Gender Budgeting and Health Spending Efficiency in Indian States

By Cheick CAMARA[1]

[1] *Université Clermont Auvergne, CNRS, IRD, CERDI, F-63000 Clermont-Ferrand, France, corresponding author: CAMARA Cheick, cheick_amadou.camara@doctorant.uca.fr*

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 multilevel Tobit 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. One key transmission channel emerges as responsible for this positive effect. Adopting gender-responsive budgeting has led to a substantial enhancement in the quality of spending forecasts. The improved accuracy of spending 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

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 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

[2] End 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

[3] 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

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 graph[4] 1 summarize the Gender Budgeting framework and how it is included in the budgetary process.

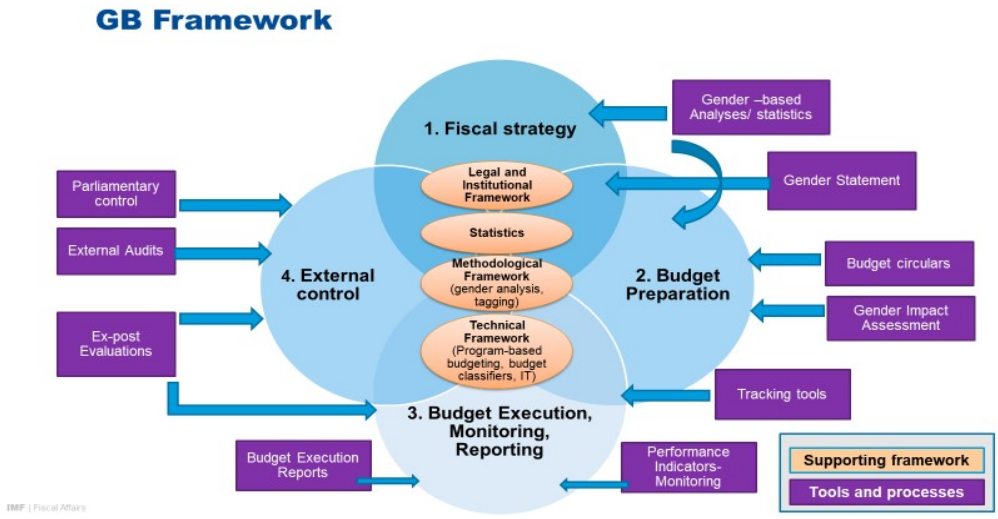


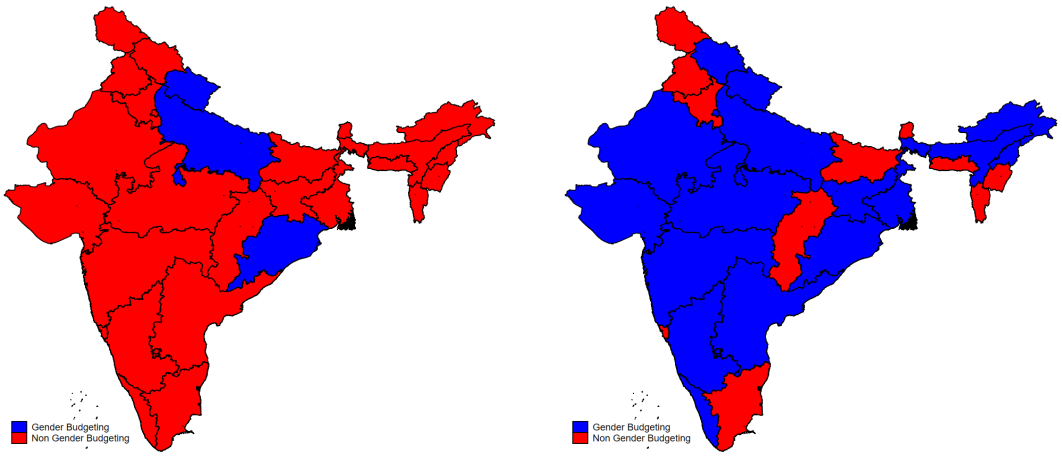
Figure 1: Gender Budgeting framework (Source: IMF PFM Blog)

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.



(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 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 Multilevel Tobit 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 estimation. Section 3 presents the first results of gender budgeting adoption with Multilevel Tobit. The results of the additional robustness analyses and entropy balancing are presented in section 4. Section 5 tries to assess the potential transmission channels. Finally, section 6 concludes.

