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

This paper investigates a novel channel of gender-responsive budgeting (GRB): its impact on the technical efficiency of public health spending. While most of the literature has focused on GRB's redistributive or participatory dimensions, I examine its role as a fiscal discipline mechanism that enhances administrative accountability. Building on a principal-agent framework, I argue that GRB, by increasing transparency and citizen oversight, can reduce moral hazard, improve resource targeting, and enhance intergovernmental coordination in decentralized settings. Empirically, I exploit the staggered adoption of GRB across Indian states using a difference-in-differences estimator suited for heterogeneous treatment timing (Callaway and Sant'Anna (2021)), combined with entropy balancing to address selection into adoption. The results show that GRB significantly improves the technical efficiency of health expenditures, particularly in reducing preventable mortality and expanding access to care. This study contributes a new perspective on GRB, not solely as a tool for gender equity, but as an institutional lever to strengthen the quality of public spending in federal systems.

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

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

Author contributions:

The article is single-authored and, thus, the author is responsible for all its contents.

List of Figures

List of Tables

Contents

1) Introduction

This paper investigates an underexplored dimension of gender-responsive budgeting (GRB): its impact on the efficiency of public health spending in a decentralized context. GRB refers to the systematic integration of gender considerations into the planning, execution, and monitoring of public budgets. It is not a separate budget, but a cross-cutting approach aimed at assessing how fiscal policies affect women and men differently, and at reallocating resources to reduce structural inequalities.

While GRB has been widely promoted as a tool to advance gender equity, its implications for public sector performance and the quality of spending have received far less attention. Yet improving not just who benefits from public resources, but how effectively those resources are used, is a central concern for governments—especially in developing and federal countries facing tight budget constraints and high service delivery demands.

This paper asks a simple but important question: can GRB, beyond promoting inclusiveness, enhance the efficiency with which governments deliver essential services? The underlying intuition is that GRB, by institutionalizing transparency, defining measurable objectives, and requiring ex-ante and ex-post evaluations, can strengthen fiscal accountability. This, in turn, may improve the way inputs are converted into outcomes, especially in sectors like health where inefficiencies are widespread. Prior research has linked fiscal transparency to improved budgetary outcomes (Chen et al. (2019); Chan and Karim (2012); De Simone et al. (2019); Gavazza and Lizzeri (2009); Montes et al. (2019)), but the potential of GRB to play a similar role has not been empirically established.

To examine this relationship, I focus on India, a large federal country where GRB was first introduced at the national level in 2000. Starting in 2005 with the state of Odisha, GRB was gradually adopted by other Indian states, using common templates developed by the Ministry of Finance and the National Institute of Public Finance and Policy (NIPFP). These included gender budget statements and analytical matrices to assess program impact. This staggered adoption across states creates a natural setting for causal inference. Figure 1 summarizes the institutional structure of GRB.[1]

