



HAL
open science

Gender Budgeting and Health Spending Efficiency in Indian States: A Staggered Difference-in-Differences Analysis

Cheick Camara

► **To cite this version:**

Cheick Camara. Gender Budgeting and Health Spending Efficiency in Indian States: A Staggered Difference-in-Differences Analysis. 2023. hal-04294262v1

HAL Id: hal-04294262

<https://uca.hal.science/hal-04294262v1>

Preprint submitted on 19 Nov 2023 (v1), last revised 12 Dec 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Gender Budgeting and Health Spending Efficiency in Indian States: A Staggered Difference-in-Differences Analysis

By Cheick CAMARA¹

¹ *Université Clermont Auvergne, CNRS, IRD, CERDI, F-63000 Clermont-Ferrand, France*

Abstract

Gender-responsive budgeting is an approach that integrates gender considerations into all stages of the budgetary process, aiming to address gender disparities and promote gender equality. This study investigates the impact of gender-responsive budgeting on fiscal efficiency in the context of Indian states. Leveraging a Staggered Difference-in-Differences (DiD) framework as the primary empirical methodology, with Entropy Balancing employed as a robustness test, we assess the consequences of gender-responsive budget adoption by Indian states. Our analysis reveals a persistent positive effect from this adoption, extending across periods. One key transmission channel emerges as responsible for this sustained positive effect. Adopting gender-responsive budgeting has led to a substantial enhancement in the quality of revenue forecasts. The improved accuracy of revenue projections equips states to make more precise and timely resource allocation decisions, further boosting fiscal efficiency. In summary, this study underscores the enduring positive influence of gender-responsive budgeting, which incorporates gender considerations into fiscal policies, on fiscal efficiency in Indian states. The increased transparency and improved revenue forecasting resulting from gender-responsive budgeting signify its crucial role in fostering equitable resource allocation, ultimately enhancing fiscal efficiency and overall governance.

JEL classifications: H1; H7; J16; R5

Keywords: Regional Government, Regional Policy, Public Economics, Public Finance

List of Figures

1	Gender Budgeting framework (Source: IMF PFM Blog)	3
2	GB adoption through the time	4
3	Stylized facts	11
4	Parrallel trend	15
5	Parallel trend with honest DiD (Rambachan and Roth (2023))	16
6	Event study	18
7	Wooldridge Results	19
8	Alternative Diff in Diff Results	22
9	Propensity score common support	40
10	Alternative Diff in Diff Results for Battese et Coelli, 1995	45

List of Tables

1	Repartition of treatment	8
2	Summary statistics	9
3	Diff in Diff results	17
4	Diff in Diff results with Wooldridge, 2021	19
5	Descriptive statistics	26
6	Covariate balancing	27
7	Results of entropy balancing	27
8	Results of entropy balancing with additional covariates	28
9	Results of entropy balancing	29
10	Statistics	30
11	Transmission channel	31
12	Pairwise correlation	31
13	Simultaneous equations	32
14	Honest DiD Rambachan and Roth (2023)	41
15	Entropy balancing full results	42

16	Entropy balancing full results	43
17	Diff in Diff results for Battese and Coelli, 1995	44
18	Entropy balancing results for Battese and Coelli.	46
19	Entropy balancing results for Battese and Coelli.	47
20	Simultaneous equations	48

Contents

1) Introduction	1
2) Methodology	5
2.1) Data	5
2.1).1 Efficiency score: the outcome variable	5
2.1).2 The treatment and control variables	8
2.1).3 Stylized facts	9
2.2) Identification strategy	11
2.3) Parallel trend assumption	14
3) Results	16
4) Robustness check: alternative estimation method	18
4.1) Alternative Diff in Diff methods	18
4.2) Entropy banalncing	22
4.2).1 Methodological concept	22
4.2).2 Correlation issue	25
4.2).3 Results	27
5) Transmission channel	29
6) Conclusion and policy implications	32
References	34
7) Appendix	40
7.1) Results for Battese and Coelli, 1995	44

1) Introduction

Gender budgeting initiatives seek to integrate gender-related goals into fiscal policies and administration to achieve gender equality and promote women's empowerment. Many countries and subnational governments have adopted gender budgeting initiatives, including many in developing countries. The most notable initiatives include those in India and the Philippines, among the developing countries, and Australia, Sweden, and the Republic of Korea, among the developed countries. Gender budgeting is not a special approach to budgeting or even an add-on to budgeting. Rather, gender budgeting is an approach to budgeting that can improve women's empowerment when fiscal policies and administrative procedures are structured to address gender inequality and women's development needs. Gender budgeting consists of explicitly taking the goal of gender equality into account in the fiscal process. It implies having a gender-based perspective during the different phases of this process and analysing the direct and indirect ramifications of public expenditure and revenue on the respective situation of men and women. It can be rolled out in different manners, either by individually examining, measure by measure, the consequences by gender of fiscal policies, or by assessing the global impact by gender of all the measures taken for a given year ¹.

Many countries consider the Sustainable Development Goals (SDG) and especially objective (5)² as an objective for their gender budgeting implementation. This goal has for objective of reducing gender inequality, improving women's economic empowerment, and access to the labour market to generate inclusive growth for all. Indeed, gender inequalities and gaps persist not only in labour force participation but also in education, unpaid care work, access to credit, technology or income, among others. gender inequality also impacts health through differential exposures, health-related behaviours, and access to care, as well as how gender-biased health research and healthcare systems reinforce and reproduce gender inequalities, with serious health implications. The cumulative consequences of structured disadvantage, mediated through discriminatory laws, policies, and institutions, as well as diet, stress, substance use, and environmental toxins, have triggered important discussions about the role of social injustice in the creation and maintenance

¹<https://www.tresor.economie.gouv.fr/Articles/f2d0994d-87f5-4c74-a1f8-0b806a4e80f2/files/7ed39a11-894a-4a84-be95-69ccf1591ac6>

²End all forms of discrimination against all women and girls everywhere; and ensure universal access to sexual and reproductive health and reproductive rights

of health inequities (Heise et al. (2019)) Gender budgeting can be a valuable tool to help tackle these challenges because it allows governments to better allocate and implement resources more inclusively. Gender budgeting is a fiscal innovation that translates gender-related goals into budgetary commitments and can help countries achieve Sustainable Development Goals regarding gender equality (Chakraborty et al. (2019)).

Gender budgeting has also been implemented at the subnational level in many developing countries like Indonesia (Salim (2016)) and India. In India, the adoption process began in 2005 with different adoption waves and following an adoption by the central government in 2000. The Union initiative was institutionalized through the development of a system of classification of budgetary transactions, and the formation of groups or cells in each Ministry of the government to lead efforts to identify gender-related goals and ways to achieve these goals through the budget (Stotsky and Zaman (2016)). Drawing upon the central government framework, states began to adopt gender budgeting, starting with Odisha in 2005. Since then, most Indian states have had some form of gender budgeting in place. Gender budgeting at the state level in India offers a suitable empirical framework for assessment of its effects because several states have adopted and sustained gender budgeting efforts (Chakraborty (2016) and Stotsky and Zaman (2016)). State-level gender budgeting in India has also used the national-level analytical matrices and templates of the National Institute of Public Finance and Policy (NIPFP) ³. This framework implies an ex-post analysis of the budget through a gender lens and a Gender Budget Statements that summarize the state (or national) effort implemented to reduce the gender gap and/or reach the objectives. Gender budgeting as previously explained requires assessments and transparency for the objectives and the results of the public policies either to reduce the gender gap and to reach Sustainable Development Goals which must be a common objective for Indian States. The agreement about the power-sharing between national and states governments in India led to the fact that States have exclusive powers over the 66 items enumerated in the State List including public health, sanitation, hospitals, and dispensaries. Indian states are responsible for health services and reproductive health to reach (jointly with the national government) the Sustainable Development Goals (SDG). For the education policies, the central government and the States

³In 2002, the Government of India commissioned the National Institute of Public Finance and Policy (NIPFP), the think tank of the Ministry of Finance, to undertake a comprehensive study on gender budgeting

can legislate any aspect of education from the primary to the university level. In case of any dispute, legislation framed by the central government will have overriding authority. By having education in the Concurrent List, the central government can directly implement any policy decision in the States. Gender inequalities and SDG are related to education (Buchmann et al. (2008); Kleven and Landais (2017)) and health (Okojie (1994); Sen and Östlin (2008)). However, in the Indian subnational context, only health public services are fully under the responsibility of Indian States. To check the effects of subnational gender budgeting adoption, this analysis will focus on the health sector efficiency.

In addition to its effects on gender inequalities, Gender Budgeting can also affect fiscal policies and fiscal transparency. Indeed, Gender Budgeting is not only a simple accounting exercise but an ongoing process of keeping a gender perspective in policy/ program formulation, implementation, and review. Its adoption implies assessing ex-ante and ex-post the spending and publishing a statement about the objectives and the results of the public spending in the related sector. The gender budgeting process needs to target the objectives of public policies in terms of gender inequality reduction through health public policies. The graph⁴ 1 summarize the Gender Budgeting framework and how it is included in the budgetary process.

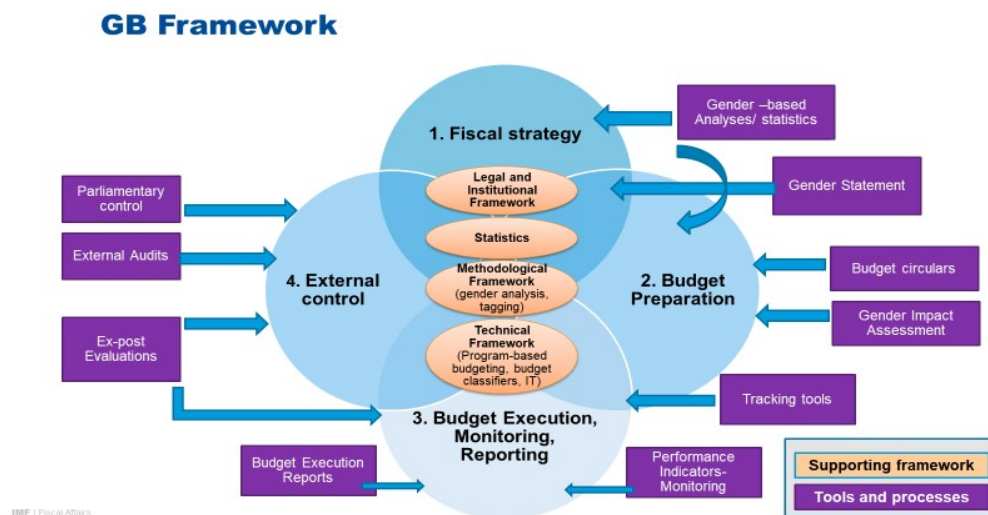


Figure 1: Gender Budgeting framework (Source: IMF PFM Blog)

The obligation to release a statement about the policy objectives improves the transparency

⁴<https://blog-pfm.imf.org/en/pfmblog/2021/02/sub-saharan-africa-course-on-gender-budgeting>

in the budgetary process. Indeed, Gender Budgeting statements summarise the governments' implications with several key indicators. It requires a high degree of coordination throughout the public sector and is essentially an accountability report by the government regarding its commitment to gender equity. [Chen et al. \(2019\)](#); [Chan and Karim \(2012\)](#); [De Simone et al. \(2019\)](#) and [Montes et al. \(2019\)](#) show that transparency in the fiscal policy improves the public spending efficiency. More precisely, [Gavazza and Lizzeri \(2009\)](#) explains that transparency of spending has more benefits for public spending efficiency than revenue transparency. This publication and the respect for Gender budgeting rules and duties led to a reinforcement of local administration and strengthening of the gender budgeting process throughout the year. This could lead to a positive effect which will remain through time and an improvement of local executives' capacities. The existence of common templates and objectives to follow could impose a constraint which will ensure that States that have adopted gender budgeting will follow the process and rules which could lead to greater transparency. In addition, India provides access to good-quality data on fiscal variables, and other demographic variables at state-level, over the period before and during the gender budgeting efforts.

