Measurement Issues in India's GDP

A Synthetic Control Approach

Advait Moharir

14-07-2022

Advisor 1: Dr. Federico Crudu

Advisor 2: Dr. Giuliano Curatola

University of Siena

The Indian GDP debate

The Indian GDP debate

India's new real GDP series, which uses base year 2011 instead of 2004, has been subject to the following critiques:

The Indian GDP debate

India's new real GDP series, which uses base year 2011 instead of 2004, has been subject to the following critiques:

- Change in methodology of measuring mfg sector output has led to "blowing up" of private corporate sector GVA (Nagaraj, 2015a; Nagaraj, 2015b).

The Indian GDP debate

India's new real GDP series, which uses base year 2011 instead of 2004, has been subject to the following critiques:

- Change in methodology of measuring mfg sector output has led to "blowing up" of private corporate sector GVA (Nagaraj, 2015a; Nagaraj, 2015b).
- Reversal of positive relationship with correlates like imports and credit growth (Subramanian, 2019).

The Indian GDP debate

India's new real GDP series, which uses base year 2011 instead of 2004, has been subject to the following critiques:

- Change in methodology of measuring mfg sector output has led to "blowing up" of private corporate sector GVA (Nagaraj, 2015a; Nagaraj, 2015b).
- Reversal of positive relationship with correlates like imports and credit growth (Subramanian, 2019).
- Other issues: upward biased revisions (Sapre and Sengupta, 2017), choice of deflators, and under-representation of unorganized sector (Nagaraj and Srinivasan, 2016).

The Indian GDP debate

India's new real GDP series, which uses base year 2011 instead of 2004, has been subject to the following critiques:

- Change in methodology of measuring mfg sector output has led to "blowing up" of private corporate sector GVA (Nagaraj, 2015a; Nagaraj, 2015b).
- Reversal of positive relationship with correlates like imports and credit growth (Subramanian, 2019).
- Other issues: upward biased revisions (Sapre and Sengupta, 2017), choice of deflators, and under-representation of unorganized sector (Nagaraj and Srinivasan, 2016).

Measurement error typically estimated by reconciling income - side and expenditure - side GDP estimates (Aruoba, Diebold, Nalewaik, et al., 2016; Chang and Li, 2018; Jacobs, Sarferaz, Sturm, et al., 2022).

The Indian GDP debate

India's new real GDP series, which uses base year 2011 instead of 2004, has been subject to the following critiques:

- Change in methodology of measuring mfg sector output has led to "blowing up" of private corporate sector GVA (Nagaraj, 2015a; Nagaraj, 2015b).
- Reversal of positive relationship with correlates like imports and credit growth (Subramanian, 2019).
- Other issues: upward biased revisions (Sapre and Sengupta, 2017), choice of deflators, and under-representation of unorganized sector (Nagaraj and Srinivasan, 2016).

Measurement error typically estimated by reconciling income - side and expenditure - side GDP estimates (Aruoba, Diebold, Nalewaik, et al., 2016; Chang and Li, 2018; Jacobs, Sarferaz, Sturm, et al., 2022).

However, no income-side data available for India.

Research Design

I construct a counterfactual trajectory of real GDP.

Research Design

I construct a counterfactual trajectory of real GDP.

The trajectory represents what India's GDP would have been had the methodology for estimation not changed post - 2011.

Research Design

I construct a counterfactual trajectory of real GDP.

The trajectory represents what India's GDP would have been had the methodology for estimation not changed post - 2011.

This is generated using an optimally chosen weighted average of countries similar to India, which did **not** have any such intervention in 2011.

Research Design

I construct a counterfactual trajectory of real GDP.

The trajectory represents what India's GDP would have been had the methodology for estimation not changed post - 2011.

This is generated using an optimally chosen weighted average of countries similar to India, which did **not** have any such intervention in 2011.

To do this, I use two methods:

Research Design

I construct a counterfactual trajectory of real GDP.

The trajectory represents what India's GDP would have been had the methodology for estimation not changed post - 2011.

This is generated using an optimally chosen weighted average of countries similar to India, which did **not** have any such intervention in 2011.

To do this, I use two methods:

- Synthetic Control Method (**SCM**) (Abadie and Gardeazabal, 2003; Abadie, Diamond, and Hainmueller, 2010)

Research Design

I construct a counterfactual trajectory of real GDP.

The trajectory represents what India's GDP would have been had the methodology for estimation not changed post - 2011.

This is generated using an optimally chosen weighted average of countries similar to India, which did **not** have any such intervention in 2011.

To do this, I use two methods:

- Synthetic Control Method (**SCM**) (Abadie and Gardeazabal, 2003; Abadie, Diamond, and Hainmueller, 2010)
- Generalized Synthetic Control Method (**GSCM**) (Xu, 2017).

Methodology

SCM

Let data be observed $J + 1$ countries, where $i = 1$ is India (treated unit). The rest of the countries are unaffected by the treatment and are referred to as the *donor* pool.

SCM

Let data be observed $J+1$ countries, where $i=1$ is India (treated unit). The rest of the countries are unaffected by the treatment and are referred to as the *donor* pool.

Let the the potential GDP for India with and without the intervention be given by Y_{1t}^{I} and Y_{1t}^{N} respectively. The former is observed, and the latter is not.

