Gender Budgeting Adoption in Indian States: Unraveling Fiscal Impacts on Health and Education Expenditures

By Cheick CAMARA[1]

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

Abstract

This paper delves into the evaluation of the repercussions of gender budgeting adoption on fiscal policies of Indian states, with a specific focus on health and education expenditures spanning from 1991 to 2020. Leveraging a Staggered Difference-in-Differences (DiD) framework as the primary empirical methodology, with Entropy Balancing employed as a robustness test, the findings reveal that Indian states embracing gender budgeting exhibit increased spending on health and education compared to their counterparts. Notably, the impact is more pronounced and robust for education expenditures, recognized as a pivotal tool for women's empowerment. The study identifies a potential transmission channel, attributing the augmented spending to increased transfers from central governments, particularly through Centrally Sponsored Schemes (CSS). Policy implications underscore that the timing of adoption matters, with early adopters experiencing more pronounced effects than later ones and non-adopters. The analysis underscores the importance of considering the economic context and local particularities during reform implementation to fully assess its effectiveness. Adopters are advised to contextualize their policy decisions, taking into account the underlying policy context for optimal outcomes.

JEL classifications*: H1; H7; J16; R5*

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

List of Figures

List of Tables

Contents

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 analyzing 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].

Gender Budgeting (GB) recognizes that government budgets impact men and women differently. Policy decisions made by governments can serve to either promote gender equality or reinforce existing inequalities. Gender Budgeting uses fiscal policies and Public Financial Management (PFM) tools to promote gender equality. It incorporates a gender lens into the budget process to ensure that governments are acutely aware of the impact of their choices on gender outcomes. Gender Budgeting is not just about funding explicit gender equality initiatives. It also entails analyzing fiscal policies and budgetary decisions to understand their impact – intended and unintended – on gender equality and using this information to design and implement more effective gender policies. In 1995, based on the discussions at the UN Fourth World Conference on Women held in Beijing, the UN Women recommended that governments systematically review public sector expenditures and adjust their budgets to ensure gender equality in access to expenditures. Since then, gender budgeting has become internationally recognized as a strategy

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

for strengthening gender equality.

Gender budgeting has also been implemented at the subnational level in many developing countries like Indonesia (Salim (2016)) and India. Regional and local governments' proximity to people's everyday lives means there is potential to respond more directly to women's and men's needs when it comes to public policy and service delivery. At these levels, there is great potential to use participatory gender budgeting approaches involving the local population.

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) [2]. 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 or to reach Sustainable Development Goals which must be a common objective for Indian States. The agreement about the power-sharing between national and state 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 can legislate any aspect of education from the primary to

[2] In 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

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)).

This paper explores the effect of gender budgeting efforts in Indian states by focusing on the effect of gender budgeting on the composition of public spending at the state-wise level. Although gender budgeting efforts, in the international context, now date back several decades, there has been little effort to assess the results of these efforts in a quantitative manner (Stotsky and Zaman (2016)). This study thus adds an important dimension to research on gender budgeting by exploring gender budgeting effects from a quantitative point of view, and by exploring the heterogenous effects according to the adoption waves.

In addition to its effects on gender inequalities, Gender Budgeting can also affect fiscal policies. 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 and education public policies for example. The graph[3] 1 summarize the Gender Budgeting framework and how it is included in the budgetary process.

[3]https://blog-pfm.imf.org/en/pfmblog/2021/02/sub-saharan-africa-course-on-gender-budgeting

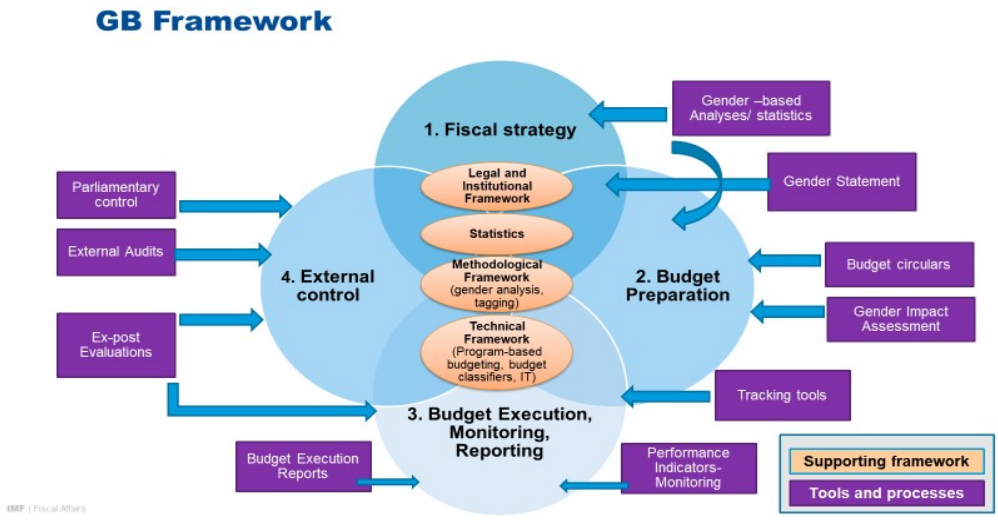
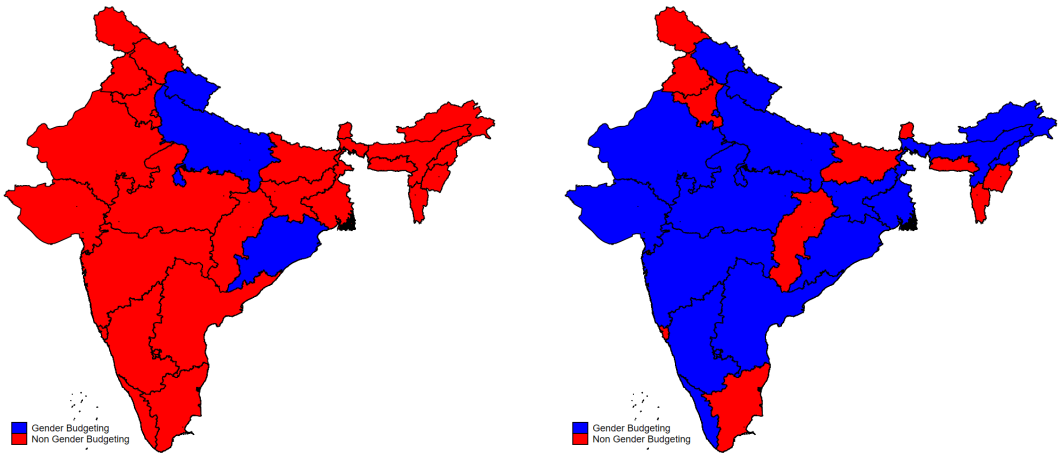


Figure 1: Gender Budgeting framework (Source: IMF PFM Blog)

The obligation to release a statement about the policy objectives improves the transparency 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. 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. The existence of common templates and objectives to follow could impose a constraint that will ensure that States that have adopted gender budgeting will follow the process and rules which could lead to a public statement of their objectives, and greater transparency in their methods to reach these objectives. 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.