The next graphs summarize the adoption of gender budgeting adoption through time and different adoption waves. These graphs clearly show a time trend effect on gender budgeting adoption.

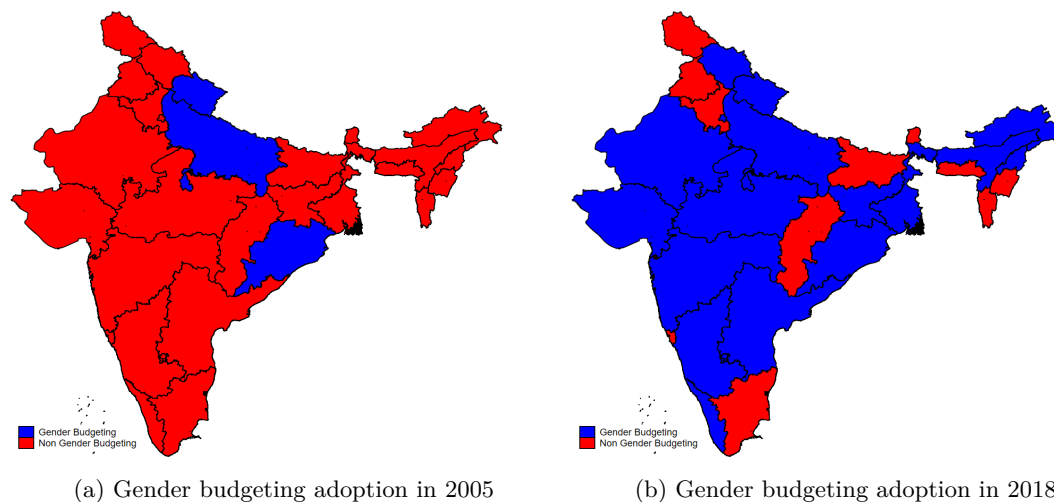


Figure 2: GB adoption through the time

The previous studies about gender budgeting have been focused either on the effects on spending composition and/or social outcome (Chakraborty (2016); Stotsky and Zaman (2016); Quinn (2016); Jung (2022)) or the determinants of success and failures of its implementation (Puig-Barrachina et al. (2017); Elomäki and Ylöstalo (2021)). None of them have empirically assessed the potential side effects of gender budgeting on spending efficiency through the transparency and budgetary process channel.

This paper is the first that wants to assess this side effect of gender budgeting on spending efficiency at the subnational level. Using state-level panel data, with Difference in Difference and entropy balancing as estimation tools, we find that health public spending efficiency improved significantly in gender budgeting states compared to states that did not put in place gender budgeting. The results remain positive and significant over time regardless of the estimation tool. This supports the conclusion that gender budgeting could lead to a better efficiency score at the state level through its positive externality.

The rest of the article is structured as follows: section 2 presents the methodology for the efficiency score and the analysis of the parallel trend assumption. Section 3 presents the first results of gender budgeting adoption with DiD staggered estimators of Callaway and Sant’Anna (2021). The results of the additional robustness analyses with alternative DiD estimators and entropy balancing are presented in section 4. Section 5 tries to assess the potential transmission channels. Finally, section 6 concludes.

2) Methodology

2.1) Data

2.1).1 Efficiency score: the outcome variable

The efficiency frontier approach relies on the computation of the production frontier curve that represents the highest output level reachable using a given set of inputs. This curve materializes the technical efficiency frontier. All Decision-Making Unit (DMU) on the frontier is technically fully efficient and the distance between a unit and the curve is a measure of inefficiency. The efficiency frontier can be estimated through parametric or non-parametric methods. We esti-

mate our efficiency score using the efficiency frontier analysis. However, our approach differs from theirs insofar as we opt for the parametric method, namely the Stochastic Frontier Analysis (SFA), rather than the non-parametric one. Several reasons motivate our strategy. First, the non-parametric techniques, especially the DEA and FDH (that are widely used), rely on linear optimization programs to build a convex curve that designs the efficiency frontier. As a deterministic method, they ignore the random variation in the data, measurement error and any stochastic influence. In the specific case of public investment, some unanticipated and noisy shocks such as fall in oil prices, political crises, etc. may influence the way that governments will provide public infrastructure independently of their "true" inefficiency. As such, for the same amount of public investment, state A, which suffers from the unexpected shocks, will have systematically a lower public infrastructure output than state B. It would be inappropriate to interpret this "bad luck" as inefficiency. Fortunately, SFA allows us to disentangle the inefficiency arising from differences in socioeconomic contexts or "bad luck" from the right efficiency related to bad public sector management. Second, the deterministic approach is very sensitive to the presence of outliers, sample size and in the case of heterogeneous units [Fiorentino et al. \(2006\)](#).

The estimation of efficiency score has been made by using the methodology of [Kumbhakar et al. \(2015\)](#) which is used by [Bamba \(2020\)](#), [Shen and Chen \(2017\)](#), [Adom et al. \(2021\)](#) and [Kang et al. \(2022\)](#) among others. The [Kumbhakar et al. \(2015\)](#) approach is suitable because it can control the unobserved heterogeneity and separate it from inefficiency. Heterogeneous characteristics of countries regarding their economic development, their political situations, or external shocks can be interpreted as inefficiency.

The use of the [Kumbhakar et al. \(2015\)](#) estimator is suitable in our case because it controls for the unobserved heterogeneity between decision-making units and separates them from the inefficiency. Especially in the panel cross-state analysis, heterogeneous characteristics of countries regarding their economic development, and their political situations may influence the public infrastructure provision without reflecting a bad or good public management.

The prediction of efficiency score followed the method of [Nguyen et al. \(2021\)](#) which is an implementation of [Kumbhakar et al. \(2015\)](#) with a segmentation of the error term " ϵ " between the pure noise, the short run inefficiency and the long-term (or persistent) inefficiency.

$$Y_{it} = \alpha + \beta X_{it} + \epsilon_{it} \quad (1)$$

Where Y_{it} is the output variable, X_{it} is the vector of our inputs variables. i refers to the state and t to the year. The error term ϵ is divided into three components. The new equation will be:

$$Y_{it} = \alpha + \beta X_{it} + v_{it} - u_{it} - \eta_{it} \quad (2)$$

In this estimation, v_i represents the pure noise, which is independent and identically distributed, ν_i is the short-run technical inefficiency and μ_i captures the long-run (persistent) inefficiency. As in [Bamba \(2020\)](#) we realize the estimation in two steps. We first estimate the next equation to get an estimation of the parameter β and the predicted value of θ_i , γ_{it} , $\hat{\theta}_i$ and $\hat{\gamma}_{it}$.

$$y_{it} = \alpha_0^* + \beta X_{it} + \theta_i + \gamma_{it} \quad (3)$$

Where

$$\alpha_0^* = \alpha_0 - E(\eta_{it}) - E(u_{it}) \quad (4)$$

$$\theta_i = \alpha_i - \eta_i + E(\eta_i) \quad (5)$$

$$\gamma_{it} = v_{it} - u_{it} + E(u_{it}) \quad (6)$$

After the first step, we realize a stochastic frontier method to estimate the persistent and transient (or short-run) technical inefficiency \hat{u}_{it} . Finally, we compute the time-varying technical efficiency and use it for the empirical analysis.

As mentioned above, the estimation of frontier analysis requires specifying at least one input and one output. In the public sector context, an output can be understood as a measurable variable. that reflects the performance or the achievement of government in a specific sector. For example in the health one, it could be the maternal or infant mortality ratio. In our case, due

to the lack of available data about other sectors, we will focus our work on the health one. The output used for the estimation of the efficiency score is the infant mortality ratio. This measure can give us a good proxy of the effectiveness of the State's health policy. Indeed, Indian States have as a mission to provide good maternal health services in their jurisdictions. In the same way, the reduction of maternal mortality ratio is one of the targets of Sustainable Development Goals used in developing countries as a target for the gender budgeting process. In addition, this measure is widely used in the literature ([Jafarov and Gunnarsson \(2008\)](#) and [Verhoeven et al. \(2007\)](#) for example).

2.1).2 The treatment and control variables

The treatment variable is a dummy which takes 1 if gender budgeting is implemented in a state and 0 otherwise. It comes from the paper of [Stotsky and Zaman \(2016\)](#) and has been updated by further research from literature and state governments' disclaims.

Table 1: Repartition of treatment

Treated	142
Untreated	381

The control variables are a set of covariates used in the literature on public spending efficiency which can also affect the likelihood to adopt or not gender budgeting.

As explained by [Boetti et al. \(2012\)](#), the subnational government's fiscal autonomy leads to some less inefficient behaviour. These states are also less dependent on central government transfers and are more autonomous in their political choices. The fiscal autonomy variable is a ratio between states' own local revenues and their total revenues (transfers and grants included). The most urbanized states can generate some scale economies, or sometimes some congestion effects which make less effective and less efficient public spending and policies related to health issues. Taxation influences public spending efficiency as explained by [Afonso et al. \(2021\)](#). So, the subnational autonomy appears to be a good control variable for the estimation process. [Sibiano](#)

and Agasisti (2013) and Rayp and Van De Sijpe (2007) highlight a link between GDP per capita and public sector efficiency. Gross domestic product per capita appears as the key determinant of efficiency in Italian regions. At the same time, GDP per capita affect the accountability of rulers and their decisions to adopt or not gender budgeting process. The share of seats held by women in local parliament influences the composition of public spending at the subnational level (Svaleryd (2009)). The presence of women in local parliament also affects the political decisions and the choice of gender budgeting adoption. All the variables have a year lag to tackle or reduce the endogeneity.

The following table summarizes the main variables used in the estimation process.