SCM

Let data be observed $J+1$ countries, where $i=1$ is India (treated unit). The rest of the countries are unaffected by the treatment and are referred to as the *donor* pool.

Let the the potential GDP for India with and without the intervention be given by Y_{1t}^I and Y_{1t}^N respectively. The former is observed, and the latter is not. The counterfactual GDP is estimated as follows:

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

where $\mathbf{W*} = (w_2^*, \ldots, w_{J+1}^*)$ is a vector of *optimal* weights. The weights are chosen such that synthetic India best represents the *predictors* of India's GDP before the treatment. Hence, the optimal weights minimize

SCM

Let data be observed $J+1$ countries, where $i=1$ is India (treated unit). The rest of the countries are unaffected by the treatment and are referred to as the *donor* pool.

Let the the potential GDP for India with and without the intervention be given by Y_{1t}^{I} and Y_{1t}^{N} respectively. The former is observed, and the latter is not. The counterfactual GDP is estimated as follows:

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

where $\mathbf{W*} = (w_2^*, \ldots, w_{J+1}^*)$ is a vector of *optimal* weights. The weights are chosen such that synthetic India best represents the *predictors* of India's GDP before the treatment. Hence, the optimal weights minimize

$$||\mathbf{X_1} - \mathbf{X_0}\mathbf{W}|| = \sqrt{(\mathbf{X_1} - \mathbf{X_0}\mathbf{W})'\mathbf{V}(\mathbf{X_1} - \mathbf{X_0}\mathbf{W})}$$

where $\mathbf{X_1}$ consists of the pre-intervention predictors of India's GDP and $\mathbf{X_0}$ consists of pre-intervention predictors of the donor countries' GDP.

GSCM

Combines synthetic control with interactive fixed effects (IFE) model (Bai, 2009).

GSCM

Combines synthetic control with interactive fixed effects (IFE) model (Bai, 2009).

The counterfactual is estimated using a linear IFE model as follows:

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

where x_{1t} is a vector of covariates, $\hat{\beta}$ consists of estimated parameters for covariates, $\hat{\lambda}_1$ are estimated common factors, and $\hat{f}_t$ are estimated factor loadings. The optimal number of factors are chosen via cross-validation.

GSCM

Combines synthetic control with interactive fixed effects (IFE) model (Bai, 2009).

The counterfactual is estimated using a linear IFE model as follows:

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

where x_{1t} is a vector of covariates, $\hat{\beta}$ consists of estimated parameters for covariates, $\hat{\lambda}_1$ are estimated common factors, and $\hat{f}_t$ are estimated factor loadings. The optimal number of factors are chosen via cross-validation.

Using both methods, we are interested in estimating the average treatment effect for India, given by

$$\hat{\alpha}_{1t} = Y_{1t}^{I} - \hat{Y}_{1t}^{N}$$

Setup

Treated Unit : India

Setup

Treated Unit : India

Treatment year : 2011

Setup

Treated Unit : India

Treatment year : 2011

Time Period : 1991-2019

Setup

Treated Unit : India

Treatment year : 2011

Time Period : 1991-2019

Donor Units : 19 Emerging Market Economies (EMEs) as defined by the IMF in their 2021 release

Setup

Treated Unit : India

Treatment year : 2011

Time Period : 1991-2019

Donor Units : 19 Emerging Market Economies (EMEs) as defined by the IMF in their 2021 release

Specifications

- SCM using predictors (covariates) of GDP
- GSCM without covariates (matching only on GDP)
- GSCM with economic covariates only
- GSCM with economic and education covariates

Data

Data Sources

Dependent variable: log real GDP in 2011 Local Currency Units

Data Sources

Dependent variable: log real GDP in 2011 Local Currency Units

Covariates

Data Sources

Dependent variable: log real GDP in 2011 Local Currency Units

Covariates

- Trade (Exports+ Imports) as a share of GDP

Data Sources

Dependent variable: log real GDP in 2011 Local Currency Units

Covariates

- Trade (Exports+ Imports) as a share of GDP
- Gross Investment as a share of GDP

Data Sources

Dependent variable: log real GDP in 2011 Local Currency Units

Covariates

- Trade (Exports+ Imports) as a share of GDP
- Gross Investment as a share of GDP
- Gross Enrollment (Primary)

Data Sources

Dependent variable: log real GDP in 2011 Local Currency Units

Covariates

- Trade (Exports+ Imports) as a share of GDP
- Gross Investment as a share of GDP
- Gross Enrollment (Primary)
- Gross Enrollment (Secondary)

Data Sources

Dependent variable: log real GDP in 2011 Local Currency Units

Covariates

- Trade (Exports+ Imports) as a share of GDP
- Gross Investment as a share of GDP
- Gross Enrollment (Primary)
- Gross Enrollment (Secondary)

All variables are taken from the World Development Indicators (WDI) database maintained by the World Bank using the package `wdi`.

Data Sources

Dependent variable: log real GDP in 2011 Local Currency Units

Covariates

- Trade (Exports+ Imports) as a share of GDP
- Gross Investment as a share of GDP
- Gross Enrollment (Primary)
- Gross Enrollment (Secondary)

All variables are taken from the World Development Indicators (WDI) database maintained by the World Bank using the package `wdi`.

