_t consists of covariates not affected by the intervention and λ_t contains unobserved common factors. θ_t and μ_t are unknown parameters and unknown factor loadings respectively. ϵ_{it} is the zero-mean error term. Now, consider a $(j \times 1)$ vector of weights $\mathbf{W} = (w_2, \dots, w_{j+1})$, where each weight is positive and the sum of weights equals one. Thus, the weighted value of the outcome variable is

$$\sum_{j=2}^{J+1} w_j Y_{jt}^N = \delta_t + \theta_t \sum_{j=2}^{J+1} w_j \mathbf{Z}_j + \lambda_t \sum_{j=2}^{J+1} w_j \mu_j + \sum_{j=2}^{J+1} w_j \epsilon_{jt} \quad (2.5)$$

Suppose there exist a set of optimal weights (w_2^*, \dots, w_j^*) , such that,

$$\sum_{j=2}^{J+1} w_j^* Y_{j1} = Y_{11}$$

$$\sum_{j=2}^{J+1} w_j^* Y_{j2} = Y_{12}$$

$$\sum_{j=2}^{J+1} w_j^* Y_{jT0} = Y_{1T0}$$

$$\sum_{j=2}^{J+1} w_j^* \mathbf{Z}_j = \mathbf{Z}_1$$

These set of weights (w_2^*, \dots, w_j^*) are such that the weighted sum of each observed outcome variable for the donor group, in the pre-intervention period $T0$ (Y_{j1}, \dots, Y_{jT0}) equals the respective outcome variable for the treated unit before the intervention (Y_{11}, \dots, Y_{1T0}). Similarly, the weighted sum of donor group covariates equals the covariates for the treated unit. In the Proposition 99 example, all the variables on the right-hand side would be the observed per capita cigarette sales in California before 1988, while the terms on the left-hand side denote the weighted sum of per capita cigarette sales in the donor states. From Equation (2.4), we have

$$\hat{Y}_{1t}^N = \sum_{j=2}^{J+1} w_j Y_{jt}^N \quad (2.6)$$

and

$$\alpha_{1t} = Y_{1t} - \hat{Y}_{1t}^N \quad (2.7)$$

Equation (2.6) states that the estimated synthetic control for the unobserved outcome is given by the weighted average of the observed outcome variables in the donor group. This is true as each observation of the estimated synthetic control (\hat{Y}_{1t}^N) is the weighted average of each observation in the donor group (Equations 6-10). Once the synthetic control is estimated, the treatment effect for the affected unit is simply the difference between the observed outcome and the estimated unobserved outcome in $t > T_0$, as shown in Equation (2.7).

How are the optimal weights determined? [Abadie, Diamond, & Hainmueller \(2010\)](#) argue that the weights should be chosen so that the synthetic control best resembles the pre-treatment values of the predictors of the outcome variable. As stated before, \mathbf{X}_1 contains the pre-treatment values of the covariates which predict the outcome variable for the treated unit, while \mathbf{X}_0 refers to the same for the donor units. Then, the vector of weights \mathbf{W}^* , minimizes

$$\|\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W}\| = \sqrt{(\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W})' \mathbf{V} (\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W})}$$

The matrix \mathbf{V} measures the discrepancy between \mathbf{X}_1 and \mathbf{X}_0 . The choice of \mathbf{V} is an important one and influences the quality of the pre-intervention fit. [Abadie, Diamond, & Hainmueller \(2010\)](#) choose \mathbf{V} , such that the synthetic control minimizes the mean squared prediction error (MSPE)

$$\sum_{t \in T_0} (Y_{1t} - w_2(\mathbf{V})Y_{2t} - \dots - w_{J+1}(\mathbf{V})Y_{J+1t})^2$$

The MSPE is a measure of the distance between the outcome variable of the treated unit (Y_{1t}) and the predicted outcome variable generated by the synthetic control. It is crucial to infer the outcome of the synthetic control, which is explained in the next section.

2.2.3 Inference

At a preliminary level, inference can be done using placebo tests. Placebo tests apply the synthetic control tests to each of the donor units i.e. the units where the intervention does not occur. If the gap between the observed and synthetic values is of a magnitude similar to those seen in the unit of interest, then there is no significant evidence that the intervention had the desired effect. However, if the

magnitudes are smaller, then it can be concluded that the intervention did have a significant effect on the unit of interest.

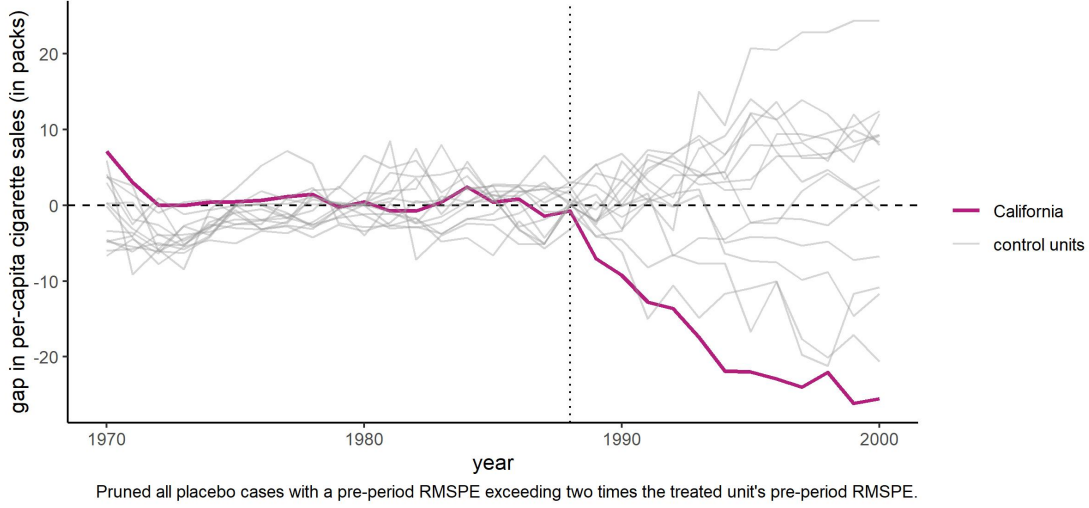


Figure 2.2: Per-capita cigarette sales gaps in California and placebo gaps in 19 control states

Figure 2.2 shows the gap in the per capita cigarette sales in the placebo states as well as California. The cigarette sale gap is highest in the state of California, showing that the effect of Proposition 99 was significant in size. The sample is restricted to 19 states, as units with a pre-intervention MSPE greater than 2 times that of California are discarded. This is done to exclude states which are too similar to California.

While being a useful starting point, placebo tests are limited as the pre-treatment fit of all the placebo units may not closely match the trajectory of the outcome variable well. For this reason, [Abadie, Diamond, & Hainmueller \(2010\)](#) specifies an exact test statistic, which measures the ratio of the post-intervention fit relative to the pre-intervention fit. This is given by

$$r_j = \frac{R_j(T_0 + 1, T)}{R_j(1, T_0)} \quad (2.8)$$

where

$$R_j(t_1, t_2) = \left(\frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} (Y_{jt} - Y_{jt}^N)^2 \right)^{\frac{1}{2}} \quad (2.9)$$

Equation (2.9) defines the mean squared prediction error (MSPE) of the synthetic control estimator over periods t_1 to t_2 . Hence, the numerator of Equation (2.8) is

the MSPE of the post-intervention period, and the denominator is the MSPE of the pre-intervention period. The p-value for this inferential procedure is given by:

$$p = \frac{1}{J+1} \sum_{j=1}^{J+1} I_+(r_j - r_1) \quad (2.10)$$

where $I_+(\cdot)$ is an indicator function taking value one for non-negative inputs and zero otherwise.

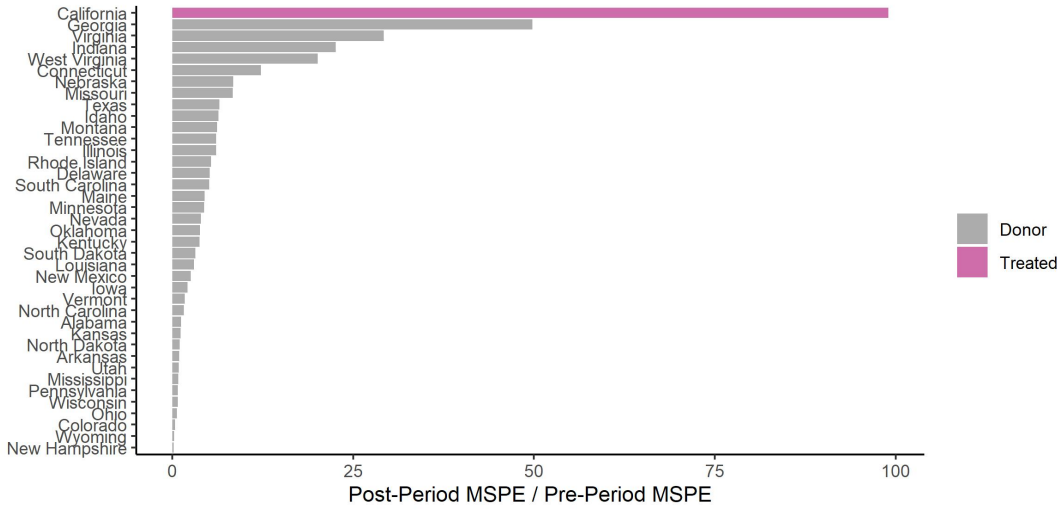


Figure 2.3: Ratio of post and pre-treatment MSPE in California and donor states

To better understand how to interpret this test statistic, consider again the Proposition 99 example. Figure 2.3 shows the value of the test statistic, i.e. the ratio of the post-treatment to the pre-treatment MSPE for the treated state (California) and each of the states belonging to the donor pool. The MSPE is the squared value of the Euclidean distance between the fitted cigarette sales and actual cigarette sales. If Proposition 99 had a significant effect on cigarette sales in California, then the pre-treatment trend of California and synthetic California overlap, and diverge in the post-treatment period. Hence, the ratio of the two MSPEs should be the largest for California, which is exactly what is seen in Figure 2.3. Furthermore, using equation (2.10) the corresponding p-values of this statistic can be computed, as shown in Table 2.1. The p-value for California is about 2.5%, making the test statistic significant at 5%.