2) Theoretical framework

The theoretical concept used for this paper is like a principal-agent-based model founded on the agency theory of Meckling and Jensen (1976). The use of the normative agency theory is based on the nature of the relationship between local leaders and their constituents. Due to the implicit delegation of power through voting, the voter is capable of positively or negatively sanctioning the elected official based on the economic situation or actions taken. In this sense, the relationship between the voter and local power is akin to a principal-agent relationship, where the principal delegates a task to the agent in exchange for compensation. Thus, in exchange for political power and associated benefits (compensation, social status, prestige, etc.), the local leader (the agent) must ensure that the implemented public policies align with the expectations

of the population. During elections, each voter negatively sanctions the local government if this is not the case or re-elects it otherwise. The agent's behavior is therefore the result of a trade-off that depends on the terms of the contract linking their compensation (in terms of total income) to their level of effort. The principal (voters) can also compare the results of their agent (local rulers) with those of neighboring jurisdictions (Revelli (2002) and Vermeir and Heyndels (2006)). We will also mobilize the concept of median voter which has already been used in India by Ghosh and Pal (2004).

Furthermore, we consider that the local government derives benefits from their elected position. These benefits can be of a monetary or non-monetary nature (Brennan and Buchanan (1980)). In the latter case, the elected official values access to power, decision-making, popularity, political support, and social status. The entirety of these benefits, which we will generalize under the term "rent," includes a fixed portion, independent of the actions taken, and a contingent portion based on the level of effort exerted. We hypothesize that this second fraction of the rent is comparable to an effort premium, as the elected official will receive political support and activist actions in exchange for the effort made in the field of gender equality. In this sense, the variable part of the rent, denoted as "s," serves as a source of utility for the local government. Over a period, the local government's utility function is then expressed as:

$$U_m = s - \theta e \tag{1}$$

Where e represents the level of effort provided by the State governments and $\theta > 0$, which corresponds to both the average cost and the marginal cost of effort. To simplify the model, we will consider that local elected officials can be divided into two distinct groups based on their level of ability, such as $\theta \in (\theta_1 \ ; \theta_2)$ with $\theta_1 < \theta_2$. When the capacity of the local government is high (it is said to be 'efficient'), the marginal cost of effort is low (θ_1), and vice versa. Consequently, for the same level of effort, the total cost of the effort of the local government of type 1 is lower than that of the local government of type 2. To get the same level of utility as a local government of type 2, the local government of type 1 must provide more effort. But, for the same level of effort (e) the local government of type 1 (the most effective) will get a greater level of utility.

In a perfect world within a context of perfect information, the voters could be aware of the

value of θ for each local government and assess them by comparing their values of θ and the efforts they realize. If the voters can observe that are not in adequation with their competencies, they can punish the government through the elections. In reality, it's impossible to assess the value of θ for each government due to the asymmetry of information. The local government of type 1 can reduce their efforts to increase their utility.

The voter's utility function could be like this:

$$U_v = \delta[L(e) - b] - \lambda(e) - s \tag{2}$$

With $\delta \in \mathbb{R}$, b>0, $\lambda > 0$, s>0.

In the equation 2, when additional public expenditures $L(e)$ are incurred, the utility of the median voter varies, on one hand, based on their preferences for gender equality (δ[L(e) – b]), and on the other hand, depending on the amount of taxation intended to finance these expenditures or the opportunity cost of forgoing certain expenses, as well as the local government's effort λe.

The degree of preference of the median voter for social housing is given by the exogenous parameter δ, which is a real number taking a negative, positive, or zero value depending on whether the voter is respectively hostile, favorable, or indifferent to an allocation of public expenditures or greater emphasis on gender issues in their state.

Furthermore, for each voter, there exists a significant level of allocation of public expenditures or importance assigned to gender denoted as b. If this level is not reached (L(e) < b), the voter who does not want additional allocation sees their satisfaction level increase, while the favorable voter experiences a decrease in their well-being.

λe represents the total financial cost of the local government's effort incurred by the median voter through local taxation (with $\lambda > 0$). The value of the marginal financial cost λ depends on the more or less favorable nature of the local context. Specifically, at a given effort level, the financial cost of the local government's effort will be higher or lower depending on the extent of financial support from the central government through various transfers or the specific characteristics of the States. Indeed, more populous states have a potentially broader tax base, while more urbanized or densely populated states, a priori, have lower collection costs. Indeed, Aizenman and Jinjarak (2008) and Andersson (2018) found that urbanization reduces collection

costs and improves the efficiency of taxes like land taxes or Value Added Tax (VAT).