The standard SC model is estimated using the `tidysynth` package and the GSC model is estimated using the `gsynth` package.

Results

SCM

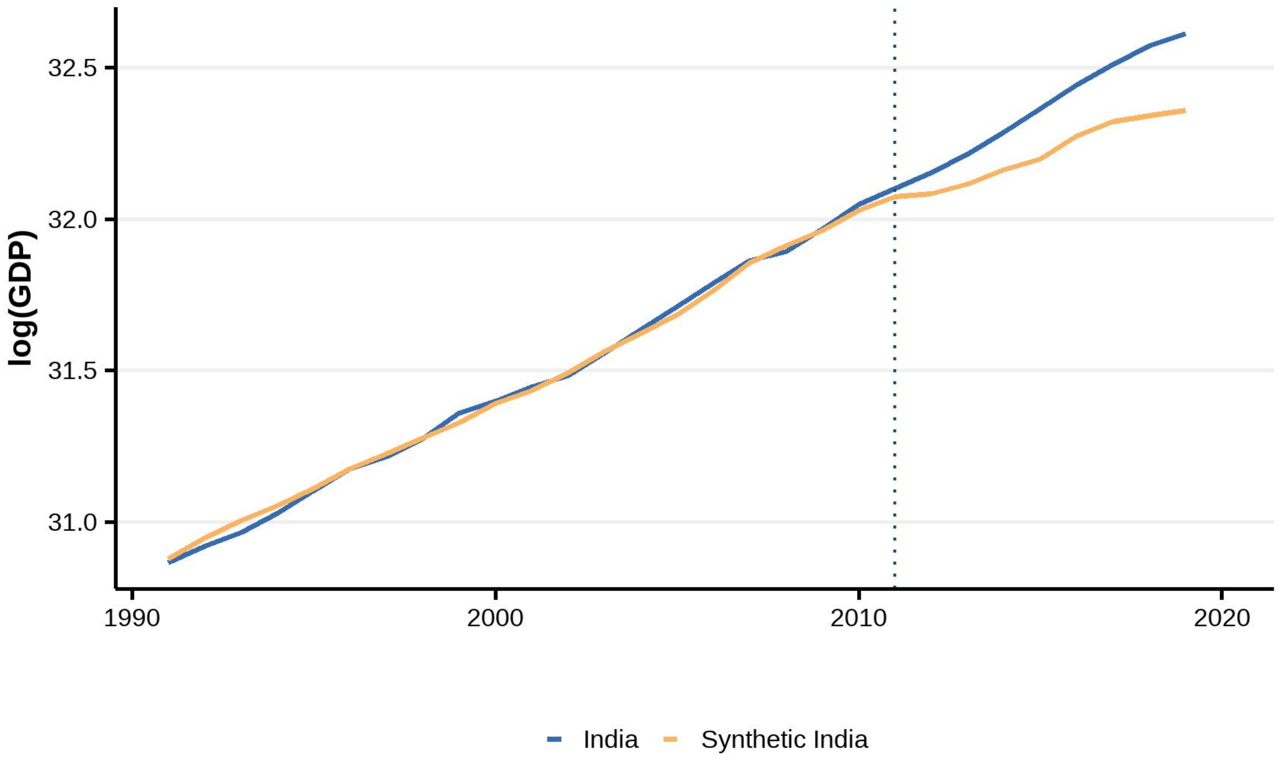

Figure - 1: Log GDP trend for actual and synthetic India (1991 - 2019).

GSCM (Baseline Model)