Unit name	Type	p-values
California	Treated	0.0256410
Georgia	Donor	0.0512821
Virginia	Donor	0.0769231
Indiana	Donor	0.1025641
West Virginia	Donor	0.1282051
Connecticut	Donor	0.1538462
Nebraska	Donor	0.1794872
Missouri	Donor	0.2051282
Texas	Donor	0.2307692
Idaho	Donor	0.2564103

Table 2.1: p-values for the Post-Pre Intervention MSPE Ratio for select states

2.3 Advantages and Limitations of SCM

The standard synthetic control method, estimated as per [Abadie, Diamond, & Hainmueller \(2010\)](#) has several advantages. The first distinct advantage of the standard SCM is that it gives a transparent fit. Specifically, the discrepancy between the treated unit and the combination of the donor units can be seen clearly. This can be done by computing the difference between the pre-treatment averages of the variables of interest for the treated unit and the corresponding synthetic control. In continuation, the weights assigned to each of the donor units are also easily accessible, which allows the researcher to understand which units are more similar to the treated unit, and which are not. Secondly, synthetic control methods preclude extrapolation due to the constraint that the sum of the weights is one. A regression model, in contrast, allows for negative weights. Finally, SCM provides a safeguard against ‘specification searches,’ which means cherry-picking models which yield a known result. This is avoided in SCM, as all the information used to compute the estimator is taken from the pre-intervention period, while the counterfactual for the post-intervention period remains unknown. Finally, SCM results are geometrically intuitive and can be easily visualized, and interpreted through placebo and exact test statistics.

However, the standard SCM has some limitations, two of which I highlight here. The first, elaborated by [Ferman, Pinto, & Possebom \(2020\)](#), relates to specification searching opportunities. The choice of contributors to the synthetic control is often a matter of discretion, and hence it is possible to try and test several combinations

which yield a desired pre-treatment fit and post-treatment trend. There is significant room for such searches when the number of pre-treatment periods is close to those used in common applications. While the distance function is minimized algorithmically, which limits the bias, the function itself is endogenously chosen by the researcher, creating the potential for bias and “p-hacking” (Cunningham, 2021).

Another common critique centres on the restriction placed on weights: that they must be positive and sum up to one. In technical terms, this is the “convex hull assumption,” wherein the weighted average (or convex combination) of the predictors of the donor pool falls close to the convex hull of these points. This restriction leads to sparse distribution of weights, with a small number of units contributing heavily to the synthetic control, and many units contributing close to zero. Since there are typically multiple predictors, and they are to be determined such that the synthetic control resembles the treated unit both in terms of the outcome and the predictors, the weights of the predictors are the control units are jointly optimized. However, multiple studies have shown that the weights produced by algorithms using such joint optimization procedures are unstable and sub-optimal (Becker & Klößner, 2017; Klößner, Kaul, Pfeifer, & Schieler, 2018). Kuosmanen, Zhou, Eskelinen, & Malo (2021) use insights from game theory to show that the true optimum of this estimation procedure is a corner solution, which assigns all the weight to a single predictor, which defeats the purpose of the SCM. These limitations, along with some others (not discussed here) have led to the development of alternatives to the SCM.

2.4 Alternative Estimators

Since its introduction in Abadie & Gardeazabal (2003), the synthetic control method has become a popular tool of causal inference. However, it has also been criticised for reasons mentioned in the previous section; the critiques have also spawned many alternatives. Some of these are improvements to the standard SCM, while others use different estimation techniques altogether. Ferman & Pinto (2021) deals with the issue of imperfect pre-treatment fit. When the matching of pre-treatment predictors of the treated and the donor units is imperfect, the SCM estimator is biased, even for a large pre-treatment period. However, using a demeaned version of the SCM reduces the bias of the estimator relative to difference-in-differences. Ben-Michael, Feller, & Rothstein (2021) also deals with the issue of bias but instead proposes a bias correction by modifying the regression

specification, called the augmented SCM. The predictor matching discrepancy is corrected in this case, by using a “penalty term,” which is a synthetic control estimator applied to residuals. Extrapolation is penalized with a ridge-linear model, which improves pre-treatment fit. The SCM can also be extended to include multiple treated units, which leads to a multiplicity of possible solutions and more scope for matching discrepancy. [Abadie & L’Hour \(2021\)](#) proposes a penalizer for pairwise discrepancies between treated and control units, in a multiple unit setup.

A subset of the literature uses matrix completion methods for SCM estimation. This involves assuming a nonlinear factor model for the untreated units, allowing for missing values, which are then estimated using matrix completion methods. Linear combinations of the estimated matrix are then used to construct the synthetic controls. This can be done in a setup without covariates ([Amjad, Shah, & Shen, 2018](#)), and with covariates ([Amjad, Misra, Shah, & Shen, 2019](#)). [Athey, Bayati, Doudchenko, Imbens, & Khosravi \(2021\)](#) assume that the untreated units are composed of a matrix and a random error term, and the matrix values are estimated using completion techniques. Their methodology allows for the inclusion of covariates and various fixed effects. Some other papers allow for regression-based weights and extrapolation. [Doudchenko & Imbens \(2016\)](#) allow for negative weights that may not sum up to one, where the weights are regularized using an elastic net (a combination of lasso and ridge) to impose penalties. The augmented SCM also allows negative weights, as does the GSCM ([Xu, 2017](#)). [Arkhangelsky, Athey, Hirshberg, Imbens, & Wager \(2021\)](#) introduces an estimator that includes not only control unit weights but also pre-intervention time weights, called the synthetic difference-in-difference estimator.

Finally, studies have proposed alternatives to the inference procedure adopted by [Abadie, Diamond, & Hainmueller \(2010\)](#), which uses a permutation test (the SC is calculated by considering every possible control unit). [Hahn & Shi \(2017\)](#) show that for the placebo tests to be valid the symmetry assumption must hold i.e. the distribution of the difference between the pre-treatment outcome variable and the estimated synthetic control must be identical. The authors show that this may not always hold for the standard SCM and propose using the following test: comparing the value of treatment effects of $T - T_0$ post-intervention periods to the distribution of the same values computed for every subset for that number of periods. [Chernozhukov, Wüthrich, & Zhu \(2021\)](#) suggests a conformal inference procedure, starting by assuming that the counterfactual

is composed of a mean unbiased proxy and an error term. With the pre and post-intervention outcomes differing only by the treatment effect, the distribution of a function of the post-intervention error terms should be the same as that of a random permutation of the same. In the case of the SCM, this can be tested by setting the proxy as the synthetic control and implementing the test on the residuals computed via least squares. [Chernozhukov, Wuthrich, & Zhu \(2021\)](#) provides a methodology for computing an exact t-test for the SC estimator.

The literature on the synthetic control method is thus rich and developing fast. For the remainder of the chapter, I delve deeper into the generalized synthetic control method, which includes many of the advances in the literature - specifically, the inclusion of multiple treated units, regression-based weights, and robust inference procedures.

2.5 Generalized Synthetic Control

The generalized synthetic control was proposed by Yiqing Xu in 2017. This method combines the synthetic control method and the interactive fixed effects model (IFE). The IFE model, proposed by [Bai \(2009\)](#), is used to model unobserved time-varying confounders. Time-varying confounders refer to unobserved variables that are correlated with both the dependent and independent variables, and which take different values over time. The IFE model interacts unit-specific intercepts, referred to as factor loadings and time-varying coefficients, referred to as latent factors. The generalized synthetic control method (GSCM), unifies the IFE model with the SCM in the following way: first, using only the control group (or donor pool) data, it estimates an IFE model to obtain a fixed number of latent factors. Then, the factor loadings for each treated unit are estimated by linearly projecting pre-treated treated outcomes on the space spanned by this factor. Finally, it computes treated counterfactuals based on the latent factors and the factor loadings.

The key difference between the SCM and GSCM is that the latter does dimension reduction before re-weighting. Specifically, the model selects the number of latent factors to be used algorithmically. This is done using a cross-validation scheme (details in the inference section), which avoids specification searches. The GSCM is also more “general” in the sense that it can be extended to cases with multiple treated units and variable treatment periods.

2.5.1 Framework

In this section and the next one, I stick to the notation used in [Xu \(2017\)](#). Let Y_{it} be the outcome of interest for unit i and time t , and let O and C denote the total number of units in the treatment and control groups, with N total units. All units are observed for T periods, with the pre-treatment periods denoted by $T_{0,i}$ for each unit. The post-treatment period is given by $q_i = T - T_{0,i}$. It is assumed that Y_{it} is given by a linear factor model

$$Y_{it} = \delta_{it}D_{it} + x'_{it}\beta + \lambda'_i f_t + \epsilon_{it} \quad (2.11)$$

where D_{it} equals 1 if i is exposed to treatment, and zero otherwise. δ_{it} is the treatment effect, and x_{it} is a $(k \times 1)$ vector of observed covariates, while β is $(k \times 1)$ vector of unobserved parameters. f_t is a $(r \times 1)$ vector of unobserved common factors and λ_i is a vector of factor loadings. ϵ_{it} is the zero-mean idiosyncratic error term. The parameter of interest is δ_{it} ; more specifically we are interested in estimating the average treatment effect on the treated unit, given by the sum of δ_{it} for each unit, divided by the treatment period.

Let $Y_{it}(1)$ and $Y_{it}(0)$ be the potential outcomes for unit i at time t , when variable D_{it} takes values 1 and 0 respectively. Therefore, we have

$$Y_{it}(1) = \delta_{it} + x'_{it}\beta + \lambda'_i f_t + \epsilon_{it}$$

$$Y_{it}(0) = x'_{it}\beta + \lambda'_i f_t + \epsilon_{it}$$

Hence, we have that $\delta_{it} = Y_{it}(1) - Y_{it}(0)$

Let the control and treated units be subscripted from 1 to N_{CO} and N_{CO+1} to N respectively. Then, stacking the control units together yields

$$Y_{CO} = X_{CO}\beta + F\Lambda'_{CO} + \epsilon_{CO} \quad (2.12)$$

where Y_{CO} and ϵ_{CO} are matrices sized $(T \times N_{CO})$, X_{CO} is of dimension $(T \times N_{CO} \times p)$, while Λ_{CO} is a $(N_{CO} \times R)$ matrix.