Table 2: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
eff_trans	99.417	0.164	98.781	99.734	562
eff_bc95	96.622	1.791	91.290	99.479	562
gender_budgeting	0.293	0.455	0	1	720
lurban	31.148	14.355	7.98	86.100	592
lloggdppc	10.417	0.878	8.02	12.728	679
lautonomy	51.23	25.163	5.466	95.34	675
llocal_wip	48.608	1.916	42.39	54.87	655
trend	12.5	6.927	1	24	720
logpop	2.864	1.731	-0.777	5.476	596
lAgri_GDP	23.891	12.309	1.79	77.888	675
lBanking_GDP	4.746	2.431	1.15	16.161	675
ldeficit	-2.248	17.373	-90.953	42.528	675

2.1).3 Stylized facts

The graph 3a highlights a comparison between the average inefficiency score for the treated (1) and untreated units (0). The efficiency score is very close to the sample, so it's difficult to apprehend the difference between treated and untreated units. This is why we choose to change the scale and focus the graph on the inefficiency score. This seems to suggest that states

that have adopted gender budgeting are less inefficient (more efficient) than those that have not adopted it. However, this correlation means nothing because a correlation does not necessarily imply causality. This result seems to confirm the intuition and provide avenues to explore for further analysis.

The graph 3b shows the relationship between the efficiency score and the time since the first adoption of the gender budgeting process. This relation seems positive and can suggest a self-enforcement process of the gender budgeting process. The positive relationship can be explained by the fact that local administrations learnt from their previous success and/or failures. In addition, the continuous disclosure of budgetary information can increase the citizens' interest and improve accountability about the use of public money in health-related public policies.

The graph 3c summarizes the average efficiency score for untreated (0) and treated units (1) by comparison to the global average of the sample (the vertical red line). The difference between both means is not important but we can see that the treated units seem to have a better efficiency score than the untreated ones, and the global sample. Finally, the last graph 3d shows the evolution of the efficiency score mean by year. It tends to confirm the hypothesis of a positive effect of time trend on the efficiency score for the sample.

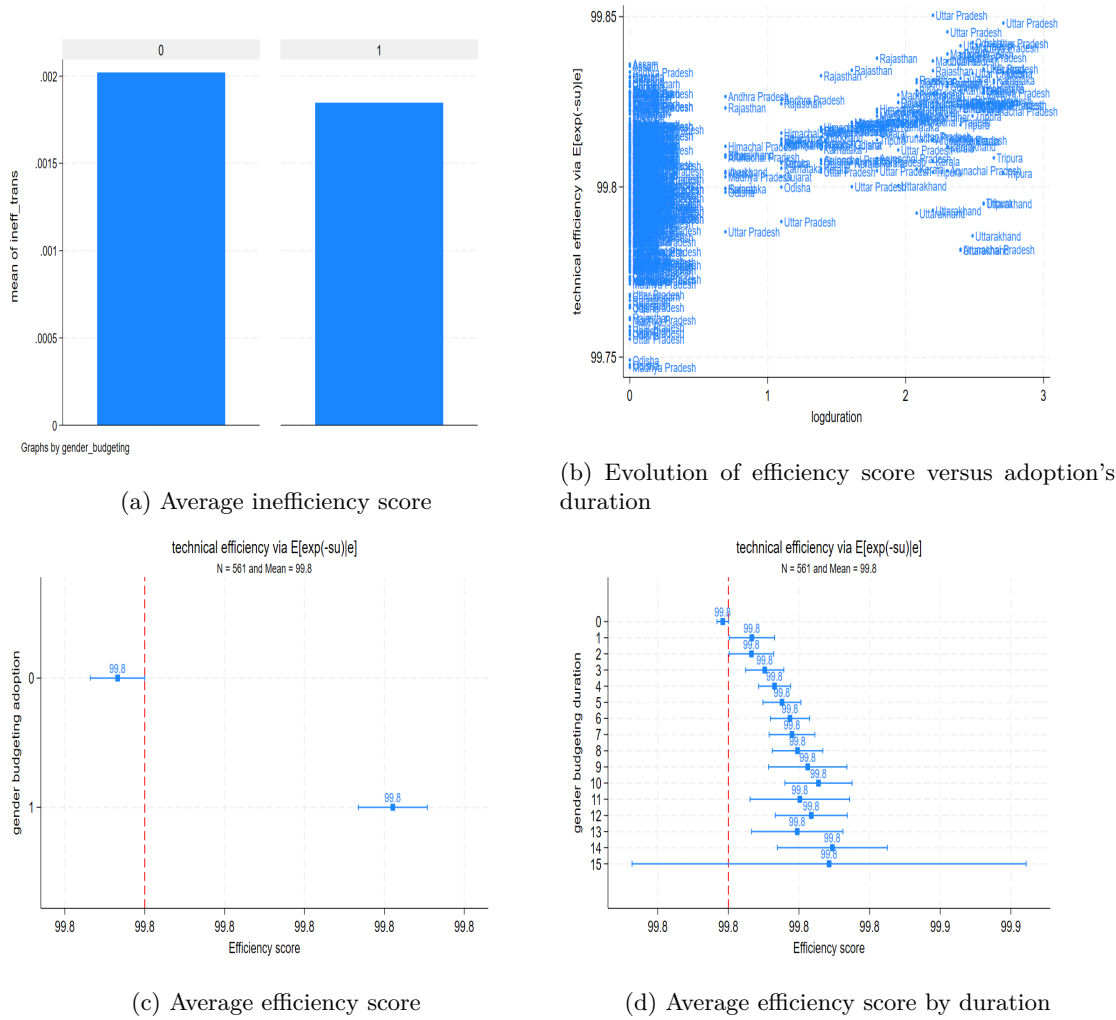


Figure 3: Stylized facts

2.2) Identification strategy

The identification method used is a Difference in Difference (DiD) strategy, using a comprehensive panel dataset. We focus on efficiency scores between 1997 and 2020, our key outcome of interest. The decision to adopt gender budgeting in each state is not random. Therefore, the main challenge is to correct for selection into the reform, i.e., to account for differences between adopter and non-adopter jurisdictions that could have influenced the outcome. The DiD identification strategy makes it possible to correct for the initial difference in efficiency score and thus estimate

the differential changes in these outcomes across states before and after each wave of adoption.

However, by using several years of data makes our approach closer to a two-way fixed effects (TWFE) linear regression. Recent methodological papers characterise the potential issues surrounding TWFE with multiple periods and multiple treatments (Callaway and Sant’Anna (2021), Borusyak et al. (2022) Goodman-Bacon (2021) and De Chaisemartin and d’Haultfoeuille (2020)). One issue addressed in this literature is the cross-unit heterogeneity of treatment. Other issues include the time-heterogeneity of treatment and the use of units that eventually become treated as control groups. When extending to 1997–2020, we try to capture longer-term effects and check if there is an increasing advantage of early adoption. We also acknowledge a group of states which have adopted gender budgeting after the first wave, which might slightly perturbate the control group as some units become treated. To address this, we suggest additional estimations where we explicitly account for the two types of treatment. In technical terms, we estimate the following equation in which y_{it} is the outcome variable, i.e., efficiency score for state i in year $t = 1, \dots, T$

$$Y_{it} = \alpha + \beta^W D_{it}^W + \rho X_{it} + \theta_i + \gamma_t + \epsilon_{it} \quad (7)$$

With the treatment dummy variable equal to 1 if the state i belongs to the group of states which have adopted gender budgeting in year k and are observed after that year.

To slightly enhance the DiD setup, we use the Callaway and Sant’Anna (2021) DiD approach. The Callaway and Sant’Anna (2021) DiD estimator allows us to use inverse probability weighting as in Abadie (2005). As with Abadie (2005), we must estimate the propensity score. However, because we have multiple treatment dates for multiple groups, there is a unique propensity score for every group. However, we do not have the luxury of a large reservoir of untreated units necessarily in many applications with multiple periods and differential timing. To create implicit pairings of units in the treatment and comparison groups, Callaway and Sant’Anna (2021) allows two options. Using a pool of units as our comparison group who never are treated during the duration of the panel. Or we may use a pool of units that have simply not yet been treated by the time of treatment. Another key concept in Callaway and Sant’Anna (2021) is the group-time ATT. The group-time ATT is a unique ATT for a cohort of units treated at the same point in

time.

The *csdid* package used for this estimation allows us to estimate with Callaway and Sant’Anna (2021) methods an estimator like Abadie (2005), but by considering the staggered adoption and heterogeneous effects. This type of approach usually brings flexibility to traditional DiD setups. Most importantly, it is used here to try to reduce unobserved time-varying differences between early- and late-gender budgeting-adopting states that could confound our results. For this, we are going to mobilise a set of variables X_{it} that are assumed to be correlated to some extent with time-varying confounders and that allow for comparing subgroups of treated and control states that are more alike.

For example, if states with the greatest GDP per capita are the ones that adopted gender budgeting first and, at the same time, are the ones that benefit from efficiency score (internal validity issue) or stand to benefit most from gender budgeting because their important GDP per capita can mean a greatest interest for central government to rule this state. So, it can increase the discretionary transfers which are targeted at specific purposes (external validity issue), and then we might overstate the benefits of the gender budgeting adoption. Assuming that the unobservable advantages (e.g. economic and cultural dynamics, political leverage or interest) are correlated with observable characteristics (e.g. population size, autonomy, GDP per capita), we could reduce the bias by comparing treated and control states that are most similar along a relevant set of observed characteristics of that sort. Rather than using matching on many different characteristics, which brings a ‘curse of dimensionality issue, we rely on a propensity score (PS) that concentrates all the useful information from these characteristics. The propensity score, denoted p hereafter, is obtained as the prediction of a first-stage estimation of a gender budgeting dummy on the set of relevant variables including key demographic dimensions such as urbanization ratio, density rate, GDP per capita, autonomy ratio (share of own revenues on total states revenues) and proportion of seats held by the women in state parliament. To consider treated and untreated states that are more like each other according to these different criteria simultaneously, we reweight observations using the inverse propensity score, as suggested by Abadie (2005) for the DiD approach. In this way, the modified estimation gives more weight to the late (early) gender budgeting adopters that are most similar to the early (late) gender budgeting adopters. We will also explore the heterogeneous impact of the reform by explicitly

zooming in on groups with similar characteristics (e.g. treated and controlled states with high wealth). All estimations are clustered at the state level to account for autocorrelation.

2.3) Parallel trend assumption

The graphs [4a](#) and [4b](#) highlight the evolution of efficiency scores for treated and untreated units through time. This concern about the parallel trend is crucial to confirm the internal validity of the Difference in Difference (DiD) Method. It provides some reassurance that untreated units could provide a reasonable counterfactual, particularly if they most closely resembled the treated ones. It also allows us to reduce the concerns about the fact that our results could be driven by pre-trend effects which would bias our results. The graph [4b](#) provides an overview of efficiency score evolution through the study period and a graphic estimation of the pre-trend assumption over this period. Graph [4a](#) provides a zoom over the period 2002-2009. This zoom allows us to check if the pre-trend assumption holds a few years before the treatment was applied as used in [Callaway and Sant'Anna \(2021\)](#) (applied in robustness check) which suggests a weaker pre-trend assumption about the duration of parallel pre-trend between treated and untreated units.