(a) Gender budgeting adoption in 2005 (b) Gender budgeting adoption in 2018

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 heterogeneous effects of its adoption. This paper is the first that wants to assess these potential heterogenous effects, to know if the context of the adoption matters or not and if the effects are the same for all the cohorts.

Using state-level panel data, with staggered Differences in Difference and entropy balancing as estimation tools, I find that education and health share in public spending improved significantly (more for education) in gender budgeting states compared to states that did not put in place gender budgeting. The results also suggest that the effects are different according to the adoption waves, so the context of the adoption period matters. The results remain positive and significant over time regardless of the estimation tool.

The rest of the article is structured as follows: section 2 presents some stylized facts about the share of health and education spending in Indian States. The methodology, with the data, identification strategy, and the analysis of the parallel trend assumption are presented in section 3. Section 4 presents the first results of gender budgeting adoption with the Difference in Difference (DiD) staggered estimators of Callaway and Sant'Anna (2021). The results of the additional

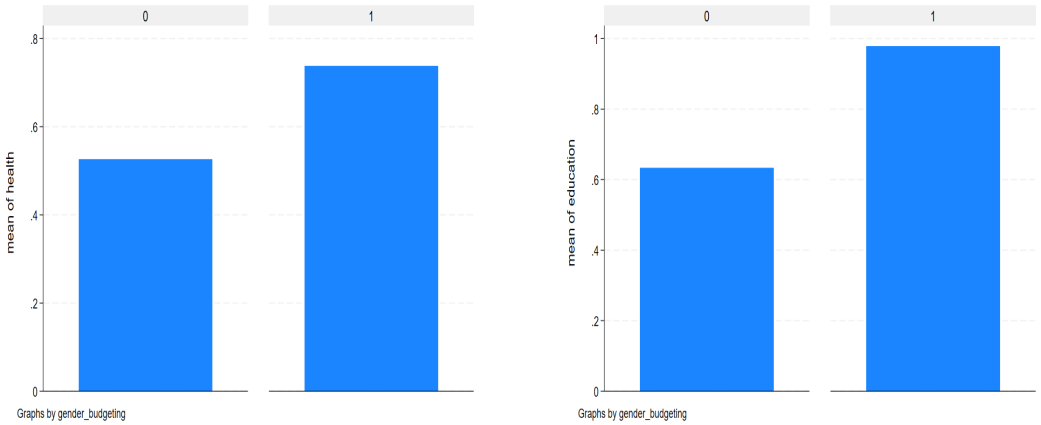
robustness analyses with alternative DiD estimators, entropy balancing, and other robustness checks are presented in section 5. Section 6 tries to assess the potential transmission channel. Finally, section 6 concludes.

2) Stylized facts

The graph 3 highlights a comparison between the average health and education expenditures share for the treated (1) and untreated units (0). This seems to suggest that states that have adopted gender budgeting spend more on health and education than those that have not adopted it. However, this correlation means nothing in terms of causes and consequences because a correlation does not necessarily imply causality. This result seems to confirm the intuition and provide avenues to explore for further analysis.

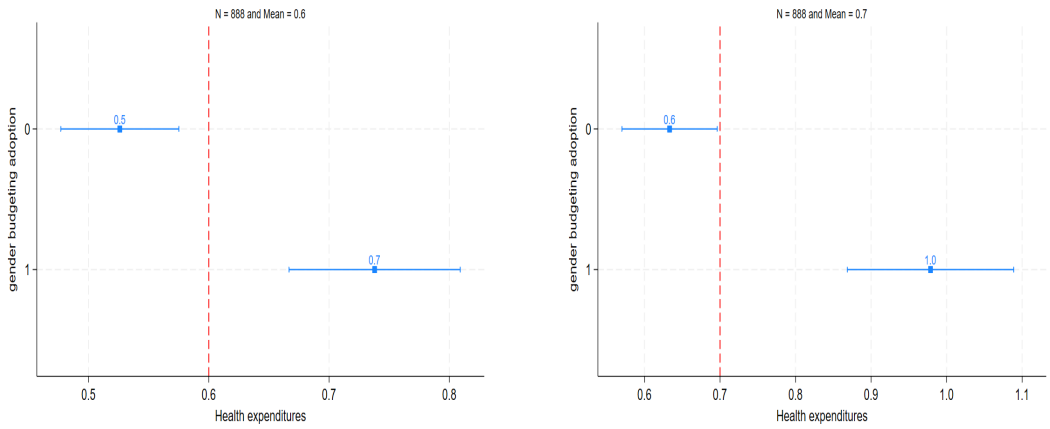
The graph 4 summarizes the average efficiency score for untreated (0) and treated units (1) by comparison to the global average of the sample (the vertical red line). The difference between both means suggests that the treated units seem to spend more for the two items (health and education) than the untreated ones. In addition, their average expenditures in these items are more important than the average of the global sample.

All these elements tend to suggest that States that have adopted gender budgeting spend significantly more than the others. However, I can conclude that this difference is due to the gender budgeting adoption. The difference can be due to a simple correlation between the variables or to the fact that States which dedicated a greater share of their expenditures to these items got more incentives to adopt gender budgeting. This is why I have to go further in the analysis.



(a) Average expenditures share for health (b) Average expenditures share for education

Figure 3: Stylized facts



(a) stylized facts for Health (b) stylized facts for Education

Figure 4: Stylized facts

3) Methodology

3.1) 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	125
Untreated	444

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

Fiscal autonomy can affect spending size and composition and is associated with a smaller public sector size at the local level for Fiva (2006). Siwińska-Gorzelak et al. (2020) shows a U-shaped relationship between the ratio of own local taxes and the share of capital expenditures and a negative relationship between the former ratio and the share of education spending for Polish municipalities.

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. To avoid this, more urbanization may call for more expenditure centralization by attracting the rural population towards big centers, favoring a certain concentration of public expenditures(Sacchi and Salotti (2016)).