2.5.2 Estimation and Inference

We are interested in estimating the treatment effect on the treated unit i , at time t , given by $\hat{\delta}_{it} = Y_{it}(1) - Y_{it}(0)$. $Y_{it}(0)$ is imputed using the following three steps:

1. Estimate an IFE model using only the control group data to obtain

$$(\hat{\beta}, \hat{F}, \hat{\Lambda}_{co}) = \arg \min_{\tilde{\beta}, \tilde{F}, \tilde{\Lambda}} \sum_{i \in C} (Y_i - X_i \tilde{\beta} - \tilde{F} \tilde{\lambda}_i)' (Y_i - X_i \tilde{\beta} - \tilde{F} \tilde{\lambda}_i)$$

$$s.t. \frac{\tilde{F}' \tilde{F}}{T} = I_r, \Lambda'_{CO} \Lambda_{CO} = \text{diagonal}$$

This generates the estimators for common factors, factor loadings and β . The imposed constraints state that all the factors are normalized and that the factor loadings are orthogonal to each other.

2. Estimate factor loadings for the treated unit by minimizing MSPE of the predicted treated outcome in the pre-treatment period:

$$\lambda_i = \arg \min_{\tilde{\Lambda}_i} (Y_i^0 - X_i^0 \hat{\beta} - \hat{F}^0 \tilde{\Lambda}_i)' (Y_i^0 - X_i^0 \hat{\beta} - \hat{F}^0 \tilde{\Lambda}_i), i \in T$$

Note that this step uses data only from the pre-treatment period, while the parameters $\hat{\beta}$ and \hat{F}^0 are taken from the first step.

3. Using the estimated parameters for covariates, common factors and factor loadings, estimate the counterfactual outcome as:

$$Y_{it}(0) = x'_{it} \hat{\beta} + \hat{\lambda}'_i \hat{f}_t, i \in T, t > T_0$$

How is the factor model used to obtain the GSC estimator chosen? [Xu \(2017\)](#) proposes a cross-validation scheme to choose the model before estimating the causal effect. Cross-validation is a resampling method i.e. it involves repeatedly drawing samples from a training set, which is a sub-sample and fitting a model of interest on each sample. In cross-validation, the data is randomly divided into a training set, and a validation set. The model of interest is then fit on the training set as used to predict observations in the validation set. The mean squared error then allows us to check which model is the most precise. For the GSC estimator, a special case is used, namely the leave one out cross-validation (LOOCV). In this case, the validation set consists of a single observation([James, Witten, Hastie, & Tibshirani, 2021](#)).

Broadly, the algorithm performing LOOCV performs the following steps. First, it estimates an IFE model with a given number of factors r , using only control group data. It then assigns one pre-treatment period of the treatment group to the validation set and uses the rest of the pre-treatment data to estimate the held back data. The corresponding MSPE is computed, and then this is repeated for

different values of r . The number of factors that minimizes MSPE (r^*) is chosen.

How is inference carried out using the GSC model? Xu (2017) uses a parametric bootstrap procedure. Bootstrapping is a widely used statistical tool used to quantify the uncertainty associated with a given estimator or statistical tool. Typically, this involves randomly selecting some observations from a dataset, performing the required estimation and repeating this procedure a large number of times. Finally, the standard error of these bootstraps is computed, yielding the required uncertainty estimates. (James, Witten, Hastie, & Tibshirani, 2021)

For the GSC estimator, the following procedure is implemented. We are interested in obtaining the conditional variance of the average treatment effect. First, LOOCV is performed by randomly assigning treatment to a control unit, resampling the rest of the control group, and applying the GSC method to obtain the residuals (or prediction error). These are *simulated* residuals. Then, the GSC method is applied to the original data, yielding the average treatment (ATT) effect, estimates for the parameters and the *predicted* residuals of the control units. Then, a bootstrapped sample is constructed using the two sets of predicted residuals, and the GSC method is applied to it, yielding a new ATT estimate. This, added to the ATT obtained in the previous step is the bootstrapped estimate. Finally, its variance and confidence intervals are constructed using the formula described in Xu (2017).

2.6 Conclusion

In this chapter, I introduced the synthetic control method and examined the formal aspects of setup, estimation and inference. This was followed by a brief overview of the advantages and limitations of the method, and a review of the literature dealing with alternative estimation techniques. Finally, I explored the generalized synthetic control method that extends the standard SCM to a multi-unit setting.

Both the SCM and GSCM estimate counterfactual trajectories for a treated unit of interest, and under a set of assumptions identify the average treatment effect on the treated unit. However, the underlying models, estimation method and inference for both techniques are completely different. The most important difference lies regarding extrapolation: the GSCM allows for extrapolation and admits negative weights, while the SCM does not. The GSCM assumes a linear

factor model with interactive fixed effects, while the SCM has no IFE. The SCM matches on predictors of the outcome of interest and does not use data on the outcome of interest itself. However, the GSCM uses both and hence allows estimating the ATT with and without covariates. Finally, the SCM relies on ratios of pre and post-treatment RMSPE and computes Fisher's exact p - values for inference, with no way to compute confidence intervals. The GSCM uses a parametric bootstrap approach to compute variance and confidence intervals for the estimated ATT. Table 2.2 summarizes the difference between the two.

Criteria	SCM	GSCM
Underlying model	Linear factor model	Linear factor model with IFE
Matching	Covariates Only	Covariates and Outcome Variable
Inference	Fisher's exact p - values	Parametric bootstrap
Extrapolation	No	Yes
Multiple Treated Units	No	Yes
Variable Treatment Periods	No	Yes

Table 2.2: Differences between the SCM and GSCM Estimators

Both the SCM and GSCM provide different but simple and powerful ways to establish the causal effects of interventions on a variable or a group of variables of interest. While the interpretation of the final estimate(s) of interest are similar, it is important to remember the vast differences in the underlying models and estimation strategies. Hence, appropriate care must be taken and the caveats kept in mind while applying these estimators to study questions of economic policy. In the next chapter, I implement the SCM and GSCM to estimate counterfactual trajectories for India's GDP.

Chapter 3

Estimating measurement error in India's GDP using SCM

Chapter - 1 introduced the issue of measurement error in GDP estimation, focusing specifically on the debate in India. The lack of income-side estimates and methodological opacity makes it difficult to assess the extent to which India's GDP is over/underestimated. Experts are forced to rely on sector-specific anomalies or rely on contradictions observed in fast-moving sectoral data, making it difficult to make causal claims about aggregate data. The key difficulty, in summary, is to estimate a counterfactual GDP series.

The synthetic control method offers to be a promising alternative in this regard. Chapter - 2 establishes the SCM to be a popular tool to find the causal effect of a policy/intervention on a large unit like a country, state or city. A clean research design, backed with minimal data is sufficient to construct a plausible counterfactual, making its usage attractive for the Indian case, where quality data is scarce.

In this chapter, I use the synthetic control method and the generalized synthetic control method to construct a counterfactual series for India's GDP. The synthetic control is constructed using a donor pool of countries similar to India - in this case, a set of emerging market economies as defined by the IMF. The treatment occurs in the year 2011 when the base year of the GDP series was changed from 2004-05 to 2011 - 12, along with the change in the methodology of measurement. I find that XXX,

The rest of this chapter is structured as follows. Section 3.1 presents a brief

overview of the literature on measurement error in GDP and application to synthetic control for GDP. Section 3.2 specifies the econometric methodology and discusses the data and the sources. Section 3.3 presents the key results, and Section 3.4 consists of robustness checks. The final section discusses the findings, and their implications and concludes.

3.1 Literature Review

Measurement error in macroeconomic data and GDP, in particular, has been studied extensively. In recent times, scholars have focused on improving measurement. Two strands of literature have emerged: the “forecast” error approach examines the difference between forecasted or revised estimates and actual GDP, and characterizes, whether revisions represent measurement errors or “noise,” or an efficient forecast accounting for all relevant information, or “news” (Aruoba, Diebold, Nalewaik, Schorfheide, & Song, 2013; Mankiw, Runkle, & Shapiro, 1984; Mankiw & Shapiro, 1986). The second strand, which I focus on, looks at the differences between the income - method and expenditure - method estimates of GDP, as described in Chapter - 1. Since both quantities are estimates of the same unobservable (actual output), the gap between the two is only due to a measurement error. Aruoba, Diebold, Nalewaik, Schorfheide, & Song (2016) show that information from *both* estimates can be used to optimally extract the true GDP. By treating both estimates as noisy, and using Bayesian estimation with multiple types of measurement error models, they construct a third estimate, which diagnostic tests show as being broadly representative. Jacobs, Sarferaz, Sturm, & Norden (2022) uses a similar framework but incorporates multiple revised estimates of noisy data, instead of a single vintage. The authors show that their estimates are less likely to undergo revisions compared to Aruoba, Diebold, Nalewaik, Schorfheide, & Song (2016). Their estimates also put more weight on expenditure-side data, which is more readily available, and whose historical decompositions show a larger share of “news” relative to the income-side estimates. Chang & Li (2018) takes a different approach, by using a pre-analysis plan to estimate the effect of measurement error on the results of empirical papers in top journals. The authors find that replacing GDP with its revised versions does not change the key results, but using GDI instead changes the results substantially. Overall, the existing literature highlights the importance of reconciliation between various sources and releases of GDP estimates.

The SCM and its various offshoots have been applied extensively to estimate counterfactual GDP. [Abadie & Gardeazabal \(2003\)](#) study the impact of terrorism in the 1960s on the per-capita GDP in Basque County. The paper finds that there was a 10% decline in growth relative to the synthetic control not affected by terrorism. [Abadie, Diamond, & Hainmueller \(2015\)](#) examines the impact of the German reunification in 1990 on West Germany's per capita GDP, and find a negative effect relative to synthetic West Germany. [Billmeier & Nannicini \(2013\)](#) looks at the impact of economic liberalization on several Latin American, Asian and African countries' real GDP, and find that while the impact was positive for most, there was no significant impact from the later liberalization episodes, especially those in Africa. Most of these studies focus on the GDP of European countries, with papers on developing economies being much more scarce.