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

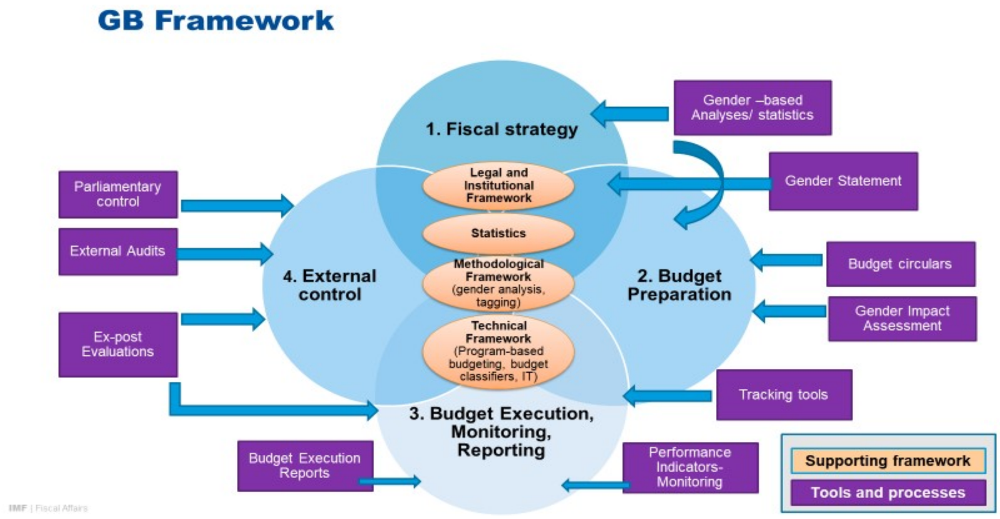


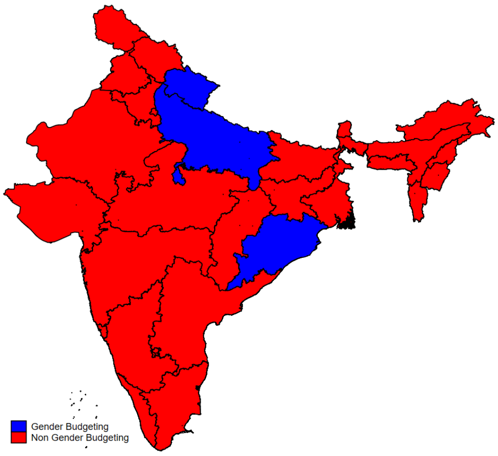
Figure 1: Gender Budgeting framework (Source: IMF PFM Blog)

This paper focuses on the health sector, which is particularly relevant for assessing subnational spending efficiency. Under the Indian Constitution, public health falls under the State List, giving state governments full responsibility over its funding and delivery. In contrast, other sectors such as education fall under the Concurrent List and are more directly shaped by central government decisions. Health therefore provides a cleaner lens for identifying the state-level effects of GRB.

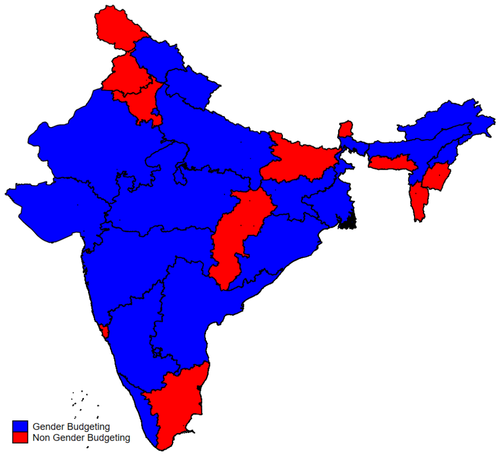
To estimate the impact of GRB on health spending efficiency, I construct a panel dataset of Indian states from 2005 to 2018. I use a difference-in-differences approach that accounts for staggered treatment timing (Callaway and Sant'Anna (2021)), combined with entropy balancing to adjust for observable differences between adopters and non-adopters. Efficiency is measured using an output-oriented frontier model comparing health outcomes to fiscal inputs.

Figure 2 illustrates the diffusion of GRB across Indian states. Adoption occurred in waves, reflecting heterogeneity in political will, administrative capacity, and institutional maturity.

Previous studies on GRB have mostly focused on its impact on gender-related spending, social outcomes, or institutional barriers to implementation (Chakraborty (2016); Stotsky and Zaman (2016)). However, they have largely overlooked whether GRB improves the actual efficiency of public service delivery. This distinction matters. An increase in health spending after GRB adoption does not automatically imply better outcomes. Governments may increase allocations to meet formal GRB guidelines or signal political alignment, without ensuring that additional



(a) Gender budgeting adoption in 2005



(b) Gender budgeting adoption in 2018

Figure 2: GRB adoption over time across Indian states

resources are well targeted or effectively managed. In such cases, GRB risks becoming a box-ticking exercise rather than a driver of meaningful fiscal reform.

This paper contributes to closing that gap by providing the first empirical assessment of whether GRB improves public spending efficiency at the state level. The results show that GRB adoption is associated with significant improvements in the efficiency of health expenditure. These effects are robust across specifications, persist over time, and vary across adoption waves—suggesting that context and implementation quality matter. The findings reframe GRB as not only a gender equity instrument, but also a tool for enhancing public sector performance in decentralized systems.

The rest of the paper is structured as follows. Section 2 presents the theoretical framework. Section 3 describes the methodology for estimating efficiency scores. Section 4 reports the main results. Section 5 provides robustness checks, including entropy balancing. Section 6 explores potential transmission mechanisms. Section 7 concludes.