GDP per capita affects the accountability of rulers and their decisions to adopt or not the 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
health	0.584	0.623	0	5.849	888
education	0.728	0.845	0	5.517	888
gender_budgeting	0.22	0.414	0	1	1325
lautonomy	48.938	25.713	5.466	100	887
lloggdppc	10.287	1.061	7.886	12.832	942
lurban	33.568	19.098	7.98	99.900	870
trend	16.815	9.352	1	33	1325
llocal_wip	48.53	1.878	42.39	54.87	902
fiscal_rule	0.381	0.486	0	1	1325
lagri	26.575	20.109	0.052	130.834	930
logpop	20.847	0.125	20.608	21.025	843

3.2) Identification strategy

The identification method used is a Difference in Difference (DiD) strategy, using a comprehensive panel dataset. I focus on the share of education and health public spending among the total expenditures for each state and each year through the period 1991-2020. 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, using several years of data makes our approach closer to two-way fixed effects (TWFE) linear regression. Recent methodological papers characterize 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, I try to capture longer-term effects and check if there is an increasing advantage of early adoption. I also acknowledge a group of states that have adopted gender budgeting after the first wave, which might slightly perturbate the control group as some units become treated. To address this, I suggest additional estimations where I explicitly account for the two types of treatment. In technical terms, I estimate the following equation in which $y_i t$ is the outcome variable, i.e., efficiency score for state i in year t = 1, ..., T

$$Y_{it} = \alpha + \beta^W {D_{it}}^W + \rho X_{it} + \theta_i + \gamma_t + \epsilon_{it} \tag{1}$$

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

To slightly enhance the DiD setup, I 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), I must estimate the propensity score. However, because I have multiple treatment dates for multiple groups, there is a unique propensity score for every group. However, I 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. I am using a pool of units as our comparison group who never are treated during the duration of the panel. Or I 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, I am going to mobilize 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.

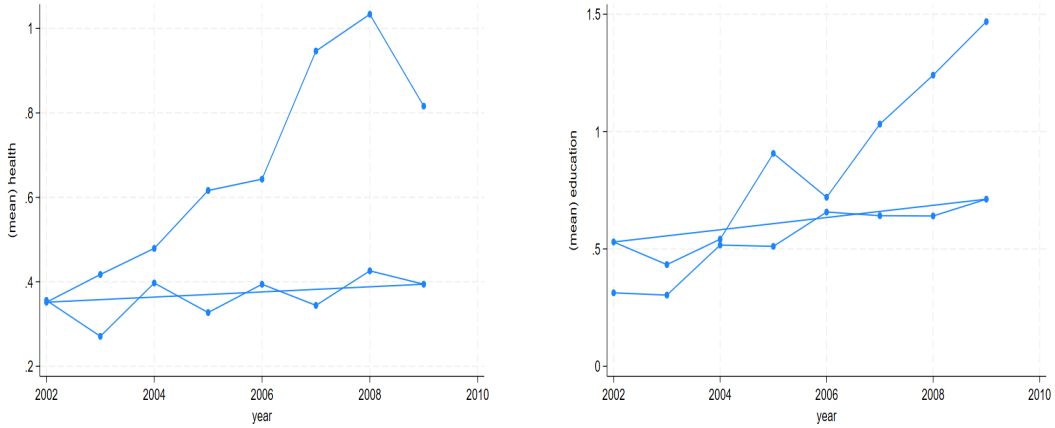
For example, if states with the greatest GDP per capita are the ones that adopted gender budgeting first and, at the same time, are the ones that benefit from efficiency score (internal validity issue) or stand to benefit most from gender budgeting because their important GDP per capita can mean greatest interest for central government to rule this state. So, it can increase the discretionary transfers that are targeted at specific purposes (external validity issue), and then I 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), I 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, I 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, I 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. I 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.

3.3) Parallel trend assumption

The graphs 5 and 6 highlight the evolution of health and education spending 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

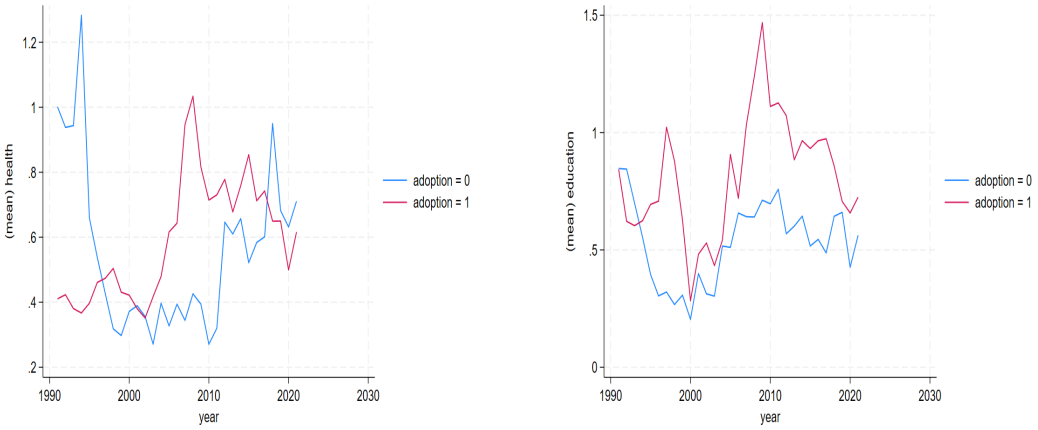
that untreated units could provide a reasonable counterfactual, particularly if they most closely resembled the treated ones. It also allows us to reduce the concerns about the fact that our results could be driven by pre-trend effects which would bias our results. The graph 6a and 6b provide an overview of efficiency score evolution through the study period and a graphic estimation of the pre-trend assumption over this period. Graph 5a and 5b provide 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, I 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 that have adopted gender budgeting at different points in time.



(a) Health (b) Education

Figure 5: Parrallel trend Parallel trend assumption until 2010



(a) Health (b) Education

Figure 6: Parrallel trend assumption

The following graph (7) 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, they impose restrictions on how different the post-treatment violations of parallel trends can be from the pre-treatment differences in trends ("pre-trends"). They recommend that researchers use their 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 7 shows robust confidence sets for the treatment effect, using different values of Mbar [4]. The figure shows that if I impose Mbar<1, meaning that I restrict the post-treatment violations of parallel trends to be no larger than the maximal pre-treatment violation of parallel trends, then I obtain a robust confidence set for the causal effect on the expenditures share. 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 expenditures share.