[Mayberry \(2022\)](#) examines the effect of the Nuclear Weapons Development programme on Pakistan's GDP, and finds that per capita GDP would have been higher by \$718 on average, had the program not been implemented. [Singhal & Nilakantan \(2016\)](#) study the impact of counterinsurgency programs on the Naxalite (Maoist) movement and find that implementing the same in one state led to a significant rise in its per capita Net State Domestic Product, relative to a synthetic control of other states where such policies were not implemented. This study differs from the existing literature in two ways: one, to my knowledge this is the first paper implementing the SCM to construct a counterfactual aggregate GDP series for India. Second, this study is also the first to use the SCM to study the impact of the base year policy change on GDP, which differentiates it from the studies carried out so far, which look at GDP revisions and sectoral data (see Chapter -1 for a detailed survey).

3.2 Econometric Approach

The empirical exercise in this chapter aims to construct a counterfactual real GDP series for India. The setup is as follows: I compare the trajectory of India's actual GDP to the trend of "synthetic" India, constructed using a weighted average of predictors and the real GDP of donor countries. The treatment happens in the year 2011, when the Central Statistical Organization (CSO) changed the base year from 2004 to 2011, as well as started using the MCA21 database to estimate manufacturing output instead of the RBI database. Hence, the specific intervention is the *change in methodology used to measure sectoral output*. The assumption

here is that no such exogenous intervention took place in the donor countries.

The gap between actual and synthetic GDP is the estimated measurement error for India. However, this error is not the same as the classical measurement error, where the unknown observable (latent variable) is measured with an additive, random error term. The treatment, in this case, is simply a change in methodology, and hence the error is the discrepancy that would have arisen had there been a comparable alternate measure of Indian GDP. The error is defined through an accounting rather than an econometric framework.

I now present the identification strategy using the SCM and GSCM. As both models have been discussed in detail in Chapter - 2, I only outline the salient details here. Let the effect of change in methodology on India's GDP be $\alpha_{1t} = Y_{1t} - Y_{1t}^N$, where Y_{1t} is India's observed real GDP, while Y_{1t}^N is the counterfactual, which is assumed to have the underlying factor model as follows:

$$Y_{1t}^N = \delta_t + \theta_t \mathbf{Z}_t + \lambda_t \mu_t + \epsilon_{it} \quad (3.1)$$

where δ_t is the time trend, \mathbf{Z}_t consists of covariates not affected by the intervention and λ_t contains unobserved common factors. θ_t and μ_t are unknown parameters and unknown factor loadings respectively. ϵ_{it} is the zero-mean error term. The covariates include gross investment share, the trade share (sum of imports and exports to GDP), as well as the gross primary school enrollment and secondary school enrollment ratios. The effect of the intervention on Indian GDP is given by

$$\hat{\alpha}_{1t} = Y_{1t} - \hat{Y}_{1t}^N = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}^N \quad (3.2)$$

where Y_{jt}^N is the observed GDP of donor country j in absence of the treatment, and w_j^* is the optimal weight assigned by the SCM, under the restriction that the sum of weights is one and that they are positive.

Maintaining consistency of notation, the counterfactual GDP is measured in the GSCM via a linear IFE model given by:

$$\hat{Y}_{1t}^N = x_{1t}' \hat{\beta} + \hat{\lambda}_1' \hat{f}_t \quad (3.3)$$

where x_{1t} consists of the covariates described above, $\hat{\lambda}_1$ are the estimated common

factors for India, and \hat{f}_t are the estimated factor loadings. The steps used to estimate the parameters are described in Chapter - 2.

3.3 Data

In this section, I introduce the data used for the construction of the synthetic control and describe the sources. The synthetic control consists of countries similar to India, without the treatment. I choose 19 countries from the International Monetary Fund's (IMF) World Economic Outlook List of Emerging Market Economies (EMEs). A country is classified as an EME if it demonstrates strong and stable growth, production of high value-added growth, and increased integration in global trade and financial markets. The following countries (other than India) were classified as being EMEs over the period 2010 - 2020: Argentina, Brazil, Chile, China, Colombia, Egypt, Hungary, Indonesia, Iran, Malaysia, Mexico, the Philippines, Poland, Russia, Saudi Arabia, South Africa, Thailand and Turkey.

The dataset consists of 551 observations over the period 1991 - 2019. The outcome variable of interest is the log real GDP in local currency units (LCU). Following the literature on growth and applications of synthetic controls to GDP ([Acemoglu, Naidu, Restrepo, & Robinson, 2019](#); [Billmeier & Nannicini, 2013](#)), I include the following covariates: Gross Investment as a share of GDP, trade (sum of exports and imports) as a share of GDP, Gross Primary School Enrollment and Gross Secondary School Enrollment. The data for all variables is obtained from the World Development Indicators (WDI) database maintained by the World Bank. Table 3.1 shows the summary statistics for the outcome variable as well as the pre-

Table 3.1: Summary Statistics

Variable	N	Mean	St. Dev.	Min	Max
GDP (log)	551	30.0	2.7	26.0	37.0
Investment	538	23.0	6.4	12.0	44.0
Trade	547	64.0	39.0	14.0	220.0
Primary Enrollment	551	106.0	9.6	78.0	166.0
Secondary Enrollment	551	83.0	17.0	30.0	121.0

dictors. The average log GDP is 30, while the average investment and trade ratios are 23% and 64% respectively. There is a large variance in the trade ratio data, as the dataset includes countries which rely heavily on trade like Saudi Arabia, as

well as relatively closed economies like Argentina. The Gross Primary Enrollment Ratio which is the total enrollment divided by the population of the relevant age group is above 100 at the mean, which happens due to the inclusion of over-aged and under-aged students, as well as grade repetition. The Secondary Enrollment Ratio measures the same, but for the secondary school age group (15-18 years).

Observations for GDP are available for all years, and trade and investment for most years. There were a significant amount of missing values (between 80 - 100), for the two enrollment ratios. However, as the gaps were not longer than a few years, and data were not missing for a significant chunk of the period for any country, the missing values were filled via linear interpolation.

The data was extracted using the R package `wdi`, while the SCM and GSCM were implemented using the packages `tidysynth` and `gsynth` respectively.

3.4 Results

I begin by presenting the counterfactual GDP estimates first using the SCM, and then the GSCM.

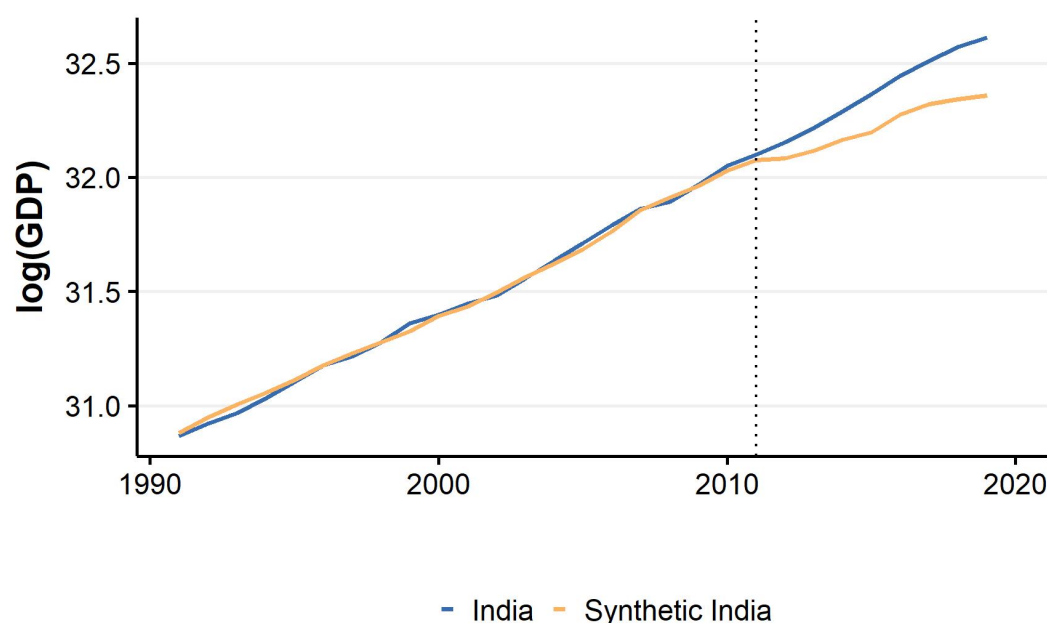


Figure 3.1: Evolution of Log GDP for India and Synthetic India (SCM)

Figure 3.1 shows the trend of log GDP for India and the estimated counterfactual over the period 1991 - 2019. The counterfactual tracks the actual log GDP trend very closely pre - 2011, and starts diverging from the actual trend post - 2011. Given the assumptions of the SCM, the synthetic control is representative of how India's GDP would have evolved in the absence of the change in measurement methodology.

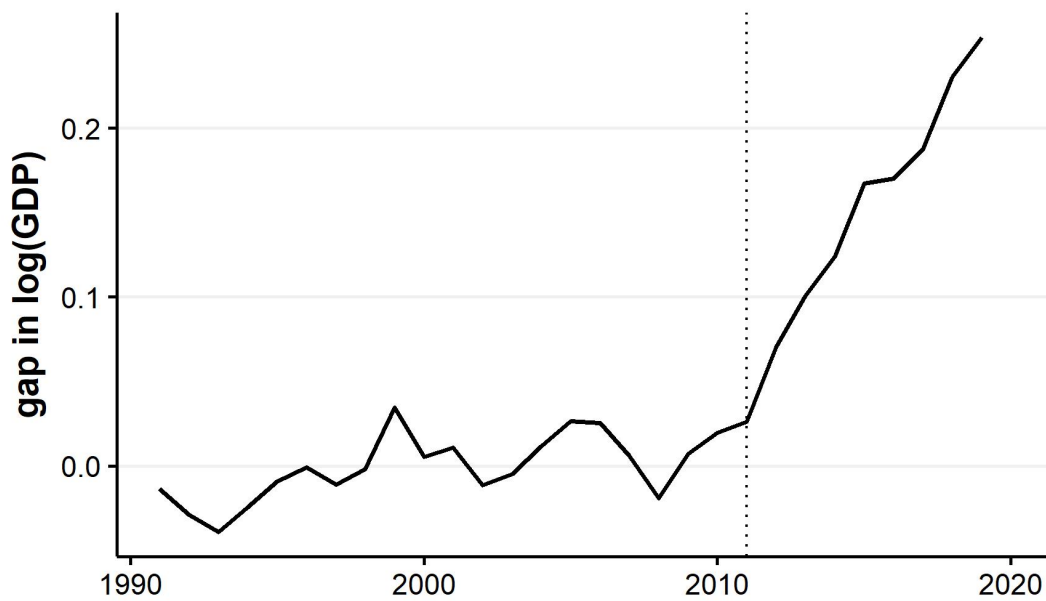


Figure 3.2: Log GDP gap between India and Synthetic India (SCM)

The same result is shown differently in Figure 3.2. This plots the difference between the log GDP trends for actual and synthetic India. As evident from Figure ??, the difference remains close to zero pre - treatment and then increases steadily post - 2011, reaching close to 0.25 log points.