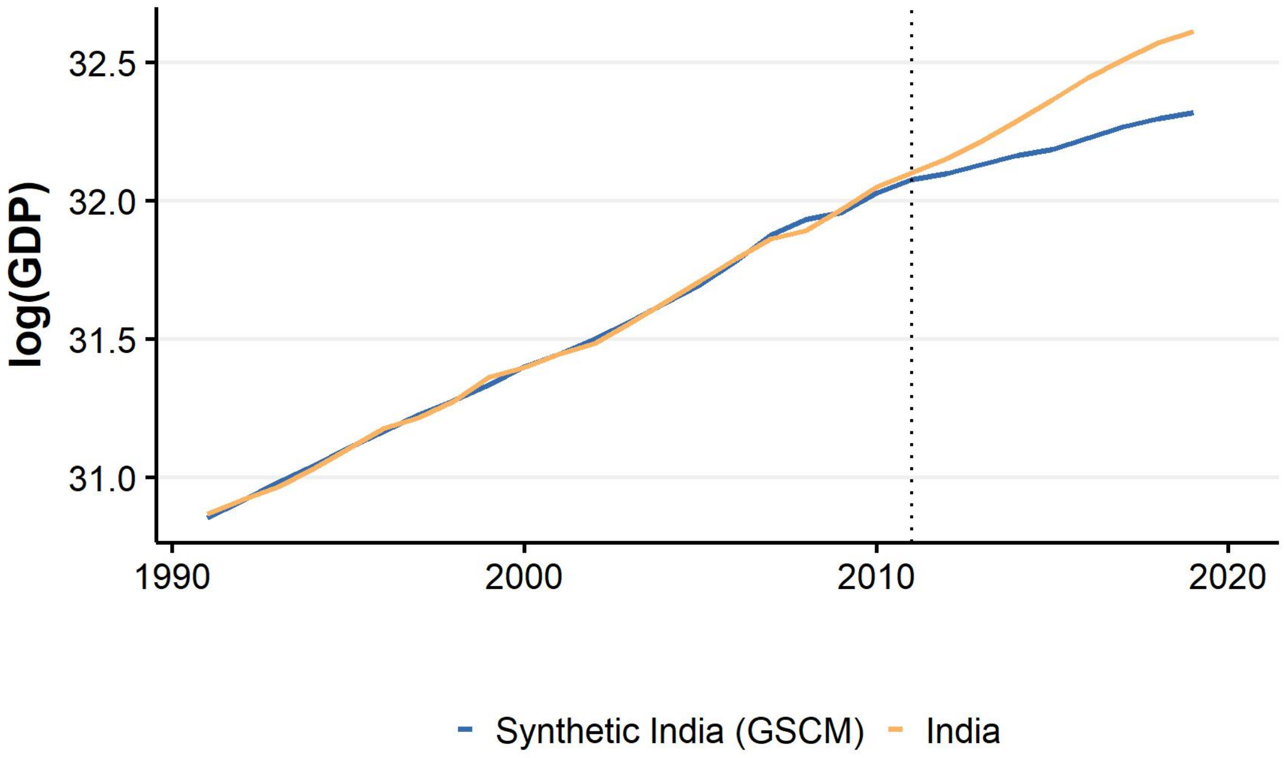

Figure - 2: Log GDP trend for actual and synthetic India (1991 - 2019).

GSCM (with economic covariates)

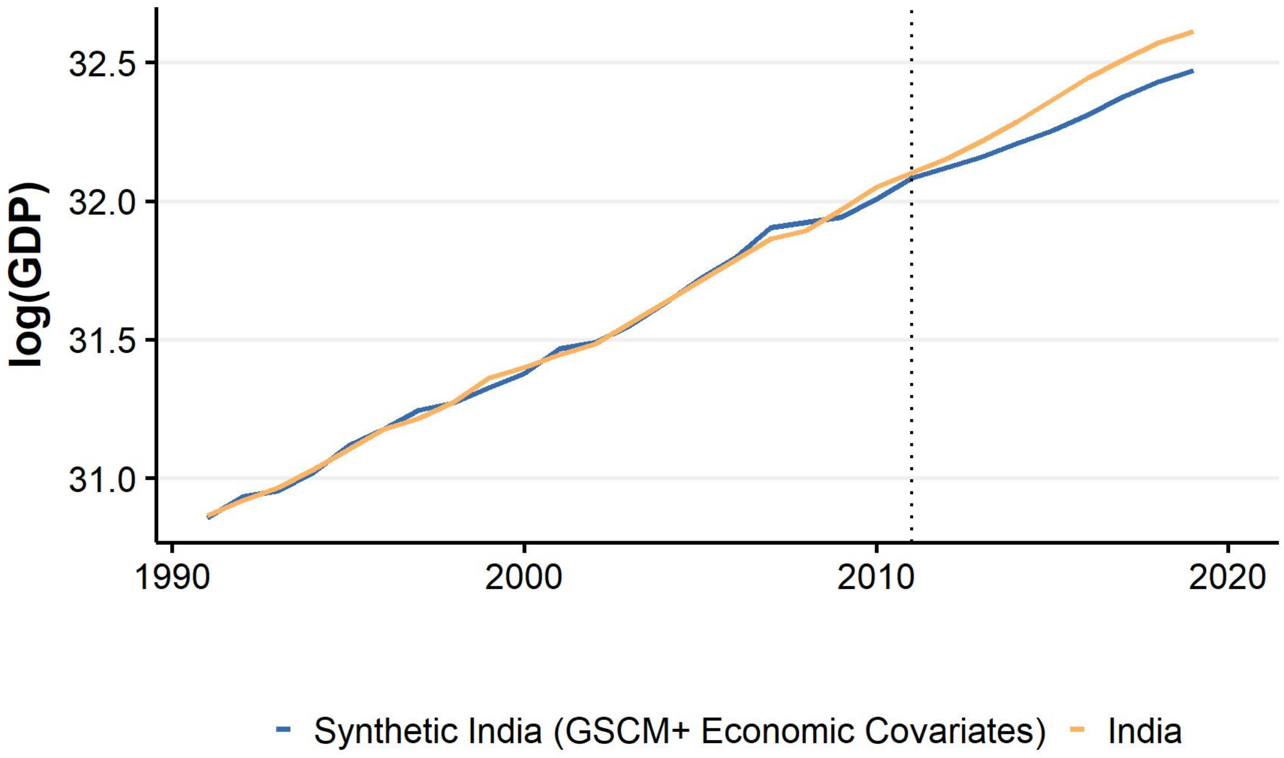

Figure - 3: Log GDP trend for actual and synthetic India (1991 - 2019).