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

[4]Mbar is a degree of smoothness, or how much I 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 I am willing to bound the magnitude of economic shocks by the max in the pre-treatment period, I will thus typically obtain wider confidence sets for parameters involving later periods. As suggested by Rambachan and Roth (2023) the table 7 available in the appendix summarizes the different bands of confidence interval according to the Mbar values.

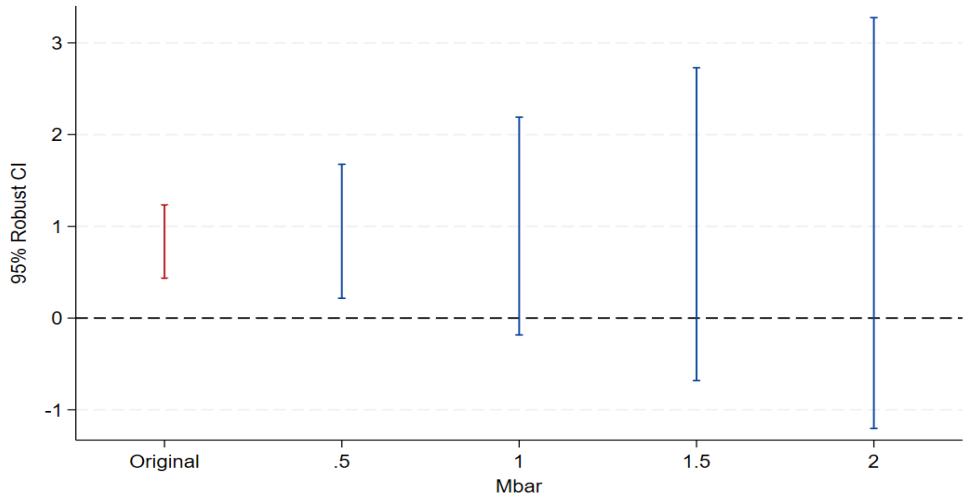


Figure 7: Parallel trend with honest DiD (Rambachan and Roth (2023))

4) Results

The results are available in the next table. These results suggest a positive effect of gender budgeting adoption on the share of health and education spending for the Indian states. This result seems coherent with the fact that education and health are some important concerns for women and gender equality. As explained by Mbodji (2023), public expenditure on education has a positive effect on gender equality in education at the primary, secondary, and tertiary levels.

Table 3: Diff in Diff results

	Health	Education
ATT	0.578***	0.718***
	(4.47)	(3.49)
N	542	542

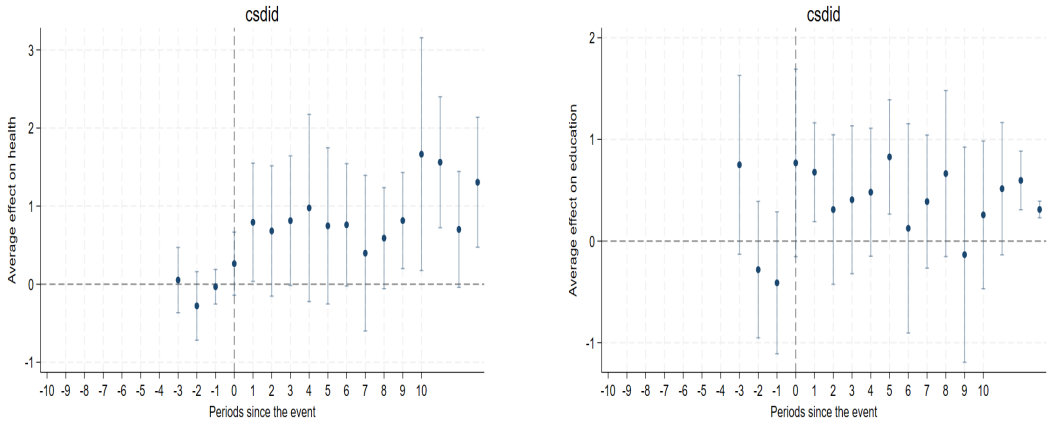
t statistics in parentheses

Source: Author calculation

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

In addition, I compute an event-study graph to assess the effects of gender budgeting adoption over time and an effect by cohort. Both results suggest an overall positive effect of gender budgeting adoption. Indeed, the DiD event study clearly shows a positive effect of gender budgeting adoption on the health and education share in total spending. However, this increase is not continuous over time. The effort to increase the fiscal space for health and education spending doesn't seem to drive a continuous and progressive increase in the share of these expenses. So, the States that adopted gender budgeting seem to allocate a greater share of their expenditures to health and education than those that have not adopted it. It's also important to notice that this positive effect is more significant for education spending.

The following graphs summarise these event-study results



(a) Event study for Health (b) Event-study for Education

Figure 8: Event-study Results

4.1) Cohort effects

One of the most popular questions that arises in DiD setups with multiple periods concerns treatment effect dynamics: How does the effect of participating in the treatment vary with the length of exposure to the treatment? For instance, do average treatment effects increase/decrease with elapsed treatment time? Indeed, answering this type of question is often the main motivation for using the event study regression. Callaway and Sant'Anna (2021) also proposes different aggregation schemes that can be used to highlight treatment effect heterogeneity across different dimensions as well as to summarize the overall effect of participating in the treatment.

consider a natural generalization of the AT T that is suitable to setups with multiple treatment groups and multiple periods. More precisely, I use the average treatment effect for units who are members of a particular group g at a particular period t, denoted by

$ATT(g,t) = \mathrm{E}[Y_t(g) - Y_t(0) \mid \mathrm{Gg} = 1]$

Callaway and Sant'Anna (2021) allows us to identify the ATT (g, t)'s by restricting treatment anticipation behavior and imposing a conditional parallel trends assumption. In many applications, the AT T (g, t)'s can be the ultimate causal parameters of interest. They can be used to highlight treatment effect heterogeneity across different groups g, at different points in time t, and across different lengths of treatment exposure, $e = t - g$.

The table 4 summarizes the results for the different adoption waves. It seems to confirm the intuition and hypothesis that the effects of gender budgeting implementation are heterogeneous and change according to the adoption waves. The effect seems to be on average positive and significant, with a more important effect for the early adopters for both outcome variables. These results can suggest a self-enforcement process of the gender budgeting process which will give more experience to the early adopters and lead to more important efforts over time. However, if I check more accurately, I will find a negative effect for both outcomes for the 2017 cohort and a negative effect for the education expenditures for the 2011 cohort.