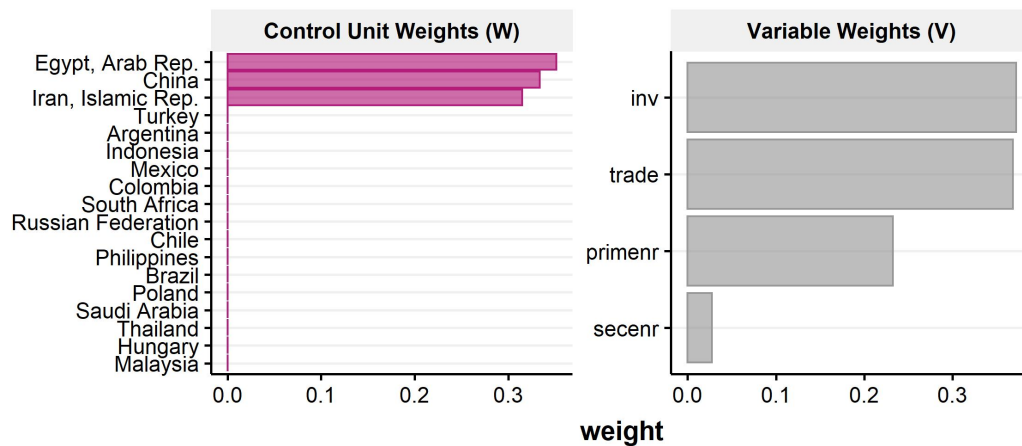


Figure 3.3: Variable weights and country weights for synthetic India

Which countries is synthetic India composed of? Figure 3.3 shows the optimal country weights, as well as the weights assigned to each predictor. From the left panel, it is clear that synthetic India is composed of Egypt, China and Iran, with each contributing about one-third to the total weights. The other countries contribute close to nothing. This sparse distribution of weights is a result of the restriction of positive weights summing to one - this prevents extrapolation, and assigns the weights such that the synthetic control lies within the convex hull. The right panel shows the contributions of the predictors: Investment and trade contribute a large share (around 70%), followed by primary enrollment.

I now present the results of the estimates generated by the GSCM below. Three specifications were run: without covariates, with economic covariates (trade and investment) only, and with all covariates.

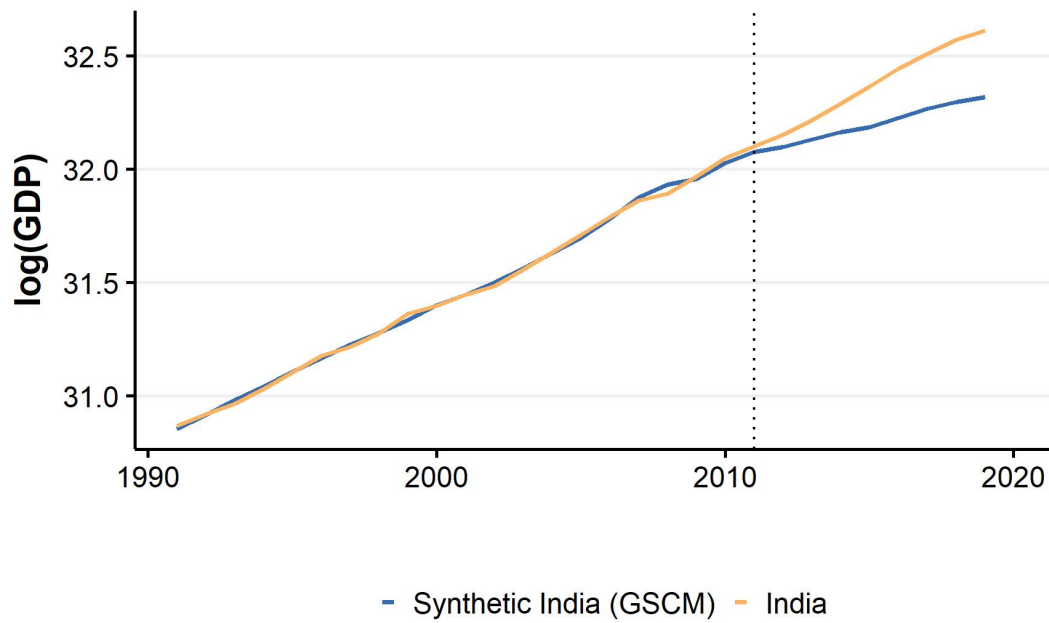


Figure 3.4: Evolution of Log GDP for India and Synthetic India (GSCM)

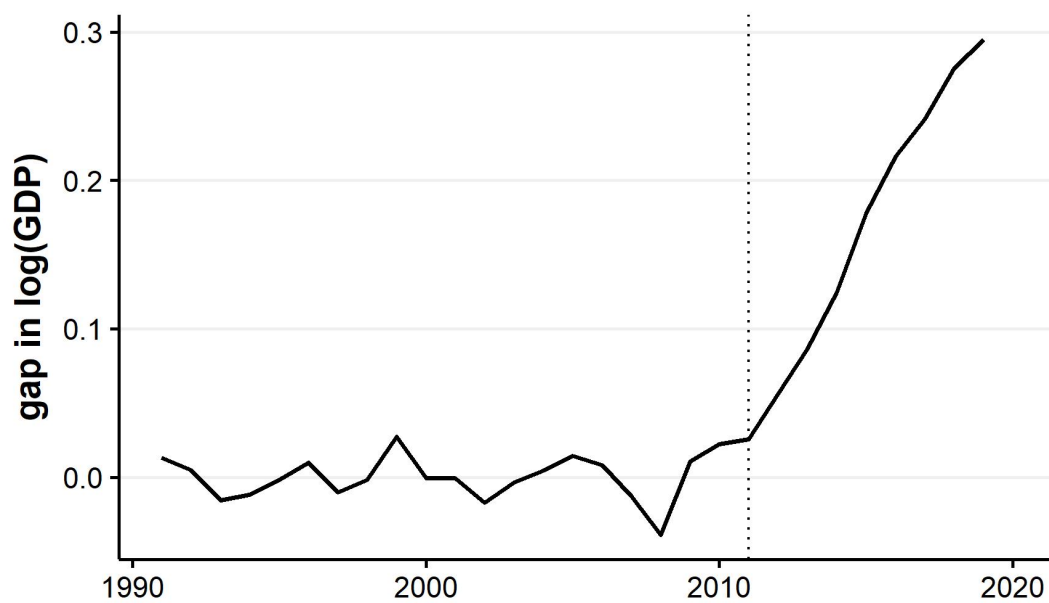


Figure 3.5: Log GDP gap between India and Synthetic India (GSCM)

Figures 3.4 and 3.5 show the estimated counterfactual log GDP and actual

GDP, and the gap between the two respectively, as measured using the GSCM. As with the SCM, the pre-intervention fit of the synthetic control matches the actual trend and then diverges post - 2011. While the two results look similar, the magnitude of the gap post-treatment is different, with the difference exceeding 0.3 log points in the case of the GSCM.

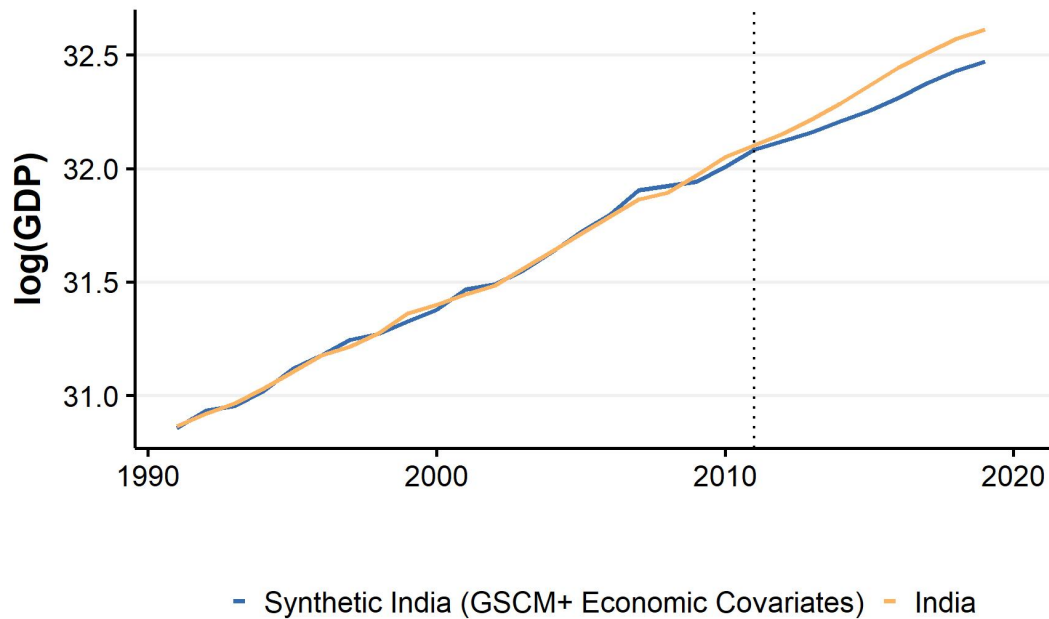


Figure 3.6: Evolution of Log GDP for India and Synthetic India with economic covariates (GSCM)

Figures 3.6 and 3.7 show the same results as the images above, but with the inclusion of economic and all covariates respectively. In the first case, the gap post-treatment persists, but at a lower magnitude (around 0.15 log points). In the second, case the post-treatment fit and the pre-treatment fit barely diverge, indicating that the treatment had no impact on India's GDP when education covariates are included. Hence, the GSCM results are robust to the inclusion of economic covariates, but not to the additional inclusion of education predictors.