2) Theoretical Framework

This section develops a theoretical model to analyze how gender budgeting transparency affects the efficiency of public health spending. The framework builds on a principal-agent approach

(Meckling and Jensen (1976)) in which a state government (the agent) allocates resources on behalf of the public (the principal) but may do so inefficiently due to asymmetric information and potential rent-seeking behavior. These benefits can be of a monetary or non-monetary nature (Brennan and Buchanan (1980)). Gender budgeting transparency is introduced as a mechanism that reduces information asymmetry and reshapes the agent's incentives, thereby influencing the efficiency of spending decisions. The reduction of information asymmetry can be due to the publication of *ex ante* and *ex post* reports by states governments with GRB adoption. The principal can also compare the results of their agent with those of neighboring jurisdictions (Revelli (2002) and Vermeir and Heyndels (2006).

I consider a repeated cross-sectional setting in which, in each period t, a new government G_t is responsible for allocating a health budget B_t. The efficiency of spending, denoted S_t, depends on two key factors: the government's effort e_t to ensure efficient allocation, and the level of transparency in gender budgeting, denoted $T_t \in [0, 1]$. The government's type, $\theta_t \in [0, 1]$, captures the degree of rent-seeking preference, with higher values indicating lower commitment to efficiency.

Efficiency is modeled as a strictly increasing function of both effort and transparency:

$$S_t = g(e_t, T_t), \quad \text{with } \frac{\partial S}{\partial e_t} > 0, \quad \frac{\partial S}{\partial T_t} > 0.$$

Effort is costly for the government and is represented by a convex cost function $C(e_t) = \frac{1}{2}ce_t^2$. This captures the idea that marginal cost of effort increases: it becomes harder (more costly) for the government to maintain higher levels of implementation effort (e.g., better targeting, program monitoring, administrative capacity). The government derives utility from three components: (i) political or reputational gains from efficient public service delivery, weighted by a parameter λ; (ii) potential rents from diverting public resources, which are inversely related to the level of transparency; and (iii) the cost of effort. The government's utility function is given by:

$$U_G = \lambda S(e_t, T_t) + (1 - T_t)\theta_t B_t - \frac{1}{2}ce_t^2.$$

The government chooses effort e_t to maximize U_G. The first-order condition for an interior maximum is:

$$\frac{\partial U_G}{\partial e_t} = \lambda \frac{\partial S}{\partial e_t} - ce_t = 0 \quad \Rightarrow \quad e_t^* = \frac{\lambda}{c} \frac{\partial S}{\partial e_t}.$$

This expression shows that the optimal effort is increasing in the marginal productivity of effort and in the weight the government places on public welfare. Since S depends positively on transparency, transparency indirectly increases optimal effort:

$$\frac{de_t^*}{dT_t} > 0.$$

Consequently, total spending efficiency is increasing in transparency both directly, via $\partial S / \partial T_t > 0$, and indirectly, through the effort channel:

$$\frac{dS_t}{dT_t} = \frac{\partial S}{\partial e_t} \cdot \frac{de_t}{dT_t} + \frac{\partial S}{\partial T_t} > 0.$$

Beyond the effect on effort, transparency reduces the marginal benefit from rent-seeking by lowering the weight on the term $(1 - T_t)\theta_t B_t$. When transparency increases, governments have fewer incentives to divert funds because the political cost of detection increases. This implies that rent-seeking behavior θ_t itself may respond endogenously to transparency, leading to a second indirect effect. Assuming θ_t is decreasing in T_t, i.e., $\frac{d\theta_t}{dT_t} < 0$, and efficiency is decreasing in θ_t, I obtain another reinforcing mechanism:

$$\frac{dS_t}{dT_t} = \frac{dS_t}{d\theta_t} \cdot \frac{d\theta_t}{dT_t} > 0.$$

Transparency also contributes to the strengthening of public financial management institutions. Through the standardization of budget classification, gender-disaggregated reporting, and monitoring mechanisms, transparency raises the institutional quality I_t, which itself enhances efficiency. If S_t increases with I_t and I_t increases with T_t, then:

$$\frac{dS_t}{dT_t} = \frac{dS_t}{dI_t} \cdot \frac{dI_t}{dT_t} > 0.$$

Finally, transparency enhances citizen awareness and electoral accountability. When information about budget allocations and outcomes is publicly available, voters can better evaluate performance and reward or sanction governments accordingly. This increases political competition C_t, which acts as another channel through which transparency raises efficiency:

$$\frac{dS_t}{dT_t} = \frac{dS_t}{dC_t} \cdot \frac{dC_t}{dT_t} > 0.$$

Summing up these effects, the total impact of transparency on efficiency is the sum of direct and multiple reinforcing indirect channels:

$$\frac{dS_t}{dT_t} = \frac{\partial S}{\partial T_t} + \frac{\partial S}{\partial e_t} \cdot \frac{de_t}{dT_t} + \frac{dS_t}{d\theta_t} \cdot \frac{d\theta_t}{dT_t} + \frac{dS_t}{dI_t} \cdot \frac{dI_t}{dT_t} + \frac{dS_t}{dC_t} \cdot \frac{dC_t}{dT_t}.$$