The results of the 2017 cohorts can be explained by the introduction of the Goods and Services Tax (GST) in India 2017. This reform negatively affects States' revenues and doesn't ensure tax neutrality and tax efforts from States. The reform has an important effect on the tax-sharing institutional arrangement between States and central government. However, it introduces some concerns about the ability of States to remain financially autonomous (Mukherjee (2023) and Van Rompuy (2016)). I assumed that this context led the State member of the cohort to reduce their health and education spending.

For the 2011 cohort, the only member is Rajasthan. For the Rajasthan Gender Budgeting Process, the information is provided neither department-wise nor major head-wise, but Budget Finalization Committee (BFC) unit-wise. I don't know how many BFC units are there in any given department and so nothing can be said about a department with any certainty. The schemes/programs are not given rank as a whole but the plan, non-plan, and CSS parts of the schemes/programs are given ranks (Ahmad (2015)). So, there is no way to say anything about any scheme with some certainty. The basis of the rank given is not explained anywhere. No explanatory notes are given about the GBS methods, meaning, etc. anywhere in the budget document. These particularities can explain the negative results of education spending. In addition, Rajasthan is a rural state, and the largest one in India. I assumed that the state government wants to prioritize health spending through reproductive health to reach its objectives of gender inequalities.

Table 4: Diff in Diff results by cohorts

Health:							
Adoption waves	2005	2006	2007	2008	2011	2016	2017
gender budgeting	1.261***	0.530***	0.173*	-0.112	0.473***	0.0698***	-0.0207***
N	542	542	542	542	542	542	542
Education:							
Adoption waves	2005	2006	2007	2008	2011	2016	2017
gender budgeting	1.025***	0.377***	0.383***	0.373***	-0.0199***	0.256***	-0.179***
N	542	542	542	542	542	542	542

5) Robustness Check

5.1) Alternative DiD estimator

As a robustness check, I use Wooldridge (2021) DiD framework which has been built to be suitable for staggered adoption. Wooldridge (2021) establishes the equivalence between the two-way fixed effects (TWFE) estimator and an estimator obtained from a pooled ordinary least square regression that includes unit-specific time averages and time-period specific cross-sectional averages, which he calls the two-way Mundlak (TWM) regression. The approach allows considerable heterogeneity in treatment effects across treatment intensity, calendar time, and covariates. The equivalence implies that standard strategies for heterogeneous trends are available to relax the common trends assumption. He concludes that there is nothing inherently wrong with using TWFE in situations such as staggered interventions – a point that is also clear from Sun and Abraham (2021). Because I know that TWFE is consistent for unbalanced panels (as the cross-sectional sample size grows with T fixed), even when selection is correlated with additive, unobserved heterogeneity, it has advantages over other estimators that include time-constant cohort indicators and time effects.

The point for him is not to conclude that other recent approaches – such De Chaisemartin and

d'Haultfoeuille (2020), Callaway and Sant'Anna (2021), Borusyak et al. (2022), among others – are not valuable and cannot improve over flexible TWFE methods. However, he is recommending not abandoning simple regression approaches because they can identify the treatment effects of interest very generally and can be made very flexible.

According to Wooldridge (2021), another nice feature of a flexible TWFE approach is that it is easily extended to allow for heterogeneous trends, which can help when one suspects the common trends assumption is violated.

Callaway (2023) concludes that the regression approaches in Wooldridge (2021) may be particularly appealing in applications where the researcher is primarily interested in estimating and conducting inference on the group-time average treatment effects themselves. Because I want this paper to assess the cohorts' effects, I consider that this approach could be interesting as a robustness check.

The results for the global average effect and the event study are available below.

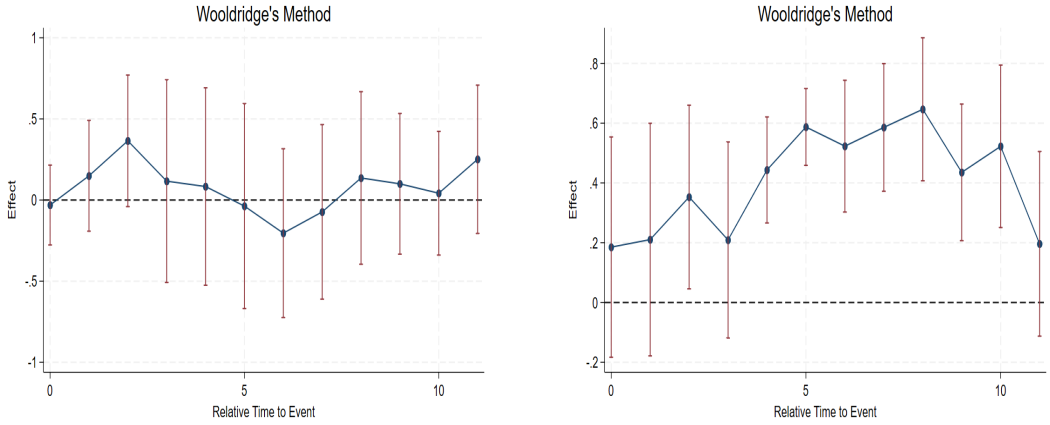
Table 5: Diff in Diff results

	Education	Health
ATT	0.416***	0.0976
	(5.75)	(0.47)
Observations	640	640

t statistics in parentheses

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

The results seem to suggest a positive and significant effect of gender budgeting adoption on education expenditures share while the results for health expenditures are not significant.



(a) Event study for Health (b) Event-study for Education

Figure 9: Event-study Results

5.2) Entropy Balancing

5.2).1 Methodological concept

For the robustness check, I 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 favor, even if they have not adopted gender budgeting. The competition effect can affect the pre-trends and bias the results. For this reason, I 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. I 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:

$$\pi ATT(x) = E[Y(1)|T = 1, X = x] - E[Y(0)|T = 0, X = x] \tag{2}$$

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 assignment 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. 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:

$$Y_{i_t} = \beta_1 T_{it} + \alpha_1 log(GDP\_pc)_{it} + \alpha_2 log(density)_{it} + \alpha_3 X_{it} + \mu_i + \psi_t + \epsilon_{it} \quad (3)$$

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, $\epsilon_i t$ is the usual idiosyncratic error term assumed to be uncorrelated with the treatment.

5.2).2 Correlation issue

Table 10 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]), I 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, I can consider the synthetic group as a perfect counterfactual of the treated group.