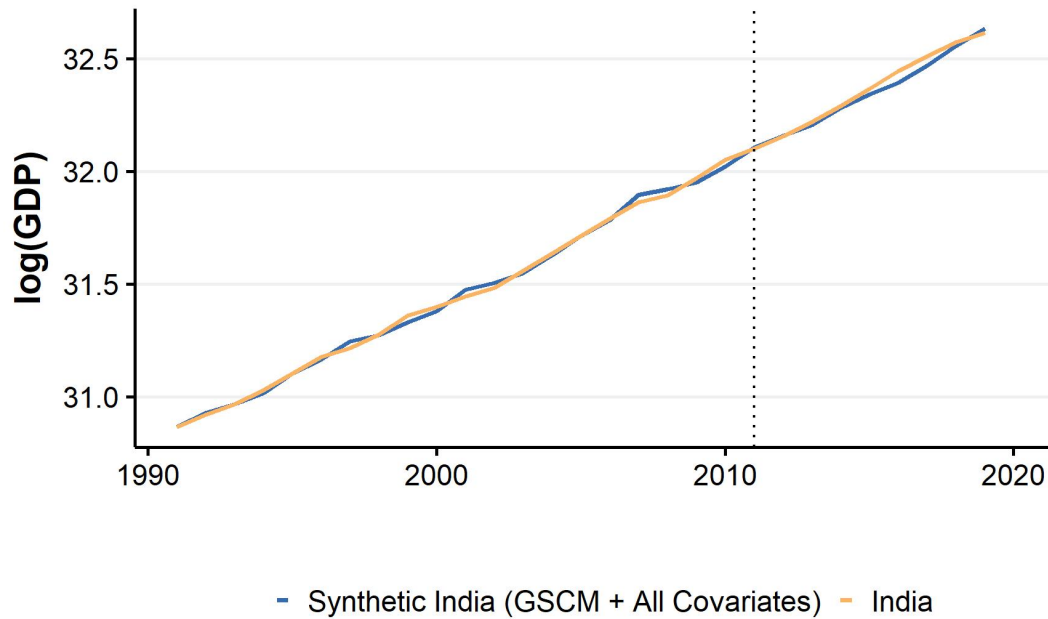


Figure 3.7: Evolution of Log GDP for India and Synthetic India with all covariates (GSCM)

Year	SCM	GSCM	GSCM+EC	GSCM+AC
2011	0.63	0.35	-2.52	-3.31
2012	4.45	3.06	1.38	-0.03
2013	3.00	3.00	2.47	1.58
2014	2.33	3.83	2.35	-0.31
2015	4.34	5.38	2.96	1.54
2016	0.27	3.79	2.26	2.90
2017	1.73	2.51	0.18	-1.23
2018	4.29	3.45	0.89	-2.32
2019	2.35	1.93	-0.25	-3.72
Avg.	2.6	3.0	1.1	-0.5

Table 3.2: Difference between growth rates (percentage) of actual and synthetic GDP (All specifications)

Table 3.2 shows the difference between the actual and synthetic GDP growth rates for the period 2011 - 2019. The first column shows the growth gap for results estimated using the SCM: except for the year 2016, the difference in growth has exceeded 1.5% every year, yielding an average gap of 2.6% over the whole period. The magnitude is larger for the GSCM estimates without covariates, shown in the second column. Post-2011, the gap has consistently been above 3%, even crossing

5% in 2015. The average growth gap in this period broadly matches the findings of Subramanian (2019). The inclusion of economic covariates does reduce the average growth gap to 1.1%, but this is largely due to the period 2017 - 2019; from 2012 - 2017, the average gap exceeded 2%. The inclusion of education covariates, leads to a significant fluctuation in the growth gap, as seen in the last column. With the gap varying between 2.9% to -3.7% over this period, the average growth gap of -0.5% suggests that there was no significant difference between actual and synthetic India's GDP growth rates.

3.4.1 Inference

The last section presented the results of the estimated counterfactual GDP using two synthetic control techniques. How significant are these findings? In this section, I present the results of the various inference tests used to assess the credibility of the results obtained using the SCM and the GSCM.

As discussed in Chapter - 2, Fisher's exact p -values evaluate the probability of getting the same result had the treatment been randomly assigned to some other country, instead of India. Table - 3.3 presents the p - values below. The p-value

Country	Type	p - value
India	Treated	0.053
Philippines	Donor	0.105
Mexico	Donor	0.158
Brazil	Donor	0.211
Argentina	Donor	0.263
Colombia	Donor	0.316
Chile	Donor	0.368
Indonesia	Donor	0.421
Russian Federation	Donor	0.474
Turkey	Donor	0.526

Table 3.3: Fisher's exact p - values

for India is 0.053, indicating that there is a 5.3% probability of getting the same results as those for India. However, the size of the p-value has a lower bound by construction, as it is simply a function of the number of donor units, in this case, 1/19. This makes it an unreliable metric for inference.

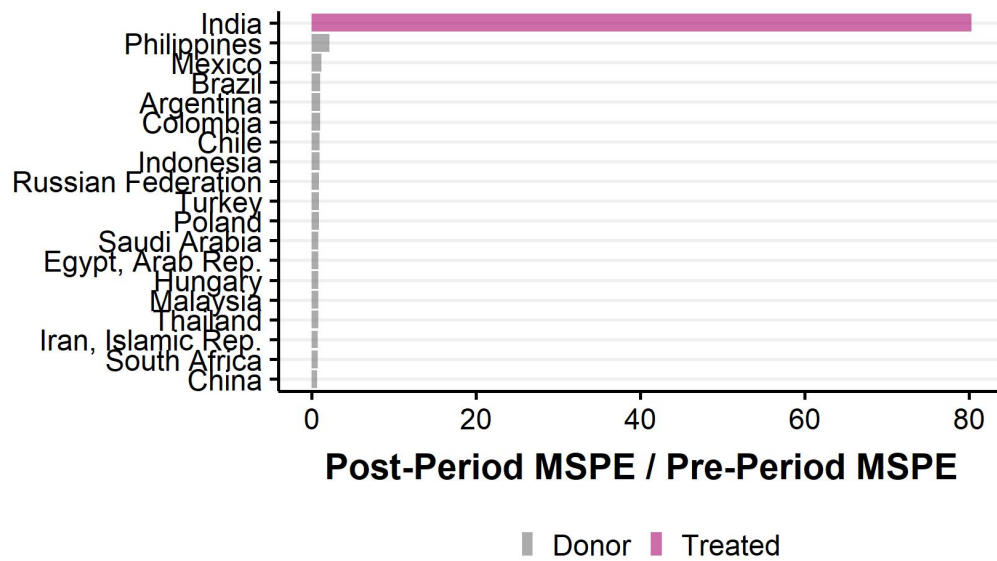


Figure 3.8: Ratio of post and pre-treatment MSPE for India and donor countries

Figure 3.8 shows the ratio of post-period and pre-period MSPE. The MSPE is a measure of how close the fit between the actual outcome variable and the synthetic control is. The figure shows the ratio of the MSPE calculated post-2011 and pre-2011 for all donor countries as well as India, serving as a placebo test. The ratio is much higher for India than for the donor pool, indicating that the post-treatment fit deviates significantly from the pre-treatment fit to a much larger magnitude relative to other countries. While this is a good starting point to establish the credibility of the synthetic control, it is not a sampling-based statistical test (like a t-test). In the case of the SCM, this is difficult to implement due to the absence of a well-defined sampling mechanism. Hence, additional robustness tests are required, which are discussed in the next section. Table 3.4 shows the average Average Treatment Effect on the

Model	Average ATT	Std. Err.	p - value
Baseline GSCM	0.166	0.256	0.516
GSCM with economic covariates	0.094	0.139	0.499
GSCM with all covariates	0.013	0.154	0.928

Table 3.4: Average ATT and p - values for GSCM specifications

Treated, which is simply the mean of the difference between the actual and counterfactual log GDP estimated via GSCM, and the associated p-values. It is important to note that the p-values are calculated in a completely different way from the SCM: rather than a permutation test, uncertainty bounds and variance are estimated using a parametric bootstrap (for details, see Chapter - 2). The results indicate that the mean ATT magnitude falls with the inclusion of covariates and that none of the specifications yields statistically significant results.

The lack of significant results could be due to two reasons: one, while the GSCM does allow the inclusion of covariates with an incomplete panel, the dropping of observations reduces the sample size by 50, which is 10% of the sample. Reduced sample size due to sparse covariate data affects the unbiasedness of the GSCM estimator significantly. Second, Xu (2017) cautions against a straightforward interpretation of results when the number of donor units is less than 40. Since I use only 19 donor countries, the results are prone to bias.

3.5 Robustness Checks

I implement two robustness checks:

Backdating: The treatment is assigned a few years/quarters before the actual intervention, and the SCM and GSCM are estimated with the new pre and post-treatment periods. The results are robust if:

- a. The pre-treatment fit does not change between the new and old treatment year, indicating the absence of any anticipation effect.
- b. The gap between the synthetic and actual outcome variable of interest starts around the same time as the actual treatment.

I implement the SCM and GSCM by backdating the treatment to the year 2005. This year is chosen to allow sufficient pre-intervention periods (15 years), as at least 10 years is the minimum recommended for effective implementation of the GSCM.