Financial costs are thus distributed over a larger number of taxpayers, and the financial cost is lower for the median voter. To simplify, states are divided into two categories based on the characteristics of their local environment, such as λ=(λ_1; λ_2) with $\lambda_1 < \lambda_2$. Consequently, for the same effort level, the total financial cost of the local government's effort in a type 1 State is lower than that of the local government in a type 2 State. Our model was inspired by Bilek et al. (2008) and Lane et al. (2013).

Vaishnav and Swanson (2015) found that Indian voters' attitude changed since the 2000s with a more retrospective voting attitude. In the hypothesis of a retrospective electoral behavior in which elected officials are judged based on their performance, the local government uses the economic policy instruments at its disposal to increase its chances of reelection. This also places us within the framework of the opportunistic political-economic cycles theory (Nordhaus (1975)), which seems relevant in the Indian context (Chowdhury (1993)).

In the context of perfectly accessible information, voters can determine the type of their local government and rely on this information for evaluation and voting sanctions. Unfortunately, reality corresponds more to a framework of information asymmetry. In this framework, the assumption of asymmetric information implies that the type and level of effort of the local government are neither observable nor verifiable, as well as the type of the municipality. The voter is therefore faced with a problem of adverse selection and moral hazard. On the one hand, the voter does not know whether the local government is "performing" or not in implementing "pro-gender" policies (adverse selection). In other words, the marginal cost θ is private information held only by the local government. Although the voter cannot identify the type of local government, they know that there is a probability p of being "performing" and 1-p of being "non-performing".

In asymmetric information settings, individuals have an incentive to misrepresent their types. Indeed, with a lower θ_1 for efficient local governments, they can always appear more effective than less efficient ones by choosing an effort level e^* such that the outcomes would be superior to those of less efficient local governments. This is done while remaining below the effort level that would constitute an optimal solution maximizing the surplus of voters and the local government. Information asymmetry thus gives local governments an informational rent that can harm the interests of voters. The implementation of gender budgeting can reduce this information

asymmetry. The transparency requirement on the means and objectives set by gender budgeting forces local governments to disclose all or part of the value of the parameters θ in their utility functions and allows voters to weight this type of θ against the means (λ) used.

What about the possibility of coordination among local governments to all align with the less efficient ones? In practice, it is perfectly possible to imagine the establishment of coordination among different local governments. The latter may be tempted to reach an agreement for all to align with the less efficient ones and increase their utility at the expense of the voters. However, this would put us in a framework similar to that of the prisoner's dilemma. In this scenario, each agent (here, the local governments) has an interest in deviating from the coordination strategy to maximize their gains. To avoid this temptation to "cheat," it is necessary to introduce mechanisms of either sanctions or incentives to encourage or compel agents to cooperate.

In the context of electoral competition in a federal system, implementing mechanisms that encourage local governments to cooperate under the threat of sanctions jeopardizes democracy itself. The establishment of a coordination strategy can be implemented on a small scale or within the framework of a coalition within a local government. Its implementation on a broader scale with all local governments as stakeholders seems less plausible.

From this analysis within this theoretical framework, it emerges that, as in the case of a principal-agent contract, local governments benefit from informational rent. They hold information about themselves that only they know. Therefore, they can calibrate their efforts to maximize their utility even at the expense of voters. The implementation of gender budgeting, by incorporating a certain obligation of transparency, helps to eliminate information asymmetry and reduce the rent of local governments. The latter, being more exposed to the risk of retrospective electoral sanctions, have every interest in appearing efficient and credible. The implementation of gender budgeting could thus result in greater importance given to "pro-gender" expenditures. Local governments adopting gender budgeting should therefore be more transparent in their use of public money and become more efficient that non-adopter ones.

The implementation of empirical analysis is necessary to verify the consistency of our results with the data. Thus, the following section will present the chosen empirical framework to verify the theoretical intuitions.