This expression highlights how transparency not only disciplines current behavior but also initiates a cascade of institutional and political responses that raise the marginal return to effort, reduce leakage, and improve service delivery outcomes over time. The framework implies that transparency is particularly effective in settings with strong citizen engagement, and competitive elections under which both the direct and indirect effects are most pronounced.

The next setion will assess empirically GRB adoption to verify the intuitions of our theoretical expectations.

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.

$$Y_{it} = \alpha + \beta X_{it} + \epsilon_{it} \tag{1}$$

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

$$Y_{it} = \alpha + \beta X_{it} + \upsilon_{it} - u_{it} - \eta_{it} \tag{2}$$

In this estimation, υ_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 $\widehat{\gamma}_{it}$.

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

Where

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

$$\theta_i = \alpha_i - \eta_i + E(\eta_i) \tag{5}$$

$$\gamma_{it} = \upsilon_{it} - u_{it} + E(u_{it}) \tag{6}$$

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	193
Untreated	560

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.

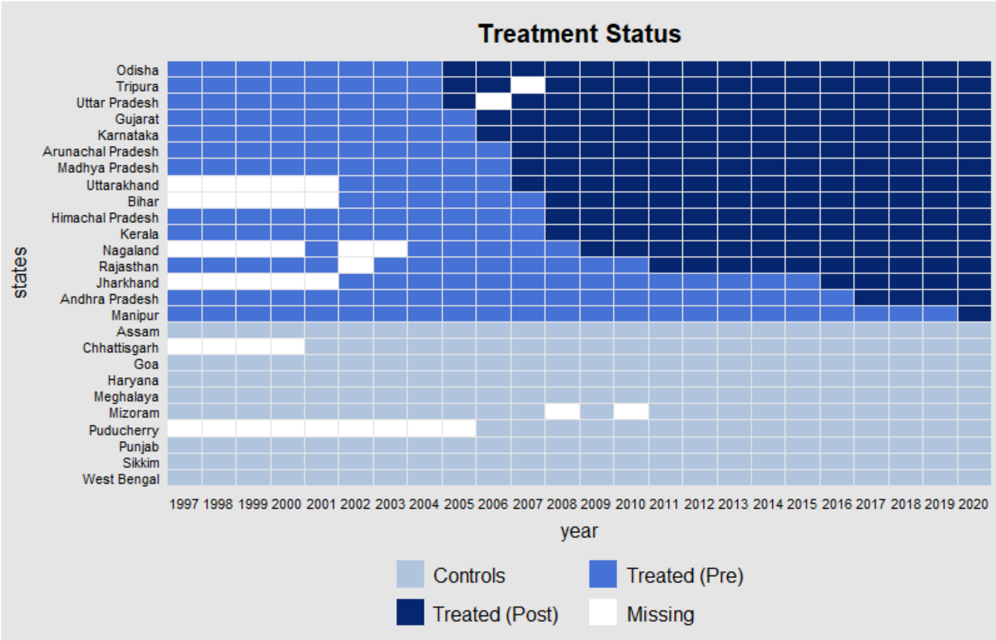


Figure 3: Treatment adoption by States

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. Throughout the remainder of the paper, all standard errors are clustered at the state level for macroeconomic outcomes and at the district level for microeconomic outcomes.