So, to compare treated and control states that are most similar, we also suggest DiD estimations adjusted by a quasi-matching strategy. Assuming that the matching variables are highly related to unobserved confounders, this approach should reduce the potential bias affecting trend differences between the groups of states which have adopted gender budgeting at different points in time.



Figure 4: Parrallel trend

The following graph (5) has been inspired by the work of [Rambachan and Roth \(2023\)](#) on a more credible approach to the parallel trend assumption. They propose some tools for robust inference in difference-in-differences and event-study designs where the parallel trends assumption may be violated. Instead of requiring that parallel trends hold exactly, we impose restrictions on how different the post-treatment violations of parallel trends can be from the pre-treatment differences in trends (“pre-trends”). We recommend that researchers use our methods to construct robust confidence intervals, under restrictions on the possible violations of parallel trends that are motivated by domain knowledge in their empirical setting. According to them, there are some key concerns about the pre-trend assumption. Despite the statistical or visual results, it’s important to consider some macroeconomic shocks that can disturb the pre-trend evolution. Figure 5 shows robust confidence sets for the treatment effect, using different values of $Mbar$ ⁵. The figure shows that if we impose $Mbar=1$, meaning that we restrict the post-treatment violations of parallel trends to be no larger than the maximal pre-treatment violation of parallel trends, then we obtain a robust confidence set for the causal effect on the efficiency score. This is wider than the original (without covariates) confidence interval, which is only valid if parallel trends hold exactly, but rule out a null effect on efficiency score.

The intuition for why the confidence sets are larger through time is that we have bound the

⁵ $Mbar$ is a degree of smoothness, or how much we allow a violation of pre-trend assumption

violation of parallel trends across consecutive periods by Mbar times the max in the pre-treatment period. Thus, the identified set will be larger for later periods, since the treatment and control groups have more time to diverge. If we are willing to bound the magnitude of economic shocks by the max in the pre-treatment period, we will thus typically obtain wider confidence sets for parameters involving later periods. As suggested by [Rambachan and Roth \(2023\)](#) the table 14 available in the appendix summarizes the different bands of confidence interval according to the Mbar values.

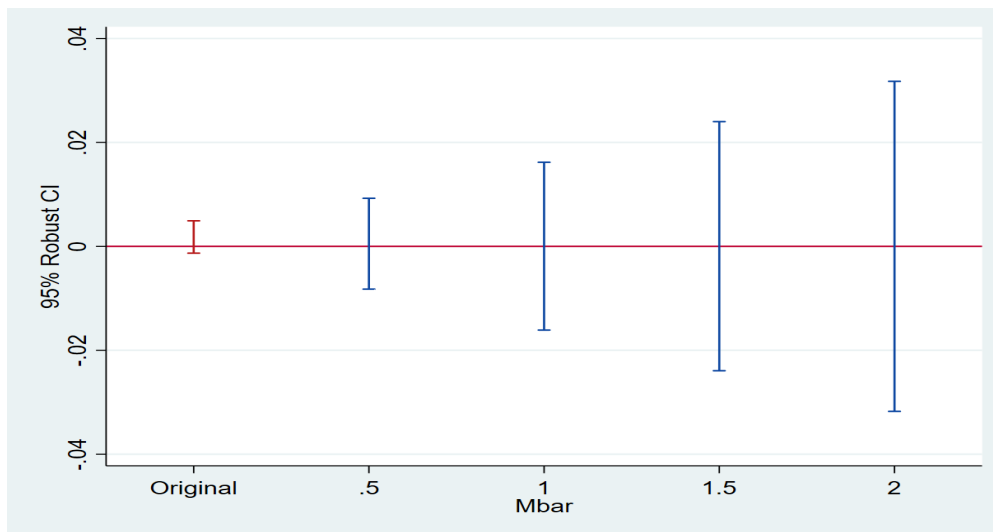


Figure 5: Parallel trend with honest DiD ([Rambachan and Roth \(2023\)](#))

3) Results

The results are available in the next table. They suggest a positive effect of gender budgeting on efficiency scores. Indeed, analysing budgets from a gender perspective is integral to gender mainstreaming. If gender considerations have been built into policies and project design, they should reflect in resource allocation; and if they have not, the outcomes are not likely to deliver substantive equality for women. Budgets are thus a critical tool for mainstreaming. Such exercises increase the transparency, disclosure of information, and participation by citizens in, economic governance. Many studies ([Zhang and Wang \(2021\)](#), [Chan and Karim \(2012\)](#), [Chen et al. \(2019\)](#), [De Simone et al. \(2019\)](#) and [Montes et al. \(2019\)](#)) have shown a clear link between

transparency and spending efficiency. So, gender budgeting by increasing transparency and local administration competencies improves the efficiency of public spending in Indian states which have adopted it.

Table 3: Diff in Diff results

	Callaway	Callaway	Callaway	Callaway
ATT	0.123***	0.137***	0.129***	0.121***
	(4.22)	(3.30)	(3.16)	(3.39)
Observations	419	401	398	388

t statistics in parentheses

Source: Author

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This effect remains positive through time as illustrated in the graph 6. The increasing trend of coefficients in the Event plot (graph 6) could suggest a continuous improvement in the efficiency score, but also in the administration capacities which can learn from their previous experiences. The results seem to confirm the learning by-doing process suspected by looking at the data and the fact that states which have adopted gender budgeting seem to improve more in the efficiency of their public spending.

In addition to these results, we use another DiD method to check the robustness of our results. The next section presents these methods and their results.

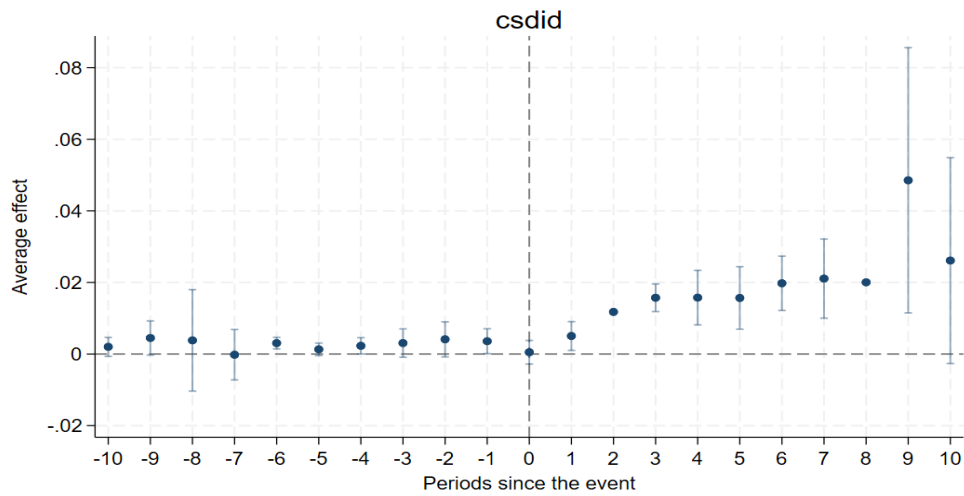


Figure 6: Event study

4) Robustness check: alternative estimation method

4.1) Alternative Diff in Diff methods

There is a huge literature about the DiD estimators, but to check the robustness of our analysis we will consider the estimators proposed by [Wooldridge \(2021\)](#). This estimator proposes an alternative regression adjustment procedure that is suitable for staggered setups. His proposed estimator differs from the [Callaway and Sant'Anna \(2021\)](#) regression adjustment estimator as he exploits additional information from pre-treatment periods, which, in turn, can lead to improvements in precision ([Roth et al. \(2023\)](#)).

The results are presented in the next table and show a positive effect of gender budgeting adoption on efficiency scores.

Table 4: Diff in Diff results with Wooldridge, 2021

	Wooldridge	Wooldridge	Wooldridge	Wooldridge
ATT	0.0101***	0.00969***	0.00620***	0.00440***
	(3.88)	(3.59)	(2.32)	(1.71)
Observations	478	477	477	477

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The results tend to confirm the first ones got with [Callaway and Sant'Anna \(2021\)](#)'s one. We also compute an event result that allows us to observe the coefficient of the treatment effects through time. These results are available in the next graph. The effect of the treatment adoption remains positive and statistically significant even after the policy intervention.

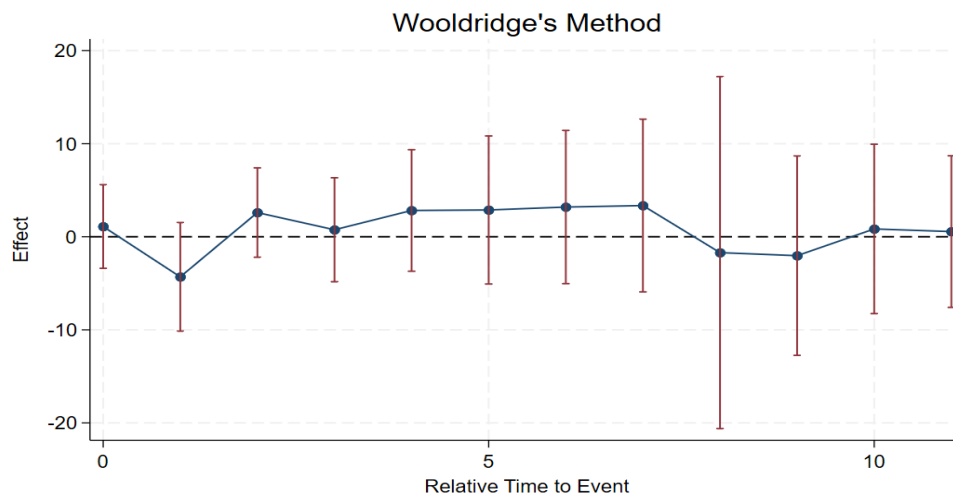


Figure 7: Wooldridge Results

Another related approach is to run a stacked regression where each treated unit is matched to 'clean' (i.e. not-yet-treated) controls and there are separate fixed effects for each set of treated units and its control, as in [Cengiz et al. \(2019\)](#) among others. This approach estimates a convex weighted average of the *ATT* under parallel trends and no anticipation, although the weights are determined by the number of treated units and variance of treatment within each stacked event,

rather than by economic considerations.

We implement the stacked event study estimator implemented in [Cengiz et al. \(2019\)](#). This method appends together individual datasets or stacks. Each stack includes all observations from a cohort of units that received treatment in the same period and all units that never received treatment. Effects are identified within each stack by comparing an individual cohort of treated units to never treated units. That approach avoids erroneous comparisons of late to early implementing units that may bias Two-Way Fixed Effects (TWFE) estimates if effects vary across treated cohorts ([Goodman-Bacon \(2021\)](#)). The stacked event study is estimated in three steps. First, individual stacks are created. Second, they are appended together. Finally, estimates an event study that includes unit-by-stack fixed effects, time-by-stack fixed effects, and standard errors clustering on the unit-by-stack.