3) Methodology

3.1) Data

3.1).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 estimate 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.

$$Y_{it} = \alpha + \beta X_{it} + \epsilon_{it} \tag{3}$$

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:

$$Y_{it} = \alpha + \beta X_{it} + v_{it} - u_{it} - \eta_{it} \tag{4}$$

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}, $\widehat{\theta_i}$ and $\widehat{\gamma_{it}}$.

$$y_{it} = \alpha_0^* + \beta X_{it} + \theta_i + \gamma_{it} \tag{5}$$

Where

$$\alpha_0^* = \alpha_0 - E(\eta_{it}) - E(u_{it}) \tag{6}$$

$$\theta_i = \alpha_i - \eta_i + E(\eta_i) \tag{7}$$

$$\gamma_{it} = v_{it} - u_{it} + E(u_{it}) \tag{8}$$

After the first step, we realize a stochastic frontier method to estimate the persistent and transient (or short-run) technical inefficiency $\widehat{u_{it}}$. 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).

3.1).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	139
Untreated	465

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
efficiency score	4.682	2.037	1.474	16.689	585
L.log(GDPpc)	10.287	1.061	7.886	12.832	942
L.urbanization	33.568	19.098	7.98	99.900	870
L.log(population)	20.847	0.125	20.608	21.025	843
fiscal rule	0.381	0.486	0	1	942
L.autonomy	48.938	25.713	5.466	100	887
trend	16.815	9.352	1	33	942

3.1).3 Stylized facts

The graph 3 illustrates the staggered adoption of gender budgeting across various states in India. It highlights the timeline and sequence in which different states implemented gender budgeting

practices, showcasing the varying pace of adoption. The data underscores how some states embraced the initiative earlier, while others followed more gradually, reflecting the diverse policy responses across the country.

Missing data often arises because some states did not exist prior to a certain point in time. Consequently, these states could not have been subject to any "treatment" immediately upon their creation. This situation ensures that there are "not yet treated" observations for all states, as newly formed states naturally fall into this category until they eventually receive treatment. This allows for a clearer comparison between treated and untreated states over time.

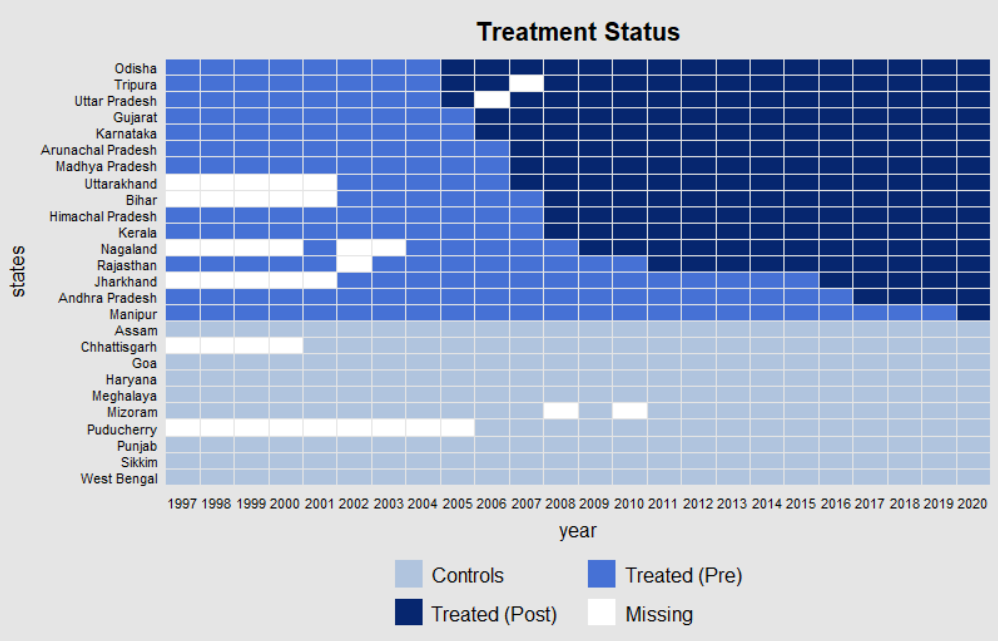


Figure 3: Treatment adoption by States

The graph 4 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. It seems to suggest that states that have adopted gender budgeting are 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.