Table 6: Balance Statistics

Variable	**treated**	**untreated**	**Difference**	**t statistics**	**p-value**
Unweighted Balance Statistics:					
	(1)	(2)	(3)	(4)	(5)
lautonomy	43.81	49.79	-5.98	2.4375	0.0156
lloggdppc	10.77	9.98	0.79	-11.2868	0.0000
lurban	27.65	29.54	-1.89	1.6248	0.1054
trend	22.76	14.18	8.58	-18.3878	0.0000
llocal_wip	48.46	48.67	-0.21	1.1481	0.2524
fiscal_rule	1.00	0.39	0.61	-26.46481	0.0000
lagri	27.33	31.13	-3.80	2.4178	0.0162
logpop	20.97	20.84	0.13	-19.3786	0.0000
Weighted Balance Statistics:					
	(1)	(2)	(3)	(4)	(5)
lautonomy	43.81	43.80	0.01	0.003	0.9922
lloggdppc	10.77	10.77	0.00	0.058	1.000
lurban	27.65	27.65	0.00	0.015	0.9473
trend	22.76	22.72	0.04	0.080	1.000
llocal_wip	48.46	48.46	0.00	0.776	0.8738
fiscal_rule	1.00	1.00	0.00	0.011	1.000
lagri	27.33	27.33	0.00	0.096	0.9919
logpop	20.97	20.96	0.01	-0.038	1.000

5.2).3 Results

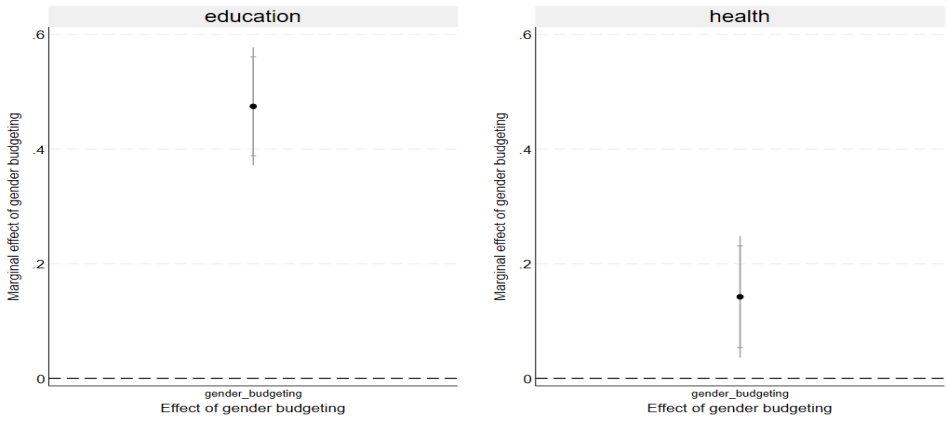


Figure 10: Entropy balancing Results

5.3) Dose Response Function

To ensure the robustness of our results, I also used the Dose Response Model (DRM). Indeed, I wanted to estimate the causal effect of the treatment variable t on an outcome within the observed sample, assuming that treated and untreated units may respond differently both to specific observable confounders (that I collect in a vector X) and to the intensity of the treatment t. In addition to the previous estimations, the DRM allows us to check if the effects of the treatment vary according to the intensity of the treatment (measured by the length of exposure) or not.

The Dose-Response Model (DRM) employed in this paper is an econometric model for estimating continuous treatments under heterogeneous responses, where selection into treatment may be endogenous and has been developed by Cerulli (2015) and used by Avenyo et al. (2019) and Janzen et al. (2023). A state's decision to adopt Gender Budgeting may not be random and may be influenced by confounders and vice versa.

In our context, to consider the "dose" of the treatment, I use the duration since the first adoption of the treatment. The intuition is that States that have adopted gender budgeting will follow the principles and common templates in an Indian context, and this application could lead to a self-enforcement of local administrations. So, a States that are part of the first adoption waves, will get more experience in the gender budgeting adoption than the members of the next

cohorts. The results are available in the table 8 in the Appendix section.

5.4) Placebo Test

I now examine whether there are confounding factors that could affect the results, which have remained stable so far (especially for education expenditures share). The empirical literature shows that the adoption of an economic policy is generally associated with parallel reforms, making the adoption of gender budgeting a non-random factor. One could therefore imagine that unobservable variables correlated with policy adoption and potentially with the outcome variable could affect the baseline results. While I am aware that the empirical — method used in this study aims to address these types of concerns, I still — strengthen the results by conducting a placebo test on gender budgeting adoption. To do this, I follow Apeti (2023) and Apeti and Edoh (2023) in setting placebo or arbitrary dates for gender budgeting, computed by randomly assigning gender budgeting episodes to countries in our sample after removing the actual adoption years. The main idea behind this test is that if the results are biased by unobservable variables, the placebo — test might also show significant effects. Random treatments within the sample do not affect both education and health expenditures share in total expenditures (Table 9, in Appendix). Therefore, I can rule out the possibility of confounding — factors influencing our results.

6) Transmission channel

To identify the potential transmission channel, I have constructed a ratio of *Centrally Sponsored Schemes* (CSS) on the state revenues and state expenditures. This construction aims to check if the increase in health and education expenditures can be due to an increase in transfers received by each state. Indeed, *Centrally Sponsored Schemes* are some transfers decided by central ministries and spent for some specific purposes such as education and health (which is on a *Concurrent List* between States and Central government). I am not able to collect data about the different schemes and only keep those related to health and education. The variables are summarised just below.

Table 7: Transmission channel

	CSS(% of revenues)	CSS (% of expenditures)
Before adoption	4.08***	4.18***
After adoption	5.82***	6.05***
Non Gender Budgeting	4.54***	4.60***

I try to estimate the potential transmission channels by using the same process as Neuenkirch and Neumeier (2016) I compute the means of the two 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 7 just above. 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, I find that the latter is characterized by a notably greater share of CSS for both measures.

Indeed, before the treatment, the treated units received less CSS in percentage of revenues (4.08% versus 4.54%) and expenditures (4.18% versus 4.60%) than the non-adopter ones, but the situation became different after the adoption for revenues (5.82 vs 4.54) and expenditures (6.05 vs 4.60) The results obtained from a t-test are also highly significative with a p-value lower than 0.01.

However, I can't conclude from this statistical test that gender budgeting adoption reduces the state's autonomy. The increase of CSS received by the States can be due to the wish of the central government to fund some projects decided by state governments to reach their objectives, but it can also be an incentive to adopt gender budgeting and mean for the central government to influence the state's decisions. I can only conclude that an increase in CSS received by the States could be a potential transmission channel to explain the greater share of health and education expenditures for the adopter States.