The estimator in [Callaway and Sant’Anna \(2021\)](#) uses the last period before units get treated (t_s-1), as the baseline outcome, while the estimator in [Borusyak et al. \(2022\)](#) uses the average outcome from period 1 to t_s-1 . This is why the latter estimator is often more precise. However, it is also more biased, when parallel trends do not exactly hold and the discrepancy between groups’ trends gets larger over longer horizons, as would, for instance, happen when there are group-specific linear trends. In such instances, [Roth \(2022\)](#) notes that leveraging earlier pre-treatment periods increases the bias of a DID estimator since one makes comparisons from earlier periods. If, on the other hand, parallel trends fail due to anticipation effects arising a few periods before t_s , the estimator in [Borusyak et al. \(2022\)](#) is less biased than [Callaway and Sant’Anna \(2021\)](#)’s one. Another difference between these approaches is that [Borusyak et al. \(2022\)](#) imposes parallel trends for every group and between every pair of consecutive periods. [Callaway and Sant’Anna \(2021\)](#), on the other hand, imposes a weaker parallel trends assumption: from period t onwards, cohort w must be on the same trend as the never-treated groups, but before that cohort w may have been on a different trend. The assumption in [Callaway and Sant’Anna \(2021\)](#) is the minimal assumption, but it is conditional on the design: which groups are required to be on parallel trend at which dates depends on groups’ realized treatments. Overall, this discussion suggests that whether the estimators in [Borusyak et al. \(2022\)](#) should be preferred to [Callaway and Sant’Anna \(2021\)](#) may depend on one’s degree of confidence in the parallel trends assumption, on the type of violations of this assumption that seems more likely to arise in the application at hand, and

on one's willingness to undertake a sensitivity analysis such as the one proposed by [Rambachan and Roth \(2023\)](#). We also use the [Gardner \(2022\)](#); [Cengiz et al. \(2019\)](#); [Dube et al. \(2023\)](#) and [De Chaisemartin and d'Haultfoeuille \(2022\)](#) estimators and combine both event studies on the next graph.

The event study graph of the alternative DiD estimators is available in the graph 8. It shows the coefficients for the different estimators and represents their evolution through time. Even the size of the effect is different according to the estimators used, but all the estimators highlight the positive effect of gender budgeting on efficiency scores. Despite the low size of the effect, it's important to consider the fact that this size represents half of the efficiency score standard deviation for the first seven years after the treatment. Despite the relatively low size of the effect *per se*, it represents an important deviation in the context of this efficiency score measure. As explained by [Rambachan and Roth \(2023\)](#) the confidence interval tends to be wider over time. This can be explained by the fact that the treated and untreated units have more time to diverge, and because it's more difficult to accurately determine the confidence interval many periods after the treatment adoption.

These results suggest that the previous results obtained with [Callaway and Sant'Anna \(2021\)](#) are robust and not driven by a bias in propensity score estimate by the method itself. To ensure that the results are not driven by a bias in the efficiency score estimation method, we use an alternative method to estimate the efficiency score. We remake the same process with an efficiency score estimated by using [Battese and Coelli \(1995\)](#) process. The results obtained with this method are summarized in the appendix.

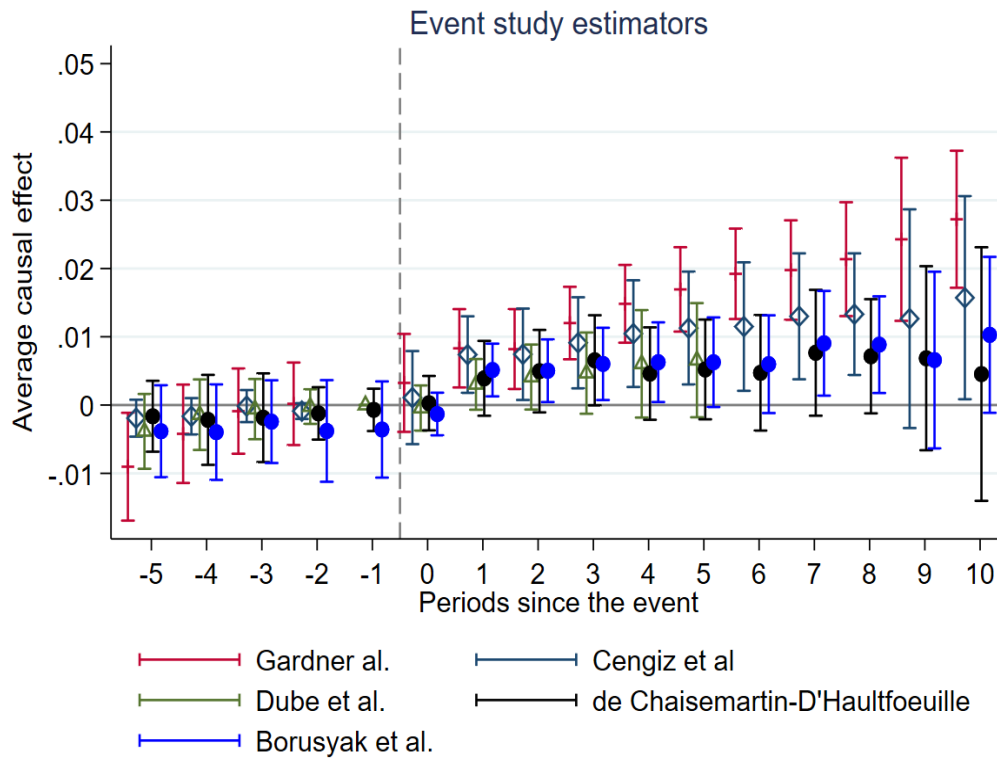


Figure 8: Alternative Diff in Diff Results

In addition, to ensure that the results are not led by a bias in the DiD method itself or a misestimation of the parallel trend assumption, we also use the entropy balancing estimation method to reinforce the robustness of our results.

4.2) Entropy balancing

4.2).1 Methodological concept

For the robustness check, we also use the entropy balancing method of [Hainmueller \(2012\)](#) like [Baccini et al. \(2018\)](#) who worked on fiscal decentralization and tax competition between local jurisdictions. Because many macroeconomic shocks have been able to change the expectations of the population, state rulers or local administrations. The announcement of the gender budgeting adoption could also raise the expectations of the population in terms of public service quality.

At the same time, the state rulers could have the incentive to improve the quality of their public spending to get the people’s favour, even if they have not adopted gender budgeting. The competition effect can affect the pre-trends and bias the results. For this reason, we use several DiD estimators which are less restrictive on the pre trends assumptions, but also the entropy balancing of [Hainmueller \(2012\)](#) which doesn’t require the pre-trend assumption.

In general, the idea of matching estimators is to mimic randomization regarding the assignment of the treatment. The unobserved counterfactual outcome is imputed by matching the treated units with untreated units that are as similar as possible regarding all pre-treatment characteristics that are associated with selection into treatment and influence the outcome of interest.

Entropy balancing is a pre-processing procedure that allows researchers to create balanced samples for the subsequent estimation of treatment effects. The pre-processing consists of a reweighting scheme that assigns a scalar weight to each sample unit such that the reweighted groups satisfy a set of balance constraints that are imposed on the sample moments of the covariate distributions. The balance constraints ensure that the reweighted groups match exactly at the specified moments. The weights that result from entropy balancing can be passed to any standard model that the researcher may want to use to model the outcomes in the reweighted data—the subsequent effect analysis proceeds just like with survey sampling weights or weights that are estimated from a logistic propensity score covariate model. The pre-processing step can reduce the model dependence for the subsequent analysis since entropy balancing orthogonalized the treatment indicator concerning the covariate moments that are included in the reweighting.

Entropy balancing is implemented in two steps. First, weights are computed that are assigned to units not subject to treatment. These weights are chosen to satisfy pre-specified balanced constraints involving sample moments of pre-treatment characteristics by remaining, at the same time, as close as possible to uniform base weights. In our analysis, the balance constraints require equal covariate means across the treatment and the control group, which ensures that the control group contains, on average, units not subject to treatment that are as similar as possible to units that received treatment. Second, the weights obtained in the first step are used in a regression analysis with the treatment indicator as an explanatory variable. This yields an estimate for the Average Treatment on Treated (ATT), that is, the conditional difference in means for the outcome

variable between the treatment and control group. The advantage of entropy balancing over the other treatment effects methods is the fact that entropy balancing is not a parametric method. Indeed, this method does not need a specific empirical model for either the outcome variable or selection into treatment needs to be specified. Hence, potential types of misspecifications like those, for instance, regarding the functional form of the empirical model, which likely leads to biased estimates, are ruled out.

Moreover, with conventional matching methods, each untreated unit either receives a weight equal to 0, in the event it does not represent a best match for a treated unit, or equal to 1, in the event it does represent a best match for one treated unit. However, when the number of untreated units is limited and the number of pre-treatment characteristics is large, this procedure does not guarantee a sufficient balance of pre-treatment characteristics across the treatment and control groups. This is a serious problem, as a low covariate balance may lead to biased treatment effect estimates where the vector of weights assigned to non-treated units is allowed to contain non-negative values.

Finally, by combining a reweighting scheme with a regression analysis, entropy balancing allows us to properly address the panel structure of our data. We can control for both state-fixed as well as time-fixed effects in the second step of the matching approach, that is, the regression analysis. The inclusion of state-fixed effects is particularly helpful in accounting for potential unobserved heterogeneity across countries. The estimation of the ATT based on the matching will be:

$$\pi ATT(x) = E[Y(1)|T = 1, X = x] - E[Y(0)|T = 0, X = x] \quad (8)$$

Where Y represents the dependant variable, x is a vector of relevant pre-treatment characteristics, $E[Y(1)|T = 1, X = x]$ is the expected outcome for the units that received treatment, and $E[Y(0)|T = 0, X = x]$ is the expected outcome for the treated units best matches.

As pointed out by [Neuenkirch and Neumeier \(2016\)](#), entropy balancing has several advantages over traditional matching methods. First, unlike the propensity score matching methods or the difference-in-differences estimator, entropy balancing is a non-parametric approach, thus requiring no specification of the functional form of the empirical model or the treatment assign-

ment procedure, which may avoid specification errors or collinearity problems. Second, entropy balancing ensures a sufficient balance of pre-treatment characteristics between treatment and control groups, even in the presence of a small sample or a limited number of untreated units. This makes it possible to construct a suitable control group, representing a near-perfect counterfactual of the treated group. Finally, in the second step, the estimator exploits the longitudinal nature of the data by including individual and time effects to control for unobserved heterogeneity across units and biases due to changes over time, independent of treatment. [Tübbicke \(2022\)](#) and [Zhao and Percival \(2017\)](#) also show that entropy balancing is doubly robust concerning linear outcome regression and logistic propensity score regression, and it reaches the asymptotic semiparametric variance bound when both regressions are correctly specified. They suggest that entropy balancing is a very appealing alternative to the conventional weighting estimators that estimate the propensity score by maximum likelihood.