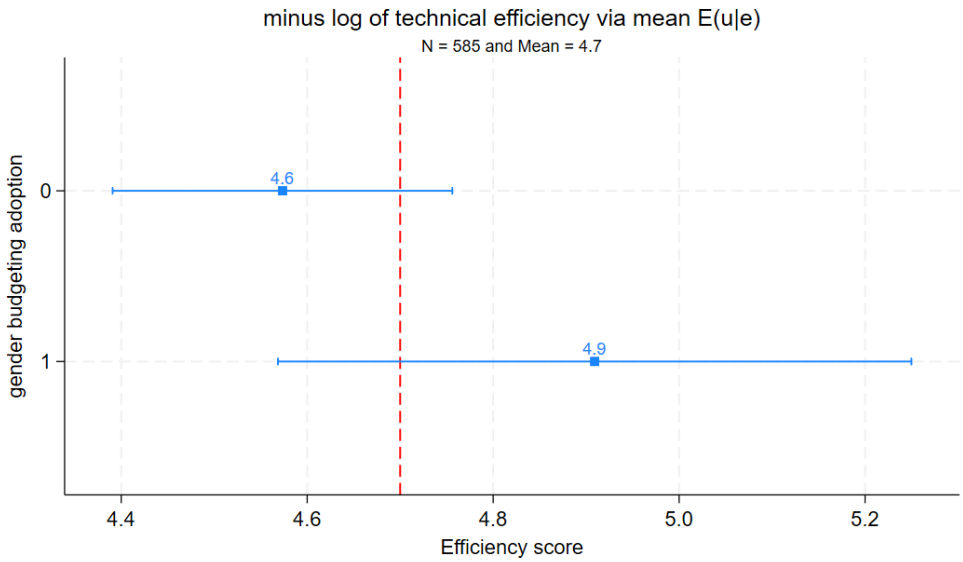


Figure 4: Efficiency score for adopters (1) and non-adopters(0)

3.2) Identification strategy

In this case, I will utilize the multilevel Tobit estimation method to investigate the effects of gender budgeting on efficiency scores, which are constrained between 0 and 100. This method has also been used by Rodden and Wibbels (2002) and Zhao et al. (2021) . The efficiency scores in my dataset exhibit censorship at both lower and upper bounds—no score can fall below 0 or exceed 100. Traditional linear regression techniques, such as Ordinary Least Squares (OLS), are inadequate for such censored data, as they do not account for the constraints imposed on the dependent variable. The Tobit model is specifically designed to address these limitations by modeling the censored nature of the dependent variable.

The Tobit model, originally developed by James Tobin, is well-suited for scenarios where the dependent variable is subject to censoring. In my analysis, the efficiency score is constrained within a specified range, and a Tobit model provides a framework that properly handles this constraint. By assuming that the underlying efficiency scores follow a normal distribution but are observed only within the bounded range, the Tobit model ensures that predicted values remain within the valid range of 0 to 100. This is crucial for accurate estimation and interpretation of the relationship between gender budgeting and efficiency scores.

Given the hierarchical nature of my data—where efficiency scores are collected across different countries or regions over multiple time periods—the multilevel Tobit model introduces random effects to account for unobserved heterogeneity at different levels. Specifically, the multilevel approach allows for the inclusion of random effects at the country level, capturing variations in efficiency that are attributable to unobserved country-specific factors. This is particularly important as countries may exhibit inherent differences in efficiency scores due to diverse institutional contexts, economic conditions, and implementation practices related to gender budgeting.

In conclusion, the multilevel Tobit estimation method is a robust tool for analyzing the effects of gender budgeting on efficiency scores constrained between 0 and 100. It effectively handles the censoring of the dependent variable and accounts for the hierarchical structure of the data, offering a comprehensive approach to understanding the complex dynamics between gender budgeting policies and efficiency outcomes across different contexts.

The Tobit model can be written as:

$$y_i^* = \alpha GB_{it} + \beta X_{it} + c_{it} + \varepsilon_{it}, \quad \varepsilon_{it} \sim N(0, \sigma^2)$$

with the observed variable y_i defined as:

$$y_i = \begin{cases} 0, & \text{if } y_i^* \leq 0, \\ y_i^*, & \text{if } y_i^* > 0 \\ 100, & \text{if } y_i^* > 100. \end{cases}$$

4) 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 (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.

	Dependent variable:
	efficiency score
gender_budgeting	0.639**
	(0.374)
Observations	479
Note:	*p<0.1; **p<0.05; ***p<0.01

5) Robustness check: alternative estimation method

5.1) Entropy banalncing

5.1).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.