7) Conclusion

Through this work, I have tried to evaluate the effects that gender budgeting adoption can have on Indian state's fiscal policies, and more precisely on health and education spending for a panel of Indian states over the period 1991-2020. To answer this question, I 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 skepticism about our conclusions. In addition to DiD and entropy balancing, I use a Dose-Response Model and perform some placebo tests to ensure that our results are robust.

Our results suggest that Indian states which adopt gender budgeting tend to spend more on health and education than those that do not. The results are more important and more robust for education expenditures which are identified as an important tool to empower women (Samarakoon and Parinduri (2015)) The potential transmission channel identified in this paper is through an increase in the transfers received from central governments and more precisely Centrally Sponsored Schemes (CSS) as denoted in the table 7.

Gender budgeting adoption by imposing a continuous assessment affects the overall fiscal framework. Gender budgeting is a useful policy tool that has some effects on fiscal policy because fiscal policy and public expenditures are the means used to reach the objectives to reduce 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 affect fiscal policy and expenditure decisions and priorities.

From the point of view of policy implications, this analysis tells us that gender budgeting adoption affects fiscal policies as expected, but the effects are different across the different adoption waves. The early adopters seem to experience a more important effect than the later ones and non-adopters. The implementation of this reform, and all reforms in general must consider the economic context and local particularities of adopters to fully assess the effectiveness of the policy implementation. So, the adopters need to consider their underlying policy context (Polzer et al. (2023)).

In the context of Indian states, the implementation of gender budgeting, particularly through improved central government transfers (CSS) to address gender-specific needs, has the potential

to reshape fiscal dynamics. While this approach seeks to rectify gender disparities and enhance social inclusion, there is a need for careful consideration. The targeted allocation of resources towards CSS on health education or other gender-related issues may impact the fiscal autonomy of states, potentially limiting their traditional decision-making flexibility. Additionally, the lack of comprehensive data to assess additional transmission channels and the potential consequences of such redistributions raises concerns and leads to caution about these effects. The delicate intergovernmental relationships among Indian states may be strained, as the perceived imbalance in resource allocation could pose challenges to the equitable distribution of funds. Recognizing the transformative potential of gender budgeting, further research is imperative to navigate these complexities and ensure that the adoption of gender budgeting at the state level aligns harmoniously with the unique fiscal landscape of Indian states.

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The review of economic studies*, 72(1):1–19.

Ahmad, N. (2015). *Status of Gender Responsive Budgeting in Rajasthan: A Study*. Budget Analysis Rajasthan Centre and Dan Church Aid.

Apeti, A. E. (2023). Household welfare in the digital age: Assessing the effect of mobile money on household consumption volatility in developing countries. *World Development*, 161:106110.

Apeti, A. E. and Edoh, E. D. (2023). Tax revenue and mobile money in developing countries. *Journal of Development Economics*, 161:103014.

Avenyo, E. K., Konte, M., and Mohnen, P. (2019). The employment impact of product innovations in sub-saharan africa: Firm-level evidence. *Research Policy*, 48(9):103806.

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.

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.

Callaway, B. (2023). Difference-in-differences for policy evaluation. *Handbook of Labor, Human Resources and Population Economics*, pages 1–61.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Cerulli, G. (2015). ctreatreg: Command for fitting dose–response models under exogenous and endogenous treatment. *The Stata Journal*, 15(4):1019–1045.

Chakraborty, L. (2016). Asia: A survey of gender budgeting efforts.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.

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.

Fiva, J. H. (2006). New evidence on the effect of fiscal decentralization on the size and composition of government spending. *FinanzArchiv/Public Finance Analysis*, pages 250–280.

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.

Janzen, J. P., Malone, T., Schaefer, K. A., and Scheitrum, D. P. (2023). Political returns to ad hoc farm payments? *Applied Economic Perspectives and Policy*, 45(1):555–578.

Jung, S.-M. (2022). Determinants of gender budgeting practices: Evidence from municipal governments in south korea. *Public Performance & Management Review*, pages 1–30.

Kleven, H. and Landais, C. (2017). Gender inequality and economic development: fertility, education and norms. *Economica*, 84(334):180–209.

Mbodji, Y. C. (2023). Effects of public expenditure on education on gender inequality in education in sub-saharan africa (ssa). *Review of Education*, 11(2):e3398.

Mukherjee, S. (2023). Revenue implications of gst on indian state finances. Technical report.

Neuenkirch, M. and Neumeier, F. (2016). The impact of us sanctions on poverty. *Journal of Development Economics*, 121:110–119.

Okojie, C. E. (1994). Gender inequalities of health in the third world. *Social science & medicine*, 39(9):1237–1247.

Polzer, T., Nolte, I. M., and Seiwald, J. (2023). Gender budgeting in public financial management: a literature review and research agenda. *International Review of Administrative Sciences*, 89(2):450–466.

Puig-Barrachina, V., Ruiz, M. E., García-Calvente, M. d. M., Malmusi, D., Sánchez, E., Camprubí, 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.

Sacchi, A. and Salotti, S. (2016). A comprehensive analysis of expenditure decentralization and of the composition of local public spending. *Regional Studies*, 50(1):93–109.

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.

Samarakoon, S. and Parinduri, R. A. (2015). Does education empower women? evidence from indonesia. *World Development*, 66:428–442.

Sen, G. and Östlin, P. (2008). Gender inequity in health: why it exists and how we can change it.

Siwińska-Gorzelak, J., Bukowska, G., and Wójcik, P. (2020). The impact of revenue autonomy on the composition of local public spending: Evidence from poland. *Local Government Studies*, 46(4):641–665.

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.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Svaleryd, H. (2009). Women's representation and public spending. *European Journal of Political Economy*, 25(2):186–198.

Tübbicke, S. (2022). Entropy balancing for continuous treatments. *Journal of Econometric Methods*, 11(1):71–89.

Van Rompuy, P. (2016). Sub-national tax autonomy and deficits: Empirical results for 27 oecd countries. *Regional studies*, 50(7):1248–1259.