GSCM (with all covariates)

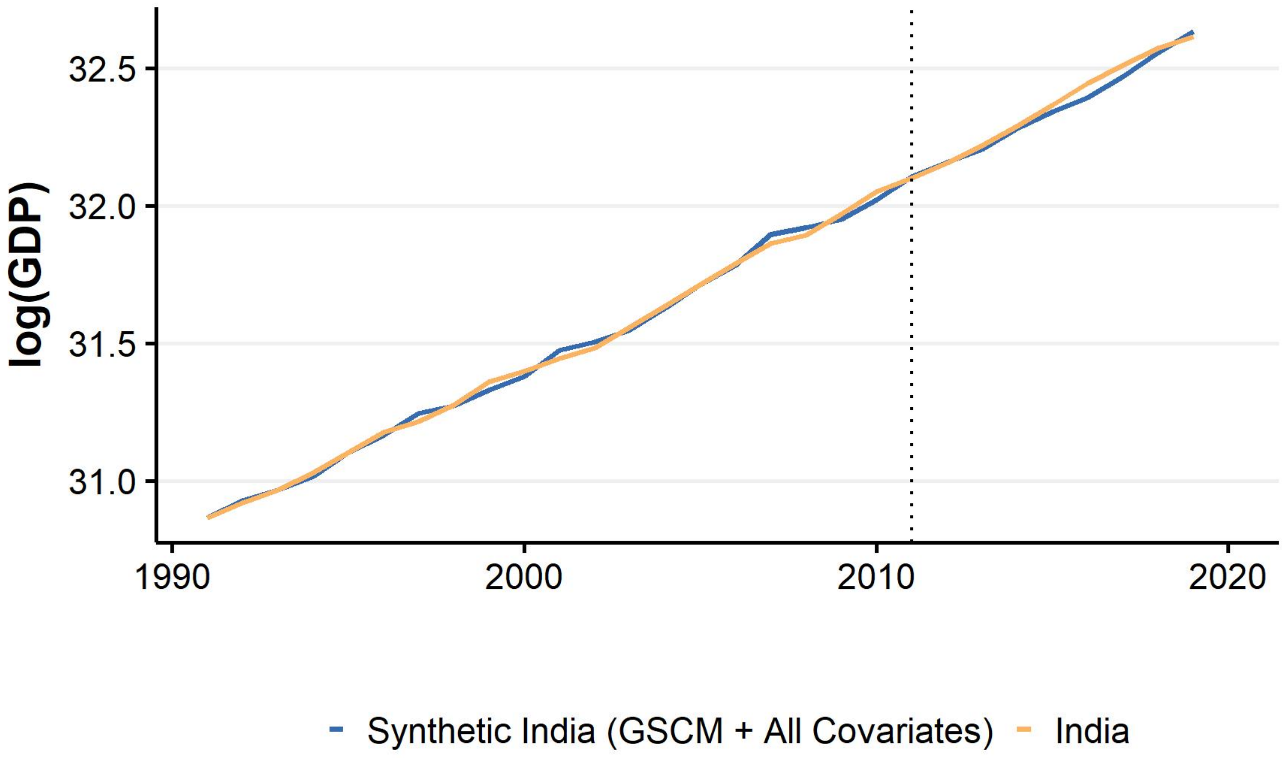

Figure - 4: Log GDP trend for actual and synthetic India (1991 - 2019).

Difference in growth rates

Year	SCM	GSCM	GSCM+EC	GSCM+AC
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 - 1: Difference between actual and synthetic GDP growth rates for various methods (in percentage).

Note: **EC** stands for "Economic Covariates", while **AC** stands for "All Covariates".

Inference

SCM: MSPE ratios

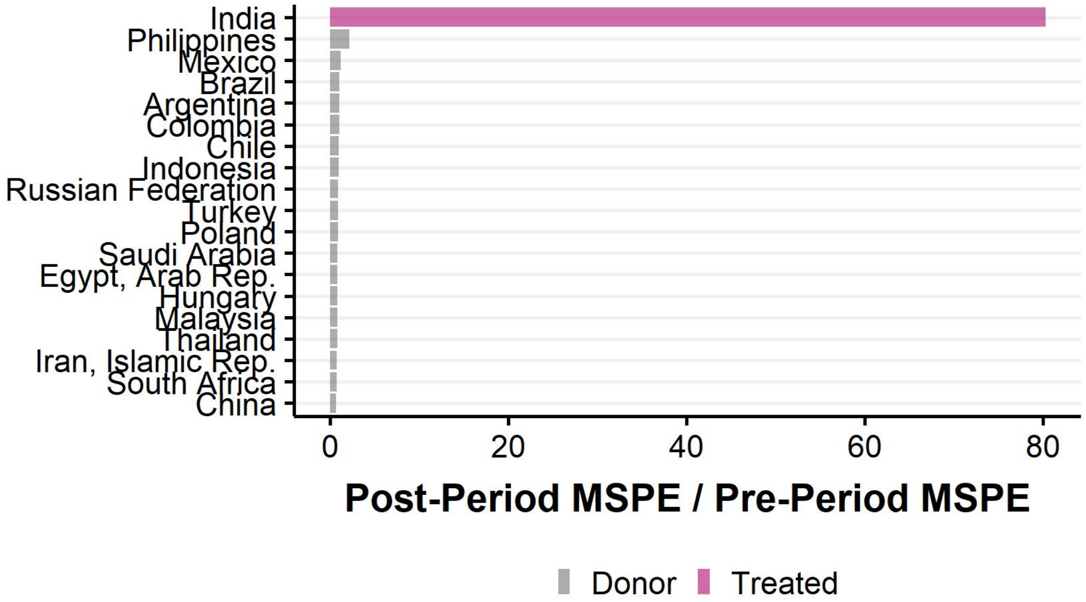

Figure - 5: Ratios of Post to Pre Intervention MSPE for India and donor countries.