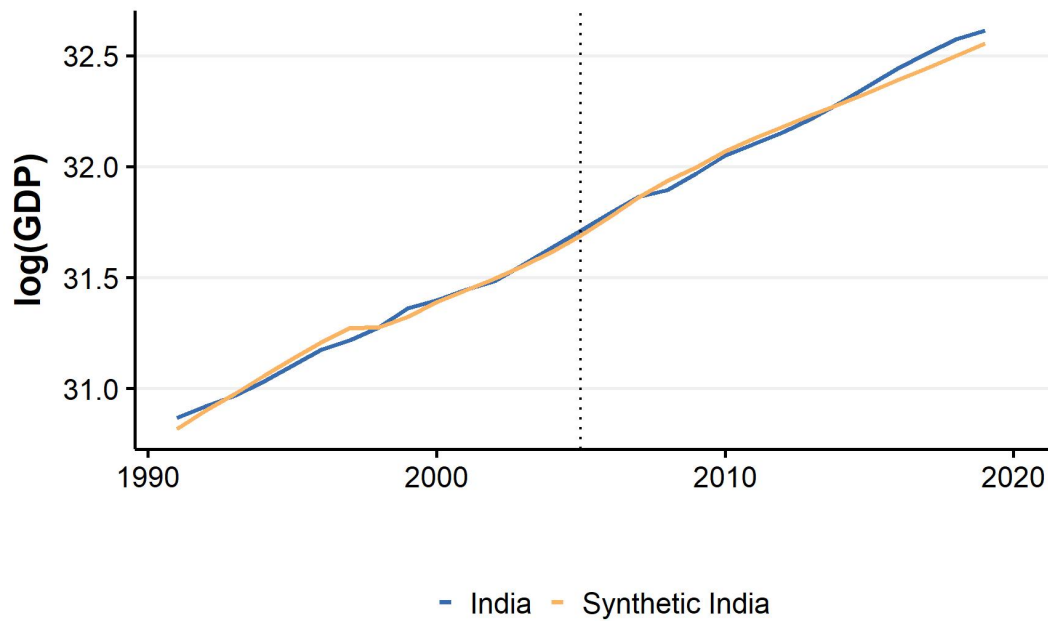


Figure 3.9: Evolution of log GDP for actual and synthetic India with backdated treatment (SCM)

Figure 3.9 shows the results of backdating implemented using the SCM. Between 2005 and 2011, the synthetic India line closely tracks the actual India line, meeting the first criterion, and ruling out anticipation effects. However, this pattern continues post - 2011 as well, and the first signs of divergence appear only near 2016, five years after the actual treatment, which raises concerns about the credibility of the SCM estimates.

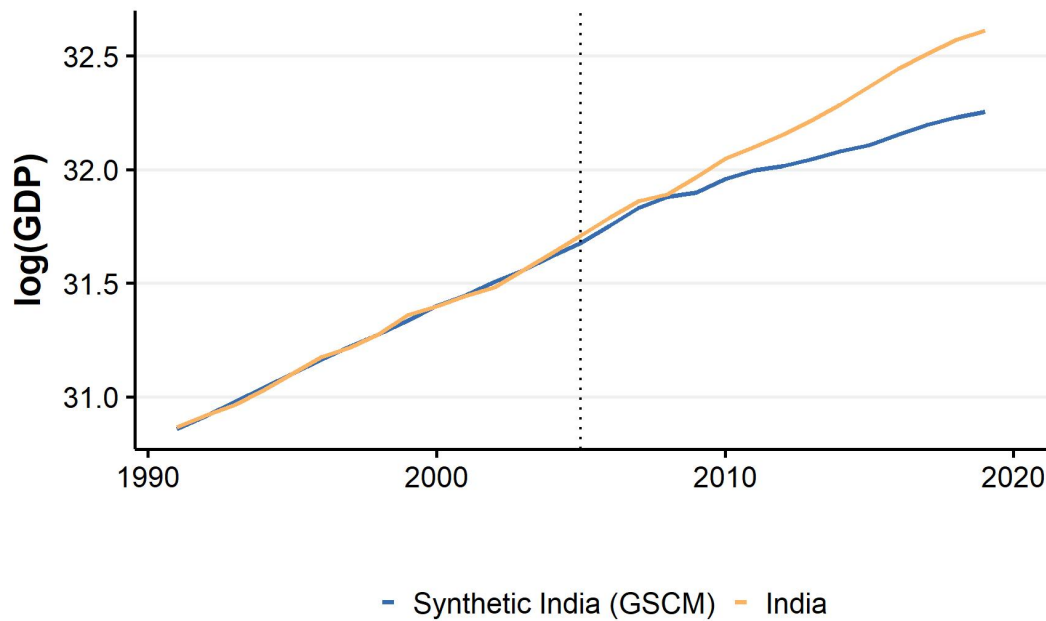


Figure 3.10: Evolution of log GDP for actual and synthetic India with backdated treatment (GSCM)

Figure 3.10 shows the results of backdating implemented using the GSCM. As previously, there is no gap between the two lines for the period 2005 - 2011. However, the two trends begin to diverge around the actual treatment year, satisfying both conditions for robustness. The backdating procedure indicates that the GSCM estimates are robust relative to SCM.

Quarterly data: This robustness check replaces yearly data with quarterly data for the outcome variable. This provides a more granular measure of the GDP, and significantly increases the number of observations, attenuating the bias of both estimators. I implement this only for the GSCM, as the SCM requires data for predictors, which was not available quarterly.

The data for quarterly real GDP is taken from the International Monetary Fund's (IMF) International Financial Statistics (IFS) database. The same donor countries were used. The period is from 1996 Q2 to 2019 Q4, yielding a dataset of 1147 observations.

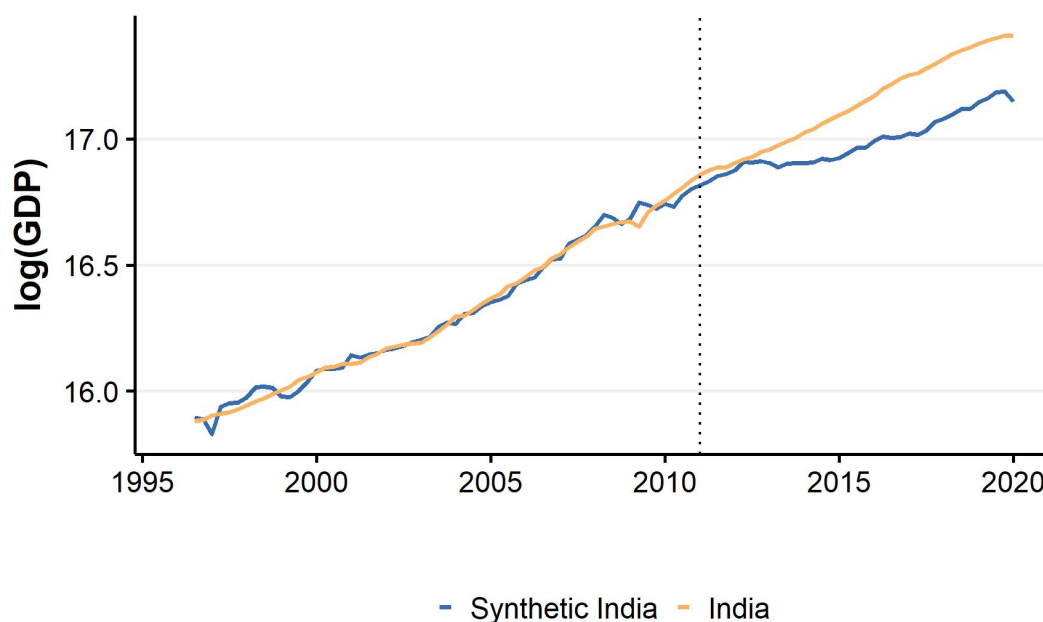


Figure 3.11: Evolution of log GDP for actual and synthetic India with quarterly data (GSCM)

Figure 3.11 shows the results of the above robustness test estimated using the GSCM. The pre-treatment fit remains good before 2011. The divergence begins not immediately in 2011 Q1, but almost a year later. However, the overall patterns evolve similar to the SCM and Baseline GSCM results, confirming the robustness of the GSCM results to the use of quarterly data.

3.6 Discussion and Conclusion

In this chapter, I use the SCM and GSCM to estimate a counterfactual trajectory of India's real GDP. In doing so, I attempt to quantify the extent of conjectured GDP overestimation post - 2011. The SCM results suggest that real GDP growth has been over-estimated by 2.6% on average since the change in the base year. However, these results are sensitive to specification, and not fully robust to backdating. The GSCM results, with and without covariates imply that the over-estimation has been between -0.5% and 3% on average. The baseline GSCM results are robust to backdating and usage of quarterly data, but statistically insignificant, as are the results with covariates.

As discussed before, both the SCM and GSCM perform poorly with a low sample size, which is the case in the specifications with predictors, as several observations are missing for many countries. Furthermore, a caveat to keep in mind while interpreting these results is that inference for synthetic control and its class of models is an evolving enterprise and that any results, significant or otherwise need to be interpreted with caution.

With regards to the two issues mentioned above, multiple promising avenues have opened up, which serve as a natural extension to this project. The first pertains to the development of methods of conformal inference, and exact t-tests by Victor Chernozhukov and co-authors ([Chernozhukov, Wuthrich, & Zhu, 2021](#); [Chernozhukov, Wüthrich, & Zhu, 2021](#)). As these methods of inference apply to a large class of causal models, their application to generate uncertainty bounds for SCM results seems like a feasible next step. The second avenue is the development of matrix completion methods for causal models ([Athey, Bayati, Doudchenko, Imbens, & Khosravi, 2021](#)), which uses observed control outcome data to impute missing data. Finally, the synthetic difference-in-difference method, which combines the two namesake techniques ([Arkhangelsky, Athey, Hirshberg, Imbens, & Wager, 2021](#)) is shown to be more efficient and unbiased relative to both individual methods. The applications of these three novelties form the basis for future research projects regarding applying the SCM to GDP data.

References

- Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2), 391–425. <http://doi.org/10.1257/jel.20191450>
- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california’s tobacco control program. *Journal of the American Statistical Association*, 105(490), 493–505. <http://doi.org/10.1198/jasa.2009.ap08746>
- Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2), 495–510. <http://doi.org/10.1111/ajps.12116>
- Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93(1), 113–132. <http://doi.org/10.1257/000282803321455188>
- Abadie, A., & L’Hour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, 116(536), 1817–1834. <http://doi.org/10.1080/01621459.2021.1971535>
- Acemoglu, D., Naidu, S., Restrepo, P., & Robinson, J. A. (2019). Democracy does cause growth. *Journal of Political Economy*, 127(1), 47–100. <http://doi.org/10.1086/700936>
- Amjad, M., Misra, V., Shah, D., & Shen, D. (2019). mRSC: Multi-dimensional robust synthetic control. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3, 1–27. <http://doi.org/10.1145/3341617.3326152>
- Amjad, M., Shah, D., & Shen, D. (2018). Robust synthetic control. *Journal of Machine Learning Research*, 19(22), 1–51. Retrieved from <http://jmlr.org/papers/v19/17-777.html>
- Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12), 4088–4118. <http://doi.org/10.1257/aer.20190159>