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

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

8) Appendix

8.1) Entropy Balancing

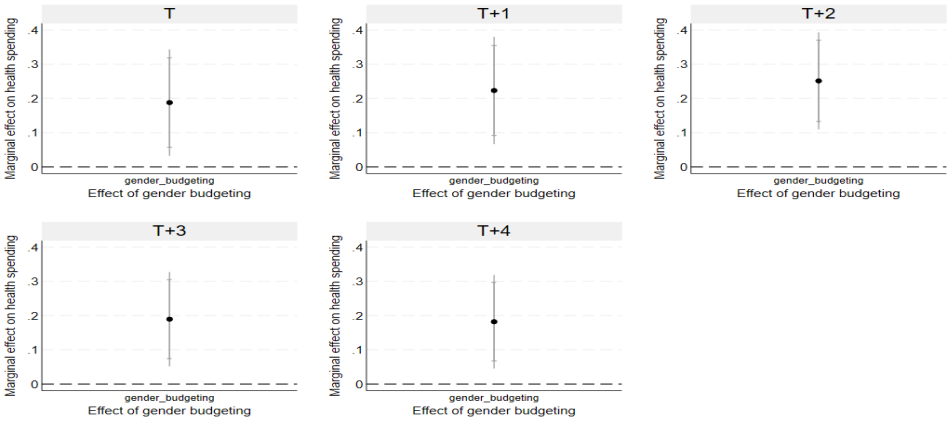


Figure 11: Entropy balancing results for health through time

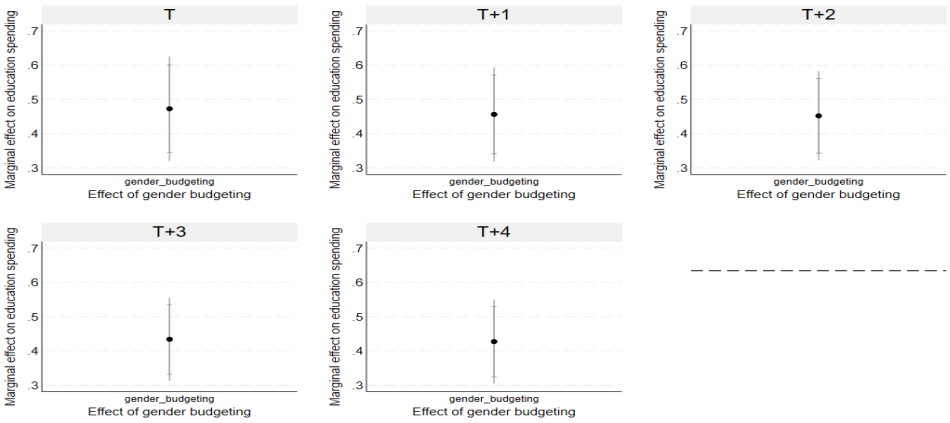
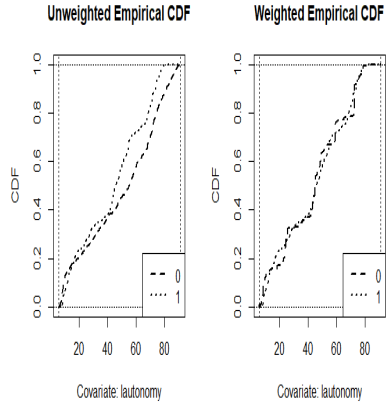
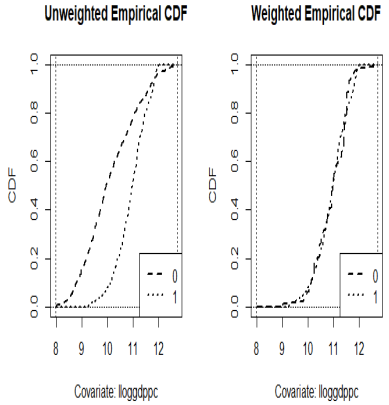


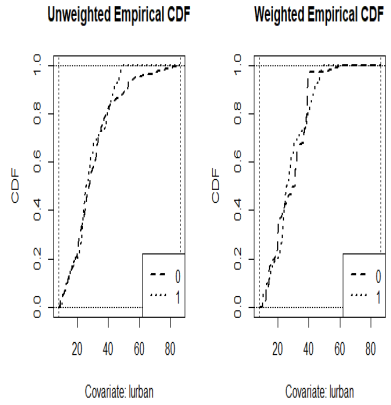
Figure 12: Entropy balancing results for education through the time



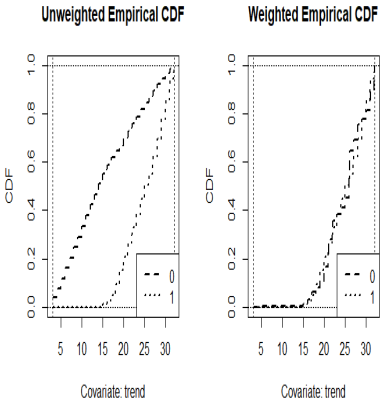
(a) Autonomy reweighted



(b) GDPpc reweighted



(c) Urbanization reweighted



(d) Trend reweighted

Figure 13: Entropy balancing results

8.2) Staggered DiD

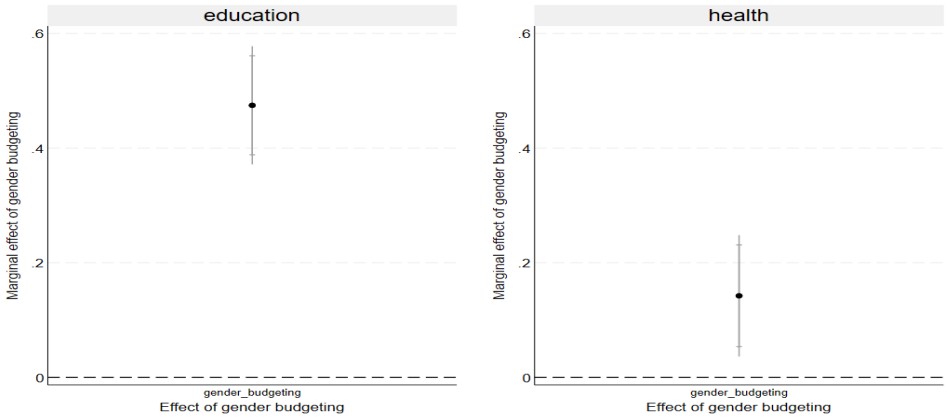


Figure 14: Visual results of Callaway and Sant'Anna

8.3) Dose Response Model

Table 8: Dose Response Model results

	Education	Health
gender_budgeting	0.709**	0.453**
	(2.64)	(2.55)
Observations	569	569
R^2	0.139	0.295

t statistics in parentheses

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

8.4) Placebo Test

Table 9: Placebo test Results with entropy balancing

	Education	Health
placebo	0.00972	-0.0331
	(0.35)	(-1.11)
Observations	640	640

t statistics in parentheses

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