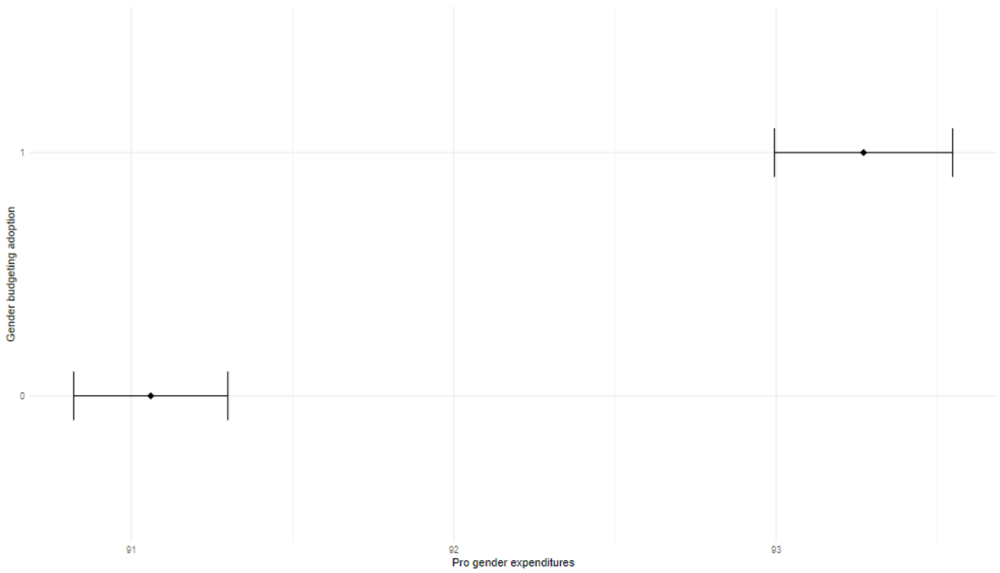


Figure 4: Efficiency score for adopters (1) and non-adopters(0)

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 "pro gender" 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 public expenditures 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 1991–2020, 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., public expenditures for State i in year t = 1, ..., T

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

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 public expenditures (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

I compute a *t-test* to compare the mean of our outcome variable for both (adopters and non-adopters) before the first year of implementation. The results available in table ?? show that the mean of the outcome variable difference is not statistically different between adopters and non adopters before the treatment was applied.

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.

	Before adoption		
Outcomes	non Adopters	Adopters	Difference
Efficiency score	90.17	92.09	
	After adoption		
Outcomes	non Adopters	Adopters	Difference
Efficiency score	91.34	93.38	***

Table 3: Outcome means before the treatment (by status)

3.4) Results

The results are presented in the following table (table 4). They suggest a positive effect of gender budgeting on efficiency scores. Indeed, analyzing budgets from a gender perspective is integral to gender mainstreaming. When gender considerations are embedded in policy and project design, they should be reflected in resource allocation. If not, outcomes are unlikely to deliver substantive equality for women. Budgets thus serve as a critical tool for gender mainstreaming. Such practices enhance transparency, information disclosure, and citizen participation in economic governance. Several 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. Therefore, by increasing transparency and strengthening local administrative capacity, gender budgeting improves the efficiency of public spending in Indian states that have adopted it.

Table 4: Diff in Diff results

	efficiency score
ATT	1.4874***
	(0.576)
Observations	668

std errors in parentheses

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

This effect remains positive over time, as illustrated in graph 5 available in appendix, although the benefits take time to materialize. The delay may reflect a learning process within public administration. While governments may initially increase spending on health or other social sectors (e.g., education or infrastructure), it takes time to improve the quality and effectiveness of such spending through better budgetary processes and public expenditure planning (including *ex ante* evaluations).

Additionally, I analyze the effects of gender-responsive budgeting by adoption cohorts. The results suggest heterogeneity across cohorts. The treatment effects appear to be stronger among early adopters, who seem to drive the overall effect. For later adopters, the effects are more modest, as shown in Table 16 in the appendix. The non-significant effects observed for the cohorts adopting between 2006 and 2009 may be due to missing data and should be interpreted with caution.

The next section presents robustness checks for our results.

4) Robustness check: alternative estimation method

4.1) Entropy banalncing

4.1).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:

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

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

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

$$Y_{i_t} = \beta_1 GB_{it} + \alpha_1 log(GDP\_pc)_{it} + \alpha_2 log(density)_{it} + \alpha_3 X_{it} + \mu_i + \psi_t + \epsilon_{it} \quad (9)$$

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

4.1).2 Correlation issue

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

Table 5: Unweighted Balance Statistics

Variable	Mean [GB=1]	Mean [GB=0]	Difference ([GB=1]-[GB=0])
L.autonomy	43.99	48.92	-4.93
L.log(GDP per capita)	10.82	10.00	0.82
L.urbanization	27.84	30.15	-2.31
trend	23.39	14.41	8.98
L.women in parliement	48.49	48.71	-0.22
fiscal_rule	1.00	0.39	0.61
L.log(population)	20.96	20.83	0.13