SCM: MSPE ratios

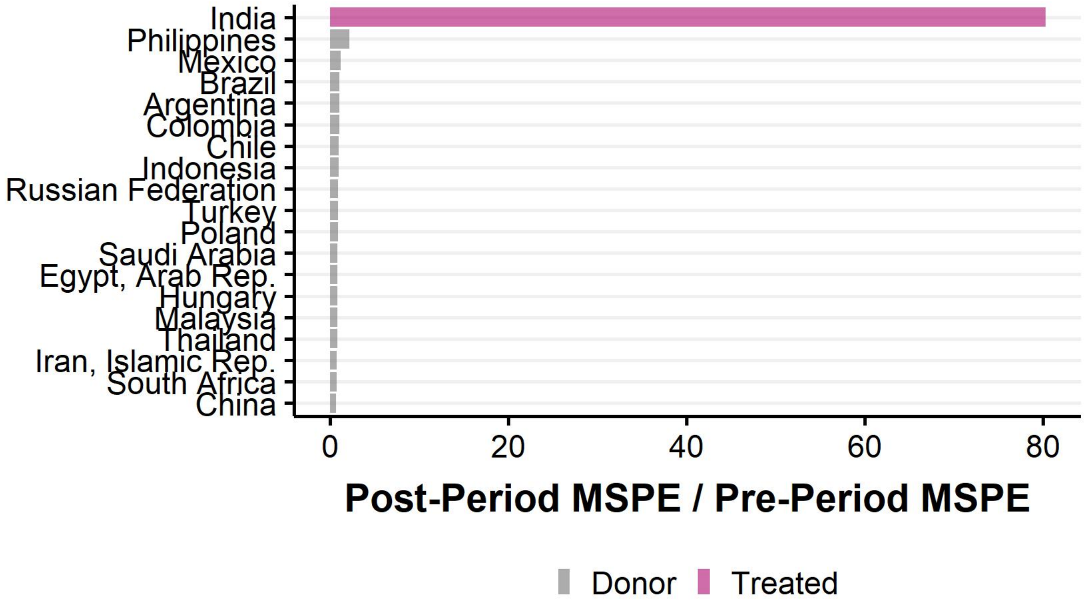

Figure - 5: Ratios of Post to Pre Intervention MSPE for India and donor countries.

Interpretation: The ratio of MSPE's is a measure of how close the two trends are before and after the intervention. If the intervention affects the unit of interest, then the ratio is the largest for that unit, as seen here.

SCM: Fisher's exact p-values

Country	Type	Fisher's exact 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 - 2: p-values for India and donor countries.

GSCM: Inference based on parametric bootstrap

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: Average ATT and p-values for GSCM models.

GSCM: Inference based on parametric bootstrap

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: Average ATT and p-values for GSCM models.

None of the ATT values are coefficient are statistically significant, indicating that there may be no over - estimation of GDP.

GSCM: Inference based on parametric bootstrap

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: Average ATT and p-values for GSCM models.

None of the ATT values are coefficient are statistically significant, indicating that there may be no over - estimation of GDP.

Possible Reasons

- *Small number of donor countries*: Xu (2017) notes that the GSCM is prone to bias if there are less than 40 donor units.

GSCM: Inference based on parametric bootstrap

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: Average ATT and p-values for GSCM models.

None of the ATT values are coefficient are statistically significant, indicating that there may be no over - estimation of GDP.

Possible Reasons

- *Small number of donor countries*: Xu (2017) notes that the GSCM is prone to bias if there are less than 40 donor units.
- *Sparse control data* : Large number of missing values for co variates lead to significantly reduced sample size.

Robustness Checks

Backdating (SCM)