Our empirical equation to estimate the effects of the treatment on the outcome variable will be:

$$Y_{it} = \beta_1 T_{it} + \alpha_1 \log(GDP_pc)_{it} + \alpha_2 \log(density)_{it} + \alpha_3 X_{it} + \mu_i + \psi_t + \epsilon_{it} \quad (9)$$

Where Y is the degree of autonomy of state i in period t , and T is the treatment variable. The treatment takes the value 1 if the state has introduced gender budgeting and 0 otherwise. X_{it} is a set of time-varying characteristics of states. μ_i and ψ_t account respectively for states and time-fixed effects, capturing specific characteristics that may be correlated with the treatment. Finally, ϵ_{it} is the usual idiosyncratic error term assumed to be uncorrelated with the treatment.

4.2).2 Correlation issue

Table 5 shows a simple comparison of pre-weighting sample means of all matching covariates between treated (Column [2]) and control (Column [1]) states, which represent the potential synthetic group. Column [5] shows significant differences between the two groups for all pre-treatment variables, as some p-values are below the threshold of 5%. Such differences could bias the true treatment effect due to a potential selection problem. Therefore, in Panel B (Column [1]), we compute a synthetic control group by re-weighting the control units, using the pre-treatment covariates from the benchmark specification. This approach allows us to make the

means of the pre-treatment covariates of the synthetic group as comparable as possible to those of the treated units. As can be seen in Column [5] of Panel B, the weighting eliminated any significant pre-treatment difference between the means of the treated and synthetic covariates. Thus, we can consider the synthetic group as a perfect counterfactual of the treated group.

Table 5: Descriptive statistics

	[1]	[2]	[3]= [1]-[2]	[4]	[5]
Panel A: Descriptive statistics	Untreated	Treated	Difference	t-test	p-value
llog(GDP per capita)	10.32	10.86	-0.54	-7.7498	0.0000
lautonomy	51.38	46.68	4.70	1.8520	0.0325
ltrend	17.10	23.99	-6.89	-13.98	0.0000
lurban	31.30	27.74	3.56	3.0999	0.0010
llocal wip	48.88	48.48	0.40	2.2479	0.0128
lAgriculture	24.89	25.18	-0.29	-0.2155	0.4148
lBanking	4.52	4.82	-0.30	-1.5957	0.0558
ldeficit	-6.50	3.55	-10.05	-7.0019	0.0000
llog(population)	2.56	3.43	-0.87	-5.4898	0.0000
Observations	381	142			

Table 6: Covariate balancing

	[1]	[2]	[3]= [1]-[2]	[4]	[5]
Panel B: Covariate balancing	Untreated	Treated	Difference	t-test	p-value
llog(GDP per capita)	10.86	10.86	0.000	0.000	0.996
lautonomy	46.68	46.68	0.000	0.000	1.000
ltrend	23.99	23.99	0.000	0.003	0.995
lurban	27.74	27.74	0.000	0.000	1.000
llocal wip	48.48	48.48	0.000	0.000	0.999
lAgriculture	25.18	25.18	0.000	-0.001	1.000
lBanking	4.819	4.82	-0.001	-0.000	0.999
ldeficit	3.557	3.55	0.007	0.004	0.997
llog(population)	3.43	3.43	0.000	0.000	0.998
Observations	142	142			

The results will be available in the next tables.

4.2).3 Results

Table 7: Results of entropy balancing

	(1)	(2)	(3)	(4)	(5)
VARIABLES	efficiency	t+1	t+2	t+3	t+4
gender_budgeting	0.053*** (0.016)	0.087*** (0.016)	0.078*** (0.017)	0.115*** (0.019)	0.133*** (0.020)
Observations	478	456	434	410	387
R-squared	0.634	0.604	0.586	0.514	0.521

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 8: Results of entropy balancing with additional covariates

	(1)	(2)	(3)	(4)	(5)
VARIABLES	efficiency	efficiency	efficiency	efficiency	efficiency
gender_budgeting	0.053*** (0.016)	0.052*** (0.016)	0.094*** (0.016)	0.095*** (0.016)	0.059*** (0.016)
Observations	478	478	478	478	478
R-squared	0.634	0.634	0.664	0.665	0.690

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 7 tends to confirm the positive effects of gender budgeting adoption on efficiency score through time. The size of the effect seems to be less important (20% of standard deviation) but confirms the previous positive effect. The results are also robust to the addition of additional covariates to check the robustness of the results as shown in the table 8. The results also seem to be positive and tend to increase over time. This tends to confirm the intuition of the graph 3b and the intuition that there is a positive self-enforcement of the gender budgeting process which can help to improve the efficiency score. The results are also robust to the addition of new covariates which can affect both, outcome, and the likelihood of adopting gender budgeting. The results tend to confirm the robustness of the results got by Callaway and Sant'Anna (2021).

The fact that entropy balancing does not require parallel trend assumption helps us to prevent the risks of bias in our results due to a breakdown in this trend despite the statistical conclusion as explained by Rambachan and Roth (2023). The results obtained by Battese and Coelli (1995) help us to ensure that the results are not led by efficiency measurement errors. The results for this efficiency score measure will be available in the appendix.

To also assess the effect and be sure that the positive effects are due to the adoption and gender budgeting process, we include a cross variable. We include a variable that combines the treatment variable and the share of seats held by the political party that rules the State i at the period t . Indeed, the results can be explained by the fact that a party held the absolute majority

or an important share of the seats in the local parliament and is free to implement its public policies without considering the opposition at the local level. To ensure that the results are due to the gender budgeting we include this variable. This variable hurts the efficiency score. This result means that the fact that a party has an important share of the seats by implementing at the same time gender budgeting does not explain the positive effects on the efficiency score. The negative sign for the variable can be explained by the lack of transparency and accountability in the front of population and local parliament. So, the transparency, and accountability induced by gender budgeting explain the positive effect of gender budgeting adoption.

Table 9: Results of entropy balancing

	(1)	(2)	(3)	(4)	(5)
VARIABLES	efficiency	efficiency	efficiency	efficiency	efficiency
gender_budgeting	0.015*** (0.006)	0.016*** (0.006)	0.023*** (0.006)	0.023*** (0.006)	0.018*** (0.006)
winner*GB	-0.000* (0.000)	-0.000* (0.000)	-0.000*** (0.000)	-0.000** (0.000)	-0.000** (0.000)
Observations	477	477	477	477	477
R-squared	0.643	0.643	0.664	0.666	0.688

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

5) Transmission channel

For this exercise, we have constructed a prevision “bias” index that is a measure of the difference between states’ health spent in the state i at period t and the share of health spending reported in the budgetary forecast made by the same state at the same period. The bias index is summarised

just below.

$$\text{bias_index}_{it} = |(health_spending_{it}/Total_expenditures_{it}) * 100 - (health_spending_forecasted_{it}/Total_expenditures_forecasted_{it}) * 100| \quad (10)$$

Table 10: Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
bias index	0.129	0.108	0.002	0.539	497

We have assumed that this bias index will allow us to apprehend the performance of subnational administration. The differences between forecast and realisation are possible and usual, but a systematic and important difference may mean a lower level of performance in its administration. We compute it as an absolute value to consider the distance (bias) between the forecast and the realisation. We made it because a systematic underestimation of expenditures in the forecast could be good news in terms of available funding, but it's not good news from the credibility and local administration capacities point of view.

To assess how gender budgeting can affect fiscal deficit and autonomy, we try to estimate the potential transmission channels by using the same process as [Neuenkirch and Neumeier \(2016\)](#). We compute the means of the three variables for (a) the treatment group during times when gender budgeting is in place, (b) the treatment group focusing only on years before gender budgeting implementation and (c) our synthetic control group obtained via entropy balancing. The results are outlined in table 11. The descriptive statistics indicate some differences between the control group obtained via entropy balancing and states which apply gender budgeting. When comparing the control group to the treatment group before gender budgeting was applied, however, we find that the latter is characterized by a notably better “credibility” (or accuracy). Indeed, before the treatment, the treated units seemed to be less credible (or accurate) than the untreated ones (with a bias of 0.13 for the treated versus 0.12 for the untreated), but this bias reduced after the adoption (0.10) for the treated units. The results obtained from a t-test are also highly significative with a p-value lower than 0.01 and seem to go in the same way as graph

4 where the treated units become better after the adoption of the policy.

Table 11: Transmission channel

	bias index
<i>Before adoption</i>	0.13***
<i>After adoption</i>	0.10***
Non Gender Budgeting	0.12***

These results seem to corroborate those of [Hory \(2016\)](#); [Fonchamnyo and Sama \(2016\)](#); [Olanubi and Olanubi \(2023\)](#); [Ouertani et al. \(2018\)](#) and [Cabezon et al. \(2015\)](#) that explain that good public financial management, better fiscal credibility, and strong tax administration⁶ positively affect spending efficiency. However, to check our transmission channel we realise another pairwise correlation between the bias measure and the efficiency score to assess if this negative expected relationship between bias measure and efficiency score exists in our data.

The results are available below.

Table 12: Pairwise correlation

	bias index
Efficiency score (GB)	0.1585***
Efficiency score (Non-GB)	0.1242

The results shown in the table 12 suggest a negative correlation between the size of the bias and the efficiency score. As expected, this means that the ability to reduce the bias could lead to an improvement in the efficiency score. The channel of “credibility” and local administration reinforcement could be one of the transmission channels by which gender budgeting can affect the efficiency of public spending at the state level.

To go further with classical correlation, we use a simultaneous model equation like [Ekoula et al. \(2023\)](#). The next table (13) summarizes the results for the two main variables. The results are highly significant and seem to confirm the previous results and the intuition about

⁶each Indian states have its own Finance ministry

the fact that the forecast credibility and better performance of local administration could be a transmission channel of the effect of gender budgeting on efficiency score.

A full table (20), with all the variables used in the simultaneous model is available in the appendix.

Table 13: Simultaneous equations

	(1)	(2)
VARIABLES	eff_trans	bias index
l.bias index	-1.048*** (0.359)	
gender_budgeting		-0.017** (0.008)
Observations	345	345
R-squared	0.470	0.546

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

6) Conclusion and policy implications

Through this work, we have tried to evaluate the effects that gender budgeting adoption can have on Indian state's health public spending efficiency for a panel of Indian states over the period 1997-2020. To answer this question, we use different DiD estimators and another novel method — entropy balancing — combining a matching approach with linear regression, thus mitigating endogeneity issues that may lead to scepticism about our conclusions. Our results suggest that Indian states which adopt gender budgeting tend to have better efficiency scores than those that do not. The potential transmission channel identified in this paper is an improvement of local administration capacities as denoted in the table 11.

Gender budgeting adoption by imposing fiscal policy and budgetary process continuous as-

assessment positively affects the overall fiscal framework. Gender budgeting is a useful policy tool which has some positive effects beyond its original purpose of tackling gender inequalities. Indeed, gender budgeting can play an important role in tackling gender inequalities and increasing social outcomes like [Chakraborty \(2016\)](#) and [Stotsky and Zaman \(2016\)](#) shown. However, it can also play a role in improving subnational public spending efficiency. This improvement may help subnational units to manage wisely their resources in the actual context of scarcity. Better spending efficiency can also help to improve the quality of public services delivered by subnational authorities.