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:

$$\pi ATT(x) = E[Y(1)|T = 1, X = x] - E[Y(0)|T = 0, X = x] \tag{9}$$

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, 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:

$$Y_{i_t} = \beta_1 GB_{it} + \alpha_1 log(GDP\_pc)_{it} + \alpha_2 log(density)_{it} + \alpha_3 X_{it} + \mu_i + \psi_t + \epsilon_{it} \tag{10}$$

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.

5.1).2 Correlation issue

Table **??** 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 3: Unweighted Balance Statistics

Variable	Mean [GB=1]	Mean [GB=0]	Difference ([GB=1]-[GB=0])
lautonomy	43.99	48.92	-4.93
lloggdppc	10.82	10.00	0.82
lurban	27.84	30.15	-2.31
trend	23.39	14.41	8.98
llocal_wip	48.49	48.71	-0.22
fiscal_rule	1.00	0.39	0.61
llogpop	20.96	20.83	0.13

Table 4: Weighted Balance Statistics

Variable	Mean [GB=1]	Mean [GB=0]	Difference ([GB=1]-[GB=0])
lautonomy	43.99	43.99	0.00
lloggdppc	10.82	10.82	0.00
lurban	27.84	27.83	0.01
trend	23.39	23.39	0.00
llocal_wip	48.49	48.49	0.00
fiscal_rule	1.00	1.00	0.00
llogpop	20.96	20.96	0.00

5.1).3 Results

The results suggest a positive and significative impact of gender budgeting adoption on health public spending quality. In the next table(**??**) I have added time and political parties fixed effects. Because of the important number of states, the addition of states fixed effects could lead to a bias in my result. In addition, in my analysis, I employed ruling parties fixed effects rather than state fixed effects due to the nature of decision-making processes regarding gender budgeting and governance reforms. Typically, these decisions are driven by the political parties in power rather than the individual states. This approach recognizes that the same party, even

when governing different states, is likely to implement consistent policies and strategies. By using ruling parties fixed effects, I account for the fact that policy decisions are influenced more by the party's agenda and ideology than by the specific characteristics of each state. This method allows for a more precise estimation of the impact of gender budgeting policies, as it isolates the effect of the party's policy choices from state-specific factors that might otherwise confound the results.

Table 5: Entropy balancing results

Variables	efficiency score			
gender budgeting	0.811**	1.201***	0.556**	0.760**
	(2.168)	(2.709)	(1.980)	(2.110)
parties FE	No	No	Yes	Yes
years FE	No	Yes	No	Yes
Covariates	Yes	Yes	Yes	Yes
Observations	502			

5.2) Placebo Test

I now examine whether there are confounding factors that could affect the results, which have remained stable so far. 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 **??**, in Appendix). Therefore, I can rule out the possibility of confounding — factors influencing our results.

5.3) Anticipation effects

Always to check the robustness of our results and be sure that the effects observed are due to the treatment adoption, we change the date of the adoption to test for potential anticipation effects. An example of anticipation effects could be the fact that the reform could be discussed in newspapers years before their adoption and that there are economic or political reasons for rulers to change spending allocation before reforms. So, the anticipation effect can have an impact on the size of the outcome and the treatment effects estimation (Mertens and Ravn (2012) and Metiu (2021)).

I change the adoption wave date by considering that the treatment has been adopted two years before the effective date of adoption to test the presence or not of anticipation effects. The results obtained are presented in the appendix section at the table 11.

The results show a non-significant effect for our alternative adoption dates. I can conclude that an absence of anticipation effects of gender budgeting adoption on the "pro gender" public spending allocation. However, I found an existing anticipation effect for education allocation spending. This effect is less important than the effect after the adoption, and the anticipation effect didn't seem to explain all the results for efficiency of health spending allocation.