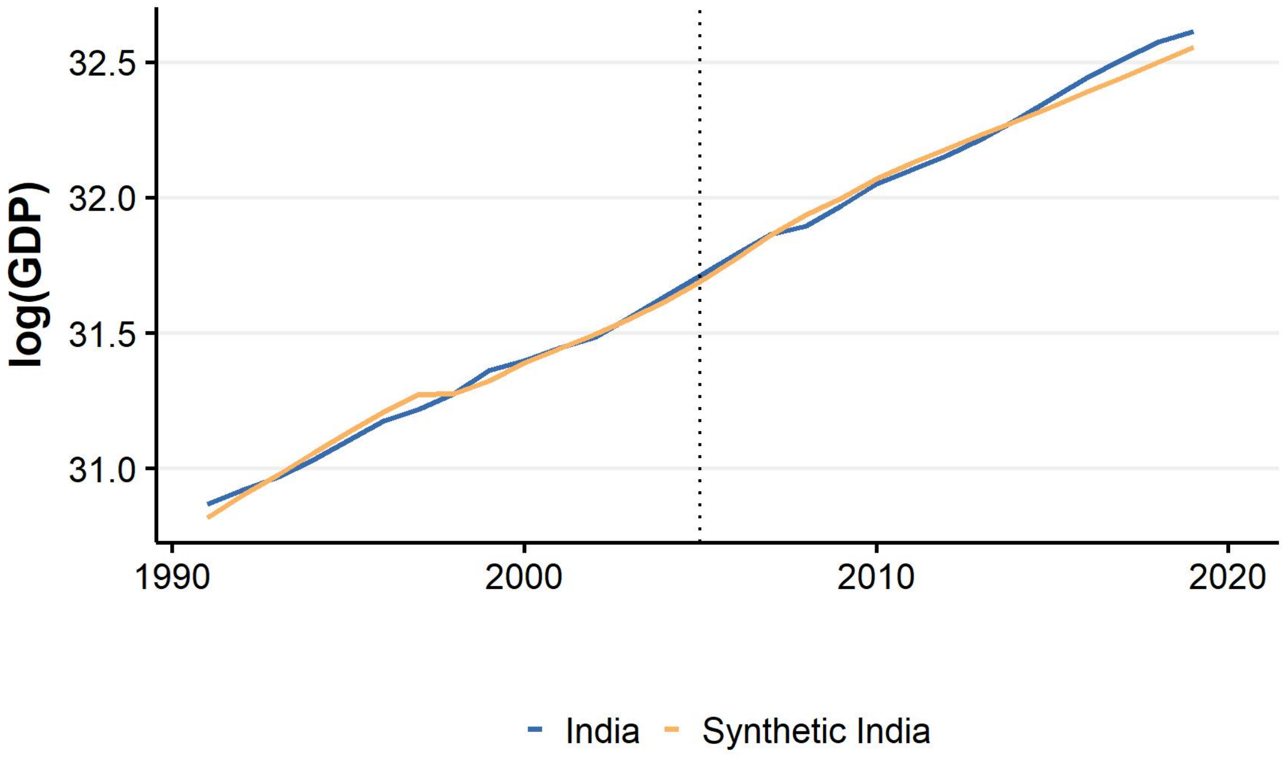

Figure - 6: Log GDP trend for actual and synthetic India with backdated treatment (1991 - 2019).

Backdating (GSCM)

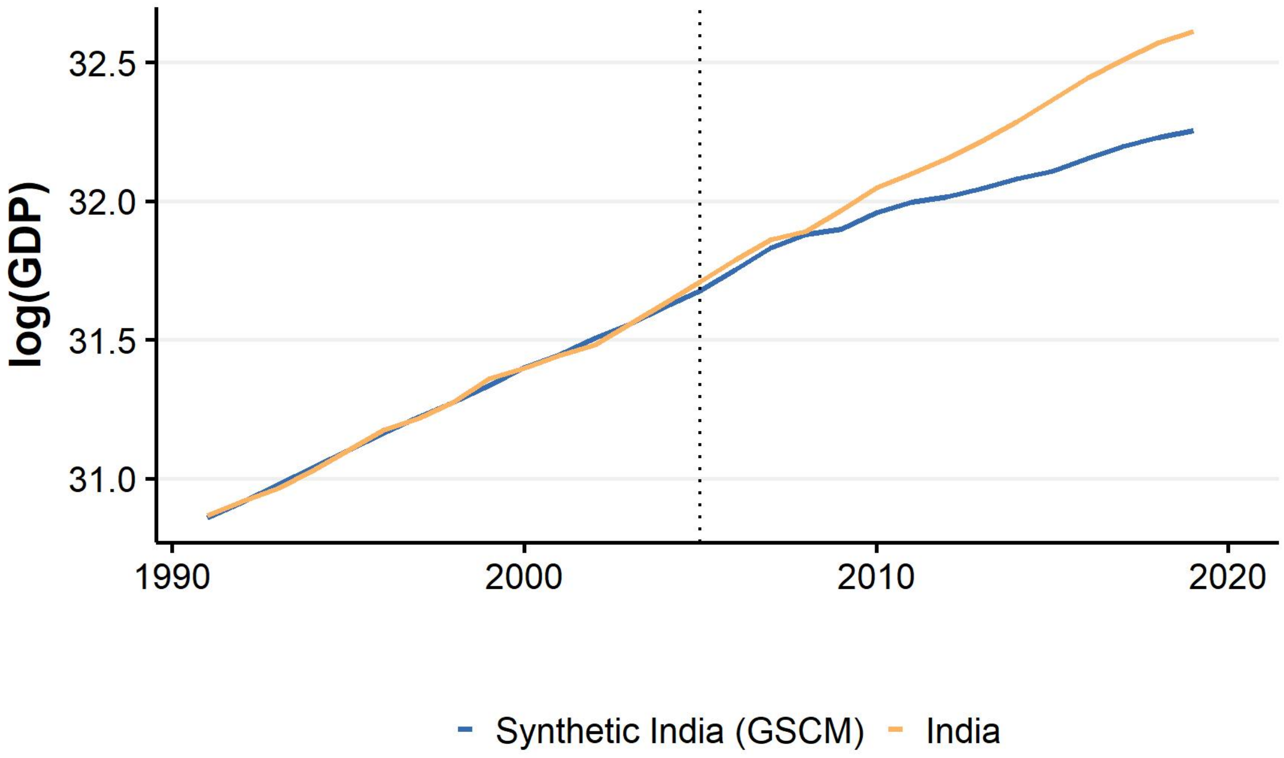

Figure - 7: Log GDP trend for actual and synthetic India with backdated treatment (1991 - 2019).