- Aruoba, S. B., Diebold, F. X., Nalewaik, J., Schorfheide, F., & Song, D. (2013). Improving u.s. GDP measurement: A forecast combination perspective. In X. Chen & N. R. Swanson (Eds.), *Recent advances and future directions in causality, prediction, and specification analysis: Essays in honor of halbert l. White jr* (pp. 1–25). New York, NY: Springer. http://doi.org/10.1007/978-1-4614-1653-1_1
- Aruoba, S. B., Diebold, F. X., Nalewaik, J., Schorfheide, F., & Song, D. (2016). Improving GDP measurement: A measurement-error perspective. *Journal of Econometrics*, 191(2), 384–397. <http://doi.org/10.1016/j.jeconom.2015.12.009>
- Athey, S., Bayati, M., Doudchenko, N., Imbens, G., & Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116(536), 1716–1730. <http://doi.org/10.1080/01621459.2021.1891924>
- Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2), 3–32. <http://doi.org/10.1257/jep.31.2.3>
- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4), 1229–1279. <http://doi.org/10.3982/ECTA6135>
- Becker, M., & Klößner, S. (2017). Estimating the economic costs of organized crime by synthetic control methods. *Journal of Applied Econometrics*, 32(7), 1367–1369. <http://doi.org/10.1002/jae.2572>
- Ben-Michael, E., Feller, A., & Rothstein, J. (2021). The augmented synthetic control method. *Journal of the American Statistical Association*, 116(536), 1789–1803. <http://doi.org/10.1080/01621459.2021.1929245>
- Bhattacharya, P. (2019, June 4). The great indian GDP debate, explained in five charts. Mint. Retrieved June 25, 2022, from <https://www.livemint.com/news/india/the-great-indian-gdp-debate-explained-in-five-charts-1559644946510.html>
- Billmeier, A., & Nannicini, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *The Review of Economics and Statistics*, 95(3), 983–1001. Retrieved from https://econpapers.repec.org/article/tprrstat/v_3a95_3ay_3a2013_3ai_3a3_3ap_3a983-1001.htm
- Callaway, B., & Sant’Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230. <http://doi.org/10.1016/j.jeconom.2020.12.001>
- Card, D. (1990). The impact of the mariel boatlift on the miami labor market. *ILR Review*, 43(2), 245–257. <http://doi.org/10.1177/001979399004300205>

- Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania. *American Economic Review*, 84(4), 772–793. Retrieved from <https://ideas.repec.org/a/aea/aecrev/v84y1994i4p772-93.html>
- Chang, A. C., & Li, P. (2018). Measurement error in macroeconomic data and economics research: Data revisions, gross domestic product, and gross domestic income. *Economic Inquiry*, 56(3), 1846–1869. <http://doi.org/10.1111/ecin.12567>
- Chernozhukov, V., Wuthrich, K., & Zhu, Y. (2021). *A \$t\$-test for synthetic controls* (No. arXiv:1812.10820). arXiv. Retrieved from <http://arxiv.org/abs/1812.10820>
- Chernozhukov, V., Wüthrich, K., & Zhu, Y. (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. *Journal of the American Statistical Association*, 116(536), 1849–1864. <http://doi.org/10.1080/01621459.2021.1920957>
- Coyle, D. (2014). *GDP: A brief but affectionate history*. Princeton University Press.
- CSO. (2015). No room for doubts on new GDP numbers. *Economic and Political Weekly*, 50(16), 7–8. Retrieved from <https://www.epw.in/journal/2015/16/discussion/no-room-doubts-new-gdp-numbers.html>
- Cunningham, S. (2021). *Causal inference; the mixtape* (1st ed.). Yale University Press. Retrieved from <https://yalebooks.yale.edu/9780300251685/causal-inference>
- Doudchenko, N., & Imbens, G. W. (2016). *Balancing, regression, difference-in-differences and synthetic control methods: A synthesis* (Working Paper No. 22791). National Bureau of Economic Research. Retrieved from <https://www.nber.org/papers/w22791>
- Dube, A., Lester, T. W., & Reich, M. (2010). Minimum wage effects across state borders: Estimates using contiguous counties. *The Review of Economics and Statistics*, 92(4), 945–964. Retrieved from <https://www.jstor.org/stable/40985804>
- Easterlin, R. A., McVey, L. A., Switek, M., Sawangfa, O., & Zweig, J. S. (2010). The happiness–income paradox revisited. *Proceedings of the National Academy of Sciences*, 107(52), 22463–22468. <http://doi.org/10.1073/pnas.1015962107>
- Ferman, B., & Pinto, C. (2021). Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, 12(4), 1197–1221. <http://doi.org/10.3982/QE1596>
- Ferman, B., Pinto, C., & Possebom, V. (2020). Cherry picking with synthetic controls. *Journal of Policy Analysis and Management*, 39(2), 510–532.

- <http://doi.org/10.1002/pam.22206>
- Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, 82(4), 540–554. <http://doi.org/10.1162/003465300559037>
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254–277. <http://doi.org/10.1016/j.jeconom.2021.03.014>
- Hahn, J., & Shi, R. (2017). Synthetic control and inference. *Econometrics*, 5(4), 1–12. Retrieved from https://econpapers.repec.org/article/gamjecnm/v_3a5_3ay_3a2017_3ai_3a4_3ap_3a52-_3ad_3a120610.htm
- Haldane, A., & Madouros, V. (2011, November 22). What is the contribution of the financial sector? VoxEU.org. Retrieved June 25, 2022, from <https://voxeu.org/article/what-contribution-financial-sector>
- Jacobs, J. P. A. M., Sarferaz, S., Sturm, J.-E., & Norden, S. van. (2022). Can GDP measurement be further improved? Data revision and reconciliation. *Journal of Business & Economic Statistics*, 40(1), 423–431. <http://doi.org/10.1080/07350015.2020.1831928>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning* (2nd ed.). NY: Springer New York. Retrieved from <https://link.springer.com/book/10.1007/978-1-0716-1418-1>
- Keynes, J. M. (2010). How to pay for the war. In J. M. Keynes (Ed.), *Essays in persuasion* (pp. 367–439). London: Palgrave Macmillan UK. http://doi.org/10.1007/978-1-349-59072-8_27
- Klößner, S., Kaul, A., Pfeifer, G., & Schieler, M. (2018). Comparative politics and the synthetic control method revisited: A note on abadie et al. (2015). *Swiss Journal of Economics and Statistics*, 154(1), 11. <http://doi.org/10.1186/s41937-017-0004-9>
- Kuosmanen, T., Zhou, X., Eskelinen, J., & Malo, P. (2021). *Design flaw of the synthetic control method* (No. 106328). University Library of Munich, Germany. Retrieved from <https://ideas.repec.org/p/pramprapa/106328.html>
- Kuznets, S. (1934). *National income, 1929-1932*. NBER. Retrieved from <https://www.nber.org/books-and-chapters/national-income-1929-1932>
- Mankiw, N. G., Runkle, D. E., & Shapiro, M. D. (1984). Are preliminary announcements of the money stock rational forecasts? *Journal of Monetary Economics*, 14(1), 15–27. [http://doi.org/10.1016/0304-3932\(84\)90024-2](http://doi.org/10.1016/0304-3932(84)90024-2)
- Mankiw, N. G., & Shapiro, M. (1986). *News or noise? An analysis of GNP revisions* ({NBER} Working Paper No. 1939). Na-

- tional Bureau of Economic Research, Inc. Retrieved from <https://econpapers.repec.org/paper/nbrnberwo/1939.htm>
- Mayberry, A. A. (2022). The economic cost of a nuclear weapon: A synthetic control approach. *Defence and Peace Economics*, 0(0), 1–20. <http://doi.org/10.1080/10242694.2022.2065186>
- Milanovic, B. (2013). Global income inequality in numbers: In history and now. *Global Policy*, 4(2), 198–208. <http://doi.org/10.1111/1758-5899.12032>
- Nagaraj, R. (2015a). Seeds of doubt on new GDP numbers: Private corporate sector overestimated?, 50, 14–17.
- Nagaraj, R. (2015b). Seeds of doubt remain: A reply to CSO’s rejoinder. *Economic and Political Weekly*, 50(18), 64–66. Retrieved from <https://www.jstor.org/stable/24481913>
- Nagaraj, R. (2021). Revisiting the GDP estimation debate. *Economic and Political Weekly*, 56(45), 10–13. Retrieved from <https://www.epw.in/journal/2021/44/commentary/revisiting-gdp-estimation-debate.html>
- Nagaraj, R., & Srinivasan, T. (2016). Measuring india’s GDP growth: Unpacking the analytics & data issues behind a controversy that refuses to go away (pp. 1–40). Presented at the India policy forum, New Delhi, India: National Council of Applied Economic Research.
- Nair, R. (2019, June 13). Base year, revised growth, overestimation & all that you need to know about calculating GDP. ThePrint. Retrieved June 25, 2022, from <https://theprint.in/theprint-essential/base-year-revised-growth-overestimation-all-that-you-need-to-know-about-calculating-gdp/249391/>
- Oulton, N. (1998). The implications of the boskin report. *National Institute Economic Review*, (165), 89–98. Retrieved from <https://www.jstor.org/stable/23872743>
- Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94(5), 1002–1037. Retrieved from <https://www.jstor.org/stable/1833190>
- Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5), S71–S102. Retrieved from <https://www.jstor.org/stable/2937632>
- Sapre, A., & Sengupta, R. (2017). Analysis of revisions in indian GDP data. *World Economics*, 18(4), 129–172. Retrieved from <https://econpapers.repec.org/article/wejwldecn/689.htm>

- Singhal, S., & Nilakantan, R. (2016). The economic effects of a counterinsurgency policy in india: A synthetic control analysis. *European Journal of Political Economy*, 45, 1–17. <http://doi.org/10.1016/j.ejpoleco.2016.08.012>
- Smith, A. (2000). *The wealth of nations*. Random House Publishing Group.
- Spearman, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology*, 15(2), 201–292. <http://doi.org/10.2307/1412107>
- Stock, J. H., & Watson, M. (2016). *Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics* (Handbook of Macroeconomics) (pp. 415–525). Elsevier. Retrieved from <https://econpapers.repec.org/bookchap/eeemacchp/v2-415.htm>
- Subramanian, A. (2019, June). *India's GDP mis-estimation: Likelihood, magnitudes, mechanisms, and implications*. {CID} Working Paper 354, CID Working Paper 354.
- Tretyakova, A., & Birman, I. (1976). Input-output analysis in the USSR. *Soviet Studies*, 28(2), 157–186. Retrieved from <https://www.jstor.org/stable/150826>
- Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1), 57–76. <http://doi.org/10.1017/pan.2016.2>