From the point of view of policy implications, this analysis tells us that imposing constraints and practices that would normally be part of any policy reform helps to improve the quality of public spending at the local level. The definition of clear and understandable public policy objectives, the design of measures to achieve these objectives and the evaluation of these measures in the light of the objectives set are facilitated here by the moral necessity to fight against gender inequalities. The ability of local administrations to repeat this cycle over time provides an opportunity for the improvement of all kinds of public policies at the local level. The process will remain the same, only the targets will have to be set according to needs (waste management, urban policies, etc.). The determining elements in the success of these measures are the involvement of civil society ([Jung \(2022\)](#) and [Elomäki and Ylöstalo \(2021\)](#)) to maintain or improve the accountability of leaders and strong political support ([Steccolini \(2019\)](#) and [Polzer et al. \(2021\)](#)). A higher authority able to control the rule of gender budgeting obligations (as is the case for the Indian states) makes it possible to make up for a lack of accountability or to reinforce it if it is pre-existing. Even in the case of federal republics, the central government has some leverage over the states. Thus, in the absence of strong accountability, or supranational authorities, the results of this analysis cannot be transposed to the national level.

References

- Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The review of economic studies*, 72(1):1–19.
- Adom, P. K., Agradi, M., and Vezzulli, A. (2021). Energy efficiency-economic growth nexus: what is the role of income inequality? *Journal of Cleaner Production*, 310:127382.
- Afonso, A., Jalles, J. T., and Venâncio, A. (2021). Taxation and public spending efficiency: An international comparison. *Comparative Economic Studies*, 63:356–383.
- Baccini, L., Li, Q., Mirkina, I., and Johnson, K. (2018). Regional competition, business politicians, and subnational fiscal policy. *Business and Politics*, 20(3):410–437.
- Bamba, M. (2020). Does fiscal consolidation improve public investment efficiency. *CERDI Working papers*.
- Battese, G. E. and Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical economics*, 20:325–332.
- Boetti, L., Piacenza, M., and Turati, G. (2012). Decentralization and local governments' performance: how does fiscal autonomy affect spending efficiency? *FinanzArchiv/Public Finance Analysis*, pages 269–302.
- Borusyak, K., Jaravel, X., and Spiess, J. (2022). Revisiting event study designs: Robust and efficient estimation. *Available at SSRN 2826228*.
- Buchmann, C., DiPrete, T. A., and McDaniel, A. (2008). Gender inequalities in education. *Annu. Rev. Sociol.*, 34:319–337.
- Cabezón, E., Tumbarello, M. P., and Wu, M. Y. (2015). *Strengthening fiscal frameworks and improving the spending mix in small states*. International Monetary Fund.
- Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.
- Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454.

- Chakraborty, L. (2016). Asia: A survey of gender budgeting efforts.
- Chakraborty, L. S. et al. (2019). *Federal fiscal policy effectiveness and Inequality: Empirical evidence on Gender Budgeting in Asia Pacific*. National Institute of Public Finance and policy.
- Chan, S.-G. and Karim, M. A. Z. (2012). Public spending efficiency and political and economic factors: Evidence from selected east asian countries. *Economic Annals*, 57(193):7–23.
- Chen, G., Kang, H., and Luna-Reyes, L. F. (2019). Key determinants of online fiscal transparency: A technology-organization-environment framework. *Public Performance & Management Review*, 42(3):606–631.
- De Chaisemartin, C. and d’Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.
- De Chaisemartin, C. and d’Haultfoeuille, X. (2022). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. Technical report, National Bureau of Economic Research.
- De Simone, E., Bonasia, M., Gaeta, G. L., and Cicatiello, L. (2019). The effect of fiscal transparency on government spending efficiency. *Journal of Economic Studies*.
- Dube, A., Girardi, D., Jorda, O., and Taylor, A. M. (2023). A local projections approach to difference-in-differences event studies. Technical report, National Bureau of Economic Research.
- Ekoula, H. W. M. A., Kamguia, B., and Ndoya, H. (2023). Do women hold the key to financial sector development in africa? *International Economics*, 173:233–248.
- Elomäki, A. and Ylöstalo, H. (2021). Gender budgeting in the crossroad of gender policy and public financial management: The finnish case. *Public Money & Management*, 41(7):516–526.
- Fiorentino, E., Karmann, A., and Koetter, M. (2006). The cost efficiency of german banks: a comparison of sfa and dea. *Available at SSRN 947340*.

- Fonchamnyo, D. C. and Sama, M. C. (2016). Determinants of public spending efficiency in education and health: evidence from selected cemas countries. *Journal of economics and Finance*, 40:199–210.
- Gardner, J. (2022). Two-stage differences in differences. *arXiv preprint arXiv:2207.05943*.
- Gavazza, A. and Lizzeri, A. (2009). Transparency and economic policy. *The Review of Economic Studies*, 76(3):1023–1048.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.
- Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political analysis*, 20(1):25–46.
- Heise, L., Greene, M. E., Opper, N., Stavropoulou, M., Harper, C., Nascimento, M., Zewdie, D., Darmstadt, G. L., Greene, M. E., Hawkes, S., et al. (2019). Gender inequality and restrictive gender norms: framing the challenges to health. *The Lancet*, 393(10189):2440–2454.
- Hory, M.-P. (2016). Fiscal multipliers in emerging market economies: can we learn something from advanced economies? *International Economics*, 146:59–84.
- Jafarov, E. and Gunnarsson, V. (2008). Government spending on health care and education in croatia: efficiency and reform options.
- Jung, S.-M. (2022). Determinants of gender budgeting practices: Evidence from municipal governments in south korea. *Public Performance & Management Review*, pages 1–30.
- Kang, J., Yu, C., Xue, R., Yang, D., and Shan, Y. (2022). Can regional integration narrow city-level energy efficiency gap in china? *Energy Policy*, 163:112820.
- Kleven, H. and Landais, C. (2017). Gender inequality and economic development: fertility, education and norms. *Economica*, 84(334):180–209.
- Kumbhakar, S., C., Wang, H.-J., and Horncastle, A. P. A. (February 2015.). Practitioner’s guide to stochastic frontier analysis using stata. *Cambridge University Press*,.

- Montes, G. C., Bastos, J. C. A., and de Oliveira, A. J. (2019). Fiscal transparency, government effectiveness and government spending efficiency: Some international evidence based on panel data approach. *Economic Modelling*, 79:211–225.
- Neuenkirch, M. and Neumeier, F. (2016). The impact of us sanctions on poverty. *Journal of Development Economics*, 121:110–119.
- Nguyen, B. H., Sickles, R. C., and Zelenyuk, V. (June 2021). Efficiency analysis with stochastic frontier models using popular statistical softwares. *International Society for Efficiency and Productivity Analysis*.
- Okojie, C. E. (1994). Gender inequalities of health in the third world. *Social science & medicine*, 39(9):1237–1247.
- Olanubi, S. O. and Olanubi, O. E. (2023). Political short termism and government spending efficiency in sub-saharan africa. In *Handbook on Public Sector Efficiency*, pages 430–442. Edward Elgar Publishing.
- Ouertani, M. N., Naifar, N., and Ben Haddad, H. (2018). Assessing government spending efficiency and explaining inefficiency scores: Dea-bootstrap analysis in the case of saudi arabia. *Cogent Economics & Finance*, 6(1):1493666.
- Polzer, T., Nolte, I. M., and Seiwald, J. (2021). Gender budgeting in public financial management: a literature review and research agenda. *International Review of Administrative Sciences*, page 00208523211031796.
- Puig-Barrachina, V., Ruiz, M. E., García-Calvente, M. d. M., Malmusi, D., Sánchez, E., Campubí, L., Muntaner, C., Cortès-Franch, I., Artazcoz, L., and Borrell, C. (2017). How to resist austerity: The case of the gender budgeting strategy in andalusia. *Gender, Work & Organization*, 24(1):34–55.
- Quinn, S. (2016). Europe: A survey of gender budgeting efforts.
- Rambachan, A. and Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, page rdad018.

- Rayp, G. and Van De Sijpe, N. (2007). Measuring and explaining government efficiency in developing countries. *The Journal of Development Studies*, 43(2):360–381.
- Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3):305–22.
- Roth, J., Sant’Anna, P. H., Bilinski, A., and Poe, J. (2023). What’s trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*.
- Salim, A. (2016). Gender responsive budgeting: State and civil society initiatives at national and sub-national levels in indonesia. *Gender Responsive and Participatory Budgeting: Imperatives for Equitable Public Expenditure*, pages 179–193.
- Sen, G. and Östlin, P. (2008). Gender inequity in health: why it exists and how we can change it.
- Shen, G. and Chen, B. (2017). Zombie firms and over-capacity in chinese manufacturing. *China Economic Review*, 44:327–342.
- Sibiano, P. and Agasisti, T. (2013). Efficiency and heterogeneity of public spending in education among italian regions. *Journal of Public Affairs*, 13(1):12–22.
- Steccolini, I. (2019). New development: Gender (responsive) budgeting—a reflection on critical issues and future challenges. *Public Money & Management*, 39(5):379–383.
- Stotsky, M. J. G. and Zaman, M. A. (2016). *The influence of gender budgeting in Indian states on gender inequality and fiscal spending*. International Monetary Fund.
- Svaleryd, H. (2009). Women’s representation and public spending. *European Journal of Political Economy*, 25(2):186–198.
- Tüblicke, S. (2022). Entropy balancing for continuous treatments. *Journal of Econometric Methods*, 11(1):71–89.
- Verhoeven, M., Gunnarsson, V., and Carcillo, S. (2007). Education and health in g7 countries: Achieving better outcomes with less spending.

Wooldridge, J. M. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Available at SSRN 3906345*.

Zhang, S. and Wang, L. (2021). The influence of government transparency on governance efficiency in information age: the environmental governance behavior of guangdong, china. *Journal of Enterprise Information Management*, 34(1):446–459.

Zhao, Q. and Percival, D. (2017). Entropy balancing is doubly robust. *Journal of Causal Inference*, 5(1).

7) Appendix

The next graph summarizes the common support for the propensity score used for the matching.