GSCM with quarterly data

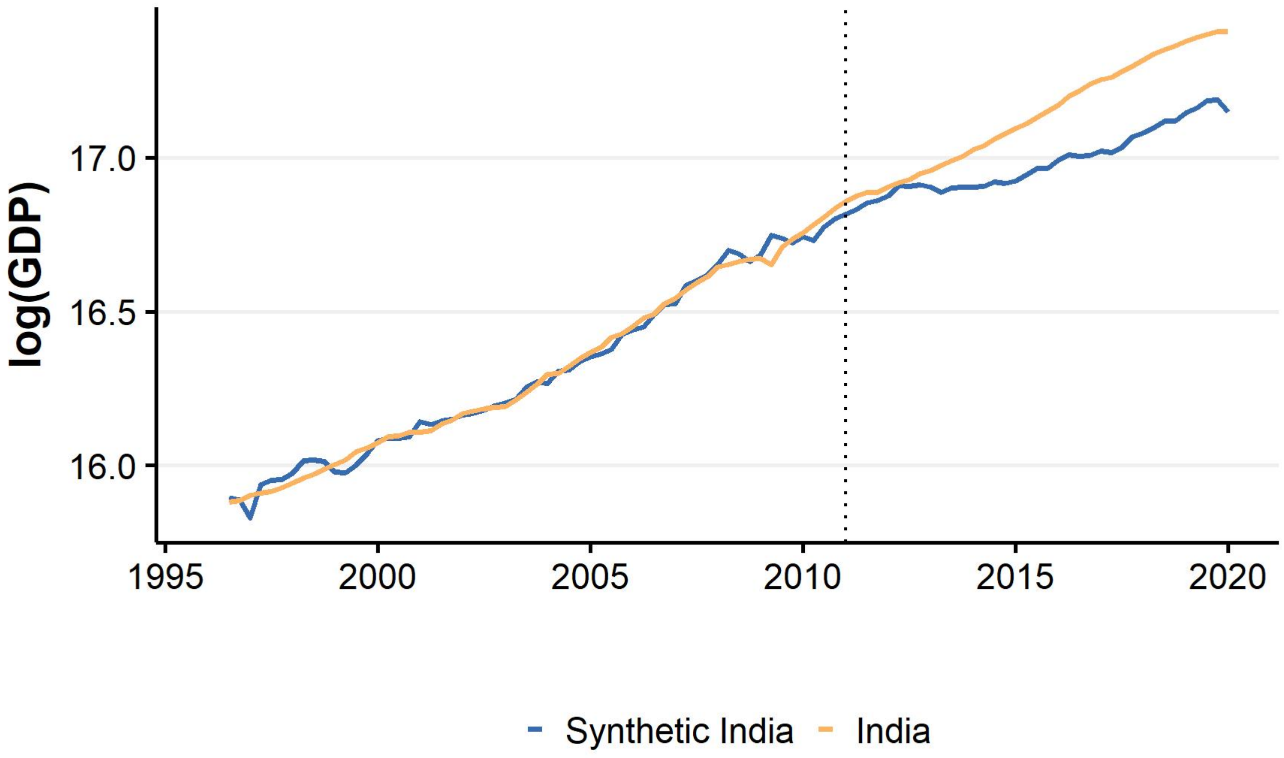

Figure - 8: Log GDP trend for actual and synthetic India with quarterly data (1996 - 2019).

Conclusion

- I construct counterfactual GDP estimates using SCM and GSCM method and find average over - estimation between **-0.5%** and **3%** p.a., depending on the identification strategy.

Conclusion

- I construct counterfactual GDP estimates using SCM and GSCM method and find average over - estimation between **-0.5%** and **3%** p.a., depending on the identification strategy.
- The SCM results are significant at 10% but not robust (sensitive to specification). GSCM results are robust, but statistically insignificant.

Conclusion

- I construct counterfactual GDP estimates using SCM and GSCM method and find average over - estimation between **-0.5%** and **3%** p.a., depending on the identification strategy.
- The SCM results are significant at 10% but not robust (sensitive to specification). GSCM results are robust, but statistically insignificant.
- Hence, using this method we cannot conclusively establish that India's GDP is over-estimated. However, inference and test statistics for SCM is a work in progress (Ben-Michael, Feller, and Rothstein, 2021; Chernozhukov, Wuthrich, and Zhu, 2021a).

Conclusion

- I construct counterfactual GDP estimates using SCM and GSCM method and find average over - estimation between **-0.5%** and **3%** p.a., depending on the identification strategy.
- The SCM results are significant at 10% but not robust (sensitive to specification). GSCM results are robust, but statistically insignificant.
- Hence, using this method we cannot conclusively establish that India's GDP is over-estimated. However, inference and test statistics for SCM is a work in progress (Ben-Michael, Feller, and Rothstein, 2021; Chernozhukov, Wuthrich, and Zhu, 2021a).
- **Future Avenues**: Using advances in causal panel data modelling, like matrix completion (Athey, Bayati, Doudchenko, et al., 2021) to deal with problem of sparse covariate data, and other estimators like synthetic difference-in-differences (Arkhangelsky, Athey, Hirshberg, et al., 2021).