5.4) Political fragmentation effects

To ensure that my results are not driven by the possibility that a single party can unilaterally decide all spending allocations due to political centralization, I compute a Herfindahl-Hirschman Index (HHI) to measure the fragmentation of votes in local parliaments. Political fragmentation refers to the dispersion of political power among multiple parties, which can lead to more inclusive decision-making, potentially increasing spending efficiency. Conversely, when one party dominates, spending decisions may be more centralized and less efficient. By generating an interaction term between this political fragmentation variable and the gender budgeting variable, I account for the potential joint effects of gender-responsive policies and political decentralization

on spending efficiency. The Herfindahl-Hirschman Index is calculated as:

$$HHI = \sum_{i=1}^{N} s_i^2$$

where s_i represents the share of votes got by party i in the local parliament. A higher HHI indicates lower fragmentation (i.e., more political centralization), while a lower HHI signals greater fragmentation. The normalized formula is:

$$HHI_{\text{norm}} = \frac{HHI - \frac{1}{N}}{1 - \frac{1}{N}}$$

where N is the total number of parties.

The results available in table 12 in appendix section show that political fragmentation can have an effect on spending efficiency. This effect is too loww to explain the previous results. In addition, this effect is due to the combination between political centralization and gender budgeting adoption. However the political centralization only, negatively affects the efficiency score.

5.5) transmission channels

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.

$$\begin{aligned} bias\_index_{it} = |(health\_spending_{it}/Total\_expenditures_{it}) * 100 - \\ (health_spending\_forecasted_{it}/Total_expenditures\_forecasted_{it}) * 100| \end{aligned} \tag{11}$$

Table 6: 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 *The absolute value refers to the fact that we multiply the negative value by -1 to get only positive values* 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 7. 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.

Table 7: Transmission channel

	bias index
Before adoption	0.13***
After adoption	0.10***
Non Gender Budgeting	0.12***

These results seem to corroborate those of Hory (2016); Olanubi and Olanubi (2023a); Olanubi and Olanubi (2023b); Ouertani et al. (2018) and Cabezon et al. (2015) that explain that good

public financial management, better fiscal credibility, and strong tax administration[5] 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 8: Pairwise correlation

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

The results shown in the table 8 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 (9) summarizes the results for the two main variables. The results are highly significant and seem to confirm the previous results and the intuition about 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.

[5]each Indian states have its own Finance ministry

Table 9: 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

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 7.

Gender budgeting adoption by imposing fiscal policy and budgetary process continuous 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

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.

Aizenman, J. and Jinjarak, Y. (2008). The collection efficiency of the value added tax: Theory and international evidence. *Journal of International Trade and Economic Development*, 17(3):391–410.

Andersson, P. F. (2018). Democracy, urbanization, and tax revenue. *Studies in Comparative International Development*, 53:111–150.

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.

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*.

Bilek, A., Costes, N., and Monmousseau, F. (2008). La loi sru incite-t-elle les maires à construire du logement social? les enseignements d'une analyse principal-agent. *Économie publique/Public economics*, (20).

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.

Brennan, G. and Buchanan, J. M. (1980). *The power to tax: Analytic foundations of a fiscal constitution*. Cambridge University Press.

Buchmann, C., DiPrete, T. A., and McDaniel, A. (2008). Gender inequalities in education. *Annu. Rev. Sociol*, 34:319–337.

Cabezon, E., Tumbarello, M. P., and Wu, M. Y. (2015). *Strengthening fiscal frameworks and improving the spending mix in small states*. International Monetary Fund.

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.

Chowdhury, A. R. (1993). Political surfing over economic waves: parliamentary election timing in india. *American Journal of Political Science*, pages 1100–1118.

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*.

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*.

Gavazza, A. and Lizzeri, A. (2009). Transparency and economic policy. *The Review of Economic Studies*, 76(3):1023–1048.

Ghosh, S. and Pal, S. (2004). The effect of inequality on growth: Theory and evidence from the indian states. *Review of Development Economics*, 8(1):164–177.

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,*.

Lane, J.-E. et al. (2013). The principal-agent approach to politics: Policy implementation and public policy-making. *Open Journal of Political Science*, 3(02):85.

Meckling, W. H. and Jensen, M. C. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial economics*, 3(4):305–360.

Mertens, K. and Ravn, M. O. (2012). Empirical evidence on the aggregate effects of anticipated and unanticipated us tax policy shocks. *American Economic Journal: Economic Policy*, 4(2):145–181.

Metiu, N. (2021). Anticipation effects of protectionist us trade policies. *Journal of International Economics*, 133:103536.

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.*

Nordhaus, W. D. (1975). The political business cycle. *The review of economic studies*, 42(2):169–190.

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. (2023a). Political short termism and government spending efficiency in sub-saharan africa. In *Handbook on Public Sector Efficiency*, pages 430–442. Edward Elgar Publishing.