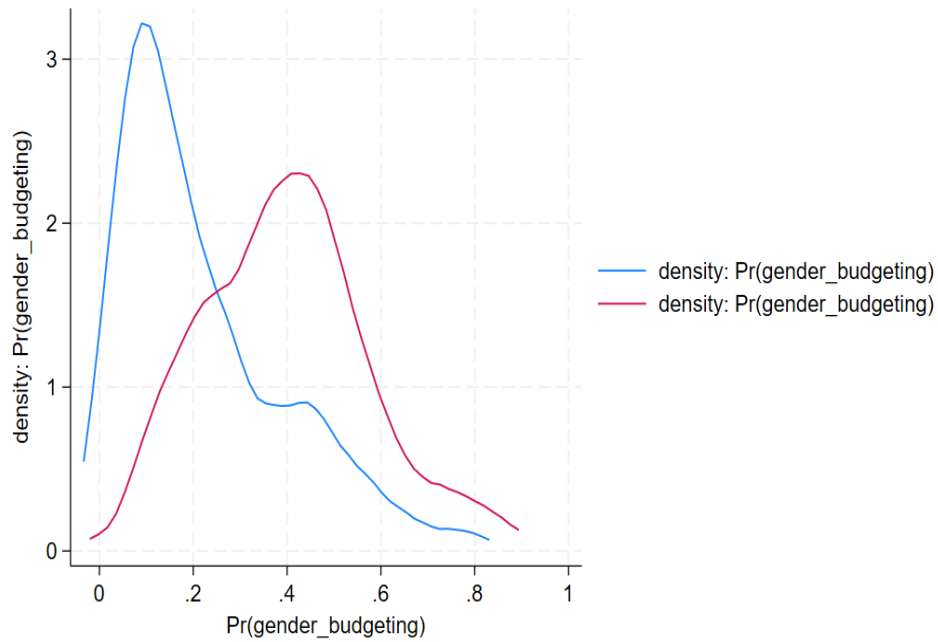


Figure 9: Propensity score common support

The next table summarises the bands of confidence set for Honest DiD from [Rambachan and Roth \(2023\)](#)

Table 14: Honest DiD [Rambachan and Roth \(2023\)](#)

Mbar	Lower band	Upper band
Original	-0.007	0.053
0.5	-0.103	0.107
1.0	-0.205	0.208
1.5	-0.307	0.309
2.0	-0.309	0.309

Table 15: Entropy balancing full results

VARIABLES	(1) efficiency	(2) t+1	(3) t+2	(4) t+3	(5) t+4
gender_budgeting	0.014** (0.006)	0.017*** (0.005)	0.021*** (0.006)	0.025*** (0.006)	0.031*** (0.007)
lurban	-0.001*** (0.000)	-0.000** (0.000)	-0.000** (0.000)	-0.001** (0.000)	-0.001*** (0.000)
lloggdppc	0.022*** (0.001)	0.021*** (0.001)	0.019*** (0.001)	0.014*** (0.001)	0.011*** (0.002)
lautonomy	0.000 (0.000)	0.000*** (0.000)	0.000*** (0.000)	-0.000 (0.000)	-0.000** (0.000)
llocal_wip	-0.003*** (0.001)	-0.005*** (0.001)	-0.003*** (0.001)	-0.002*** (0.001)	-0.001 (0.001)
winner*GB	-0.000* (0.000)	-0.000 (0.000)	-0.000** (0.000)	-0.000** (0.000)	-0.000** (0.000)
winner	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.001*** (0.000)
Constant	99.718*** (0.032)	99.834*** (0.053)	99.746*** (0.033)	99.798*** (0.037)	99.798*** (0.038)
Observations	477	451	429	405	383
R-squared	0.643	0.643	0.626	0.562	0.634

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 16: Entropy balancing full results

	(1)	(2)	(3)	(4)	(5)
VARIABLES	eff_trans	eff_trans	eff_trans	eff_trans	eff_trans
gender_budgeting	0.015*** (0.006)	0.016*** (0.006)	0.023*** (0.006)	0.023*** (0.006)	0.018*** (0.006)
lurban	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
lloggdppc	0.022*** (0.001)	0.022*** (0.004)	0.025*** (0.003)	0.021*** (0.005)	0.011** (0.005)
lautonomy	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000** (0.000)
llocal_wip	-0.002*** (0.001)	-0.002*** (0.001)	-0.003*** (0.001)	-0.003*** (0.001)	-0.002*** (0.001)
essai	-0.000* (0.000)	-0.000* (0.000)	-0.000*** (0.000)	-0.000** (0.000)	-0.000** (0.000)
winner	0.000 (0.000)	0.000 (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
trend		-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.001* (0.001)
llogpop			0.028*** (0.005)	0.028*** (0.005)	0.025*** (0.005)
lAgri_GDP				0.000 (0.000)	-0.000 (0.000)
lBanking_GDP					0.003*** (0.001)
Constant	99.708*** (0.034)	99.704*** (0.047)	99.578*** (0.051)	99.612*** (0.059)	99.687*** (0.058)
Observations	477	477	477	477	477
R-squared	0.643	0.643 ₄₃	0.664	0.666	0.688

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

7.1) Results for Battese and Coelli, 1995

Battese and Coelli (1995) propose a model that relaxes this assumption and suggests that the determinants of inefficiency can be expressed as a linear function of a set of explanatory variables that reflect the inherent characteristics of each state. Therefore, the model of ? enables estimation of the efficiency for each state and of the factors that explain the efficiency differences between states in a single-stage estimation procedure. However, the model does not allow us to divide the efficiency score between structural (long run) and conjunctural (short run) ones. By using another estimation method to determine the efficiency score, we want to check If our results are driven by a bias in the efficiency score estimation or not.

Table 17: Diff in Diff results for Battese and Coelli, 1995

	Callaway	Callaway	Callaway	Callaway
ATT	1.000***	1.106***	1.152***	1.155***
	(4.40)	(4.10)	(3.62)	(3.68)
Observations	419	401	398	388

t statistics in parentheses

Source: Author

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

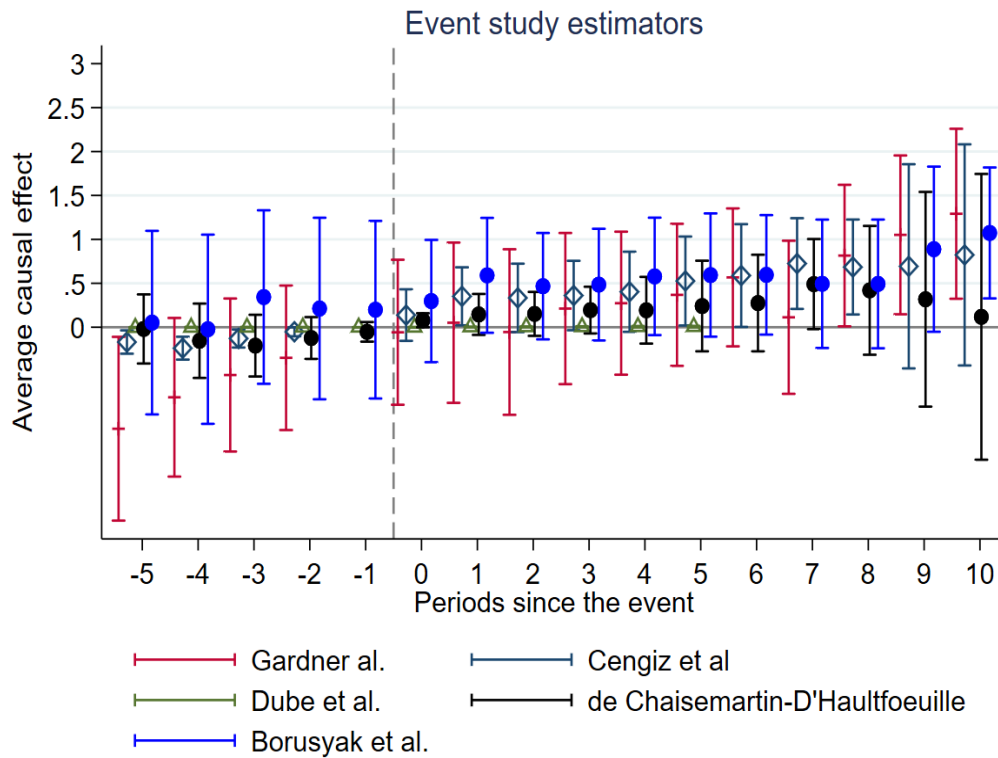


Figure 10: Alternative Diff in Diff Results for Battese et Coelli, 1995

The next tables summarise the results obtained by using the entropy balancing for the efficiency score estimated by Battese and Coelli (1995). The full tables are displayed just below.

Table 18: Entropy balancing results for Battese and Coelli.

VARIABLES	(1) efficiency	(2) t+1	(3) t+2	(4) t+3	(5) t+4
gender_budgeting	0.468*** (0.114)	0.599*** (0.117)	0.635*** (0.121)	0.836*** (0.130)	0.882*** (0.145)
lurban	-0.020* (0.012)	-0.023** (0.012)	-0.028** (0.012)	-0.035*** (0.012)	-0.040*** (0.013)
lloggdppc	1.363*** (0.071)	1.318*** (0.072)	1.177*** (0.076)	0.979*** (0.082)	0.878*** (0.089)
lautonomy	-0.005 (0.006)	0.012* (0.007)	0.010 (0.007)	-0.003 (0.008)	-0.016 (0.010)
llocal_wip	-0.214*** (0.036)	-0.241*** (0.035)	-0.197*** (0.027)	-0.242*** (0.028)	-0.239*** (0.028)
Constant	93.460*** (1.951)	94.389*** (1.932)	94.383*** (1.697)	99.991*** (1.777)	102.413*** (1.863)
Observations	478	456	434	410	387
R-squared	0.839	0.843	0.847	0.827	0.814

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 19: Entropy balancing results for Battese and Coelli.

VARIABLES	(1) efficiency	(2) efficiency	(3) efficiency	(4) efficiency	(5) efficiency
gender_budgeting	0.468*** (0.114)	0.417*** (0.113)	0.532*** (0.121)	0.544*** (0.121)	0.386*** (0.125)
lurban	-0.020* (0.012)	-0.027** (0.012)	-0.028** (0.012)	-0.029** (0.012)	-0.036*** (0.011)
lloggdppc	1.363*** (0.071)	0.536** (0.207)	0.571*** (0.207)	0.304 (0.268)	-0.130 (0.284)
lautonomy	-0.005 (0.006)	0.004 (0.007)	-0.000 (0.007)	-0.002 (0.007)	-0.011 (0.007)
llocal_wip	-0.214*** (0.036)	-0.233*** (0.035)	-0.244*** (0.035)	-0.249*** (0.035)	-0.224*** (0.035)
trend		0.092*** (0.022)	0.083*** (0.022)	0.111*** (0.028)	0.158*** (0.030)
llogpop			0.719** (0.294)	0.761** (0.294)	0.586** (0.292)
lAgri_GDP				0.005 (0.003)	-0.001 (0.004)
lBanking_GDP					0.129*** (0.031)
Constant	93.460*** (1.951)	101.054*** (2.624)	98.639*** (2.789)	100.948*** (3.153)	104.546*** (3.221)
Observations	478	478	478	478	478
R-squared	0.839	0.846	0.848	0.848	0.854

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 20: Simultaneous equations

VARIABLES	(1)	(2)
	eff_trans	bias index
lbiais index	-1.048*** (0.359)	
lurban	-0.001** (0.000)	0.002*** (0.000)
llogdppc	0.032*** (0.011)	-0.001 (0.006)
lautonomy	0.000 (0.000)	0.002*** (0.000)
llocal_wip	-0.011*** (0.003)	-0.011*** (0.003)
trend	0.010*** (0.001)	
gender_budgeting		-0.017** (0.008)
Constant	99.527*** (0.151)	0.481*** (0.138)
Observations	345	345
R-squared	0.470	0.546

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1