Olanubi, S. O. and Olanubi, O. E. (2023b). 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., 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.

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.

Revelli, F. (2002). Local taxes, national politics and spatial interactions in english district election results. *European journal of political economy*, 18(2):281–299.

Rodden, J. and Wibbels, E. (2002). Beyond the fiction of federalism: Macroeconomic management in multitiered systems. *World Politics*, 54(4):494–531.

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.

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

Vaishnav, M. and Swanson, R. (2015). Does good economics make for good politics? evidence from indian states. *India Review*, 14(3):279–311.

Verhoeven, M., Gunnarsson, V., and Carcillo, S. (2007). Education and health in g7 countries: Achieving better outcomes with less spending.

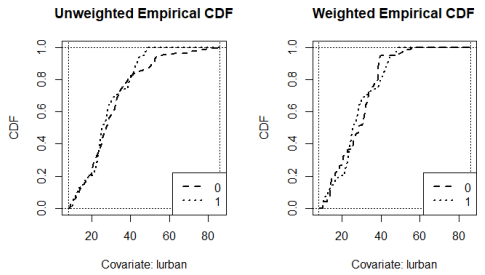
Vermeir, J. and Heyndels, B. (2006). Tax policy and yardstick voting in flemish municipal elections. *Applied Economics*, 38(19):2285–2298.

Zhao, C., Feng, F., Chen, Y., and Li, X. (2021). Local government competition and regional innovation efficiency: From the perspective of china-style fiscal federalism. *Science and Public Policy*, 48(4):488–489.

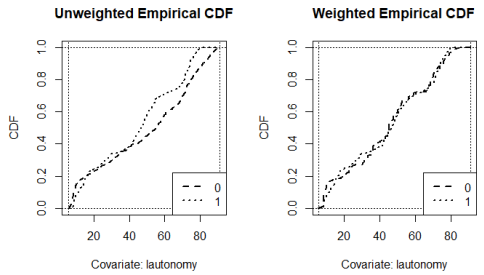
Zhao, Q. and Percival, D. (2017). Entropy balancing is doubly robust. *Journal of Causal Inference*, 5(1).

7) Appendix

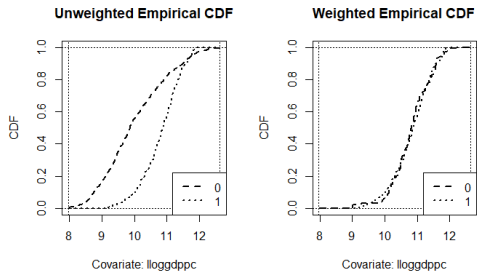
7.1) Entropy balancing



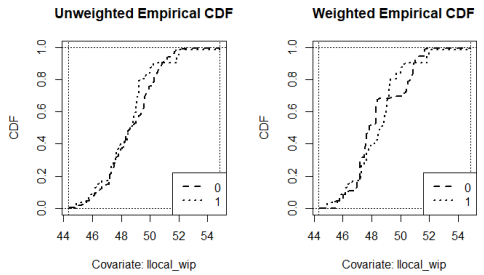
(a) Urbanization balance



(b) Autonomy balance



(c) GDPpc balance



(d) Wip Balance

Figure 5: Entropy banlancing graphs

7.2) Placebo test

Table 10: Results for the placebo test

	Dependent variable:
	efficiency score
placebo	0.122
	(0.436)
Observations	479
Note:t statistics in parentheses	*p<0.1; **p<0.05; ***p<0.01

7.3) Anticipation effects

Table 11: Results for the anticipation test

	Dependent variable:
	efficiency score
two_years_before_policy	−0.871
	(-1.122)
Observations	479
Note:t statistics in parentheses	*p<0.1; **p<0.05; ***p<0.01

7.4) Political fragmentation

Table 12: Political fragmentation results

Variables	efficiency score			
HHI	-0.002***	−0.002***	−0.002***	−0.002***
	(-4.749)	(-6.840)	(-4.845)	(-8.568)
HHI*GB	0.0007**	0.001***	0.001***	0.002***
	(2.566)	(2.888)	(2.693)	(8.006)
parties FE	No	No	Yes	Yes
years FE	No	Yes	No	Yes
Covariates	Yes	Yes	Yes	Yes
Observations	379			