Table 6: Weighted Balance Statistics

Variable	Mean [GB=1]	Mean [GB=0]	Difference ([GB=1]-[GB=0])
L.autonomy	43.99	43.99	0.00
L.log(GDP per capita)	10.82	10.82	0.00
L.urbanization	27.84	27.83	0.01
trend	23.39	23.39	0.00
L. women in parliement	48.49	48.49	0.00
fiscal_rule	1.00	1.00	0.00
L.log(population)	20.96	20.96	0.00

4.1).3 Results

The results suggest a positive and significative impact of gender budgeting adoption on health public spending quality. In the next table(table 7) 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 7: Entropy balancing results

Variables	efficiency score			
gender budgeting	1.21 **	1.02***	1.45 ***	1.03 ***
	(2.03)	(3.71)	(3.06)	(3.33)
parties FE	No	No	Yes	Yes
years FE	No	Yes	No	Yes
Covariates	Yes	Yes	Yes	Yes
Observations	541			

4.2) 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 17, in Appendix). Therefore, I can rule out the possibility of confounding — factors influencing our results.

4.3) 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 18.

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.

4.4) 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 seats 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 are summarized below

Table 8: Summary statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
HHI	0.422	0.177	0.151	1	730

The results available in table 19 in appendix section show that political fragmentation can influence spending efficiency. Indeed, the regression results highlight a nonlinear relationship between gender budgeting (GRB) and political concentration, measured by the Herfindahl-Hirschman Index (HHI), in shaping fiscal transparency outcomes. In models controlling for state and/or year fixed effects, the interaction term GRB $\times$ HHI is positive and significant, while the squared term (GRB $\times$ HHI) ^{2} is negative and marginally significant, indicating an inverted U-shaped effect. This suggests that the adoption of GRB has the strongest impact on efficiency when the political environment exhibits moderate concentration. In highly concentrated systems—where one party or a dominant coalition controls the legislature—the incentive to promote transparency and increase efficiency diminishes due to limited electoral competition and reduced accountability pressures. Conversely, in extremely fragmented systems, where power is dispersed across many small parties, fragile coalitions emerge that struggle to coordinate and sustain coherent transparency initiatives and/or coherent fiscal policy. These dynamics undermine the effective implementation of GRB. The turning point estimates range between 0.27 and 0.42 across models, with a consistent pattern around 0.40, suggesting that optimal conditions for transparency gains from gender budgeting occur when political concentration is balanced—neither too high nor too low. The magnitude of the coefficients should be interpreted in light of the fact that the HHI is normalized between 0 and 1; even small coefficient values imply substantial marginal effects across the range of political concentration.

4.5) 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.

$$\begin{gathered} bias\_index_{it} = |(health\_spending_{it}/Total\_expenditures_{it}) * 100 - \\ (health_spending\_forecasted_{it}/Total_expenditures\_forecasted_{it}) * 100| \end{gathered} \tag{10}$$

Table 9: 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 10. 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 10: 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[2] 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.

[2]each Indian states have its own Finance ministry

Table 11: Pairwise correlation

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

The results shown in the table 11 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 (12) 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.

Table 12: Simultaneous equations

	(1)	(2)
VARIABLES	eff_trans	bias index
l.bias index	-1.048***	
	(0.359)	
gender_budgeting		-0.017**
		(0.008)
Observations	541	541
R-squared	0.470	0.546

Standard errors in parentheses

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

5) Microeconomic effects

Beyond the effects that gender budgeting adoption can have on fiscal policy strategies and the resulting increase in the efficiency of public spending dedicated to health, I also examine how it can impact individuals' daily lives, particularly by enhancing their protection against health-related risks. Improved quality of health expenditures can strengthen effective health coverage by ensuring better access to care, increased prevention efforts, and reduced exposure to financial hardship due to illness (Erlangga et al. (2019)). In this context, expanding health insurance coverage plays a key role, as it facilitates access to healthcare services and reinforces the protective function of public health systems. For example, Clots-Figueras (2011) finds that politicians' gender affects policy, but that their social position, i.e., their caste, should be consider as well. Female legislators in seats reserved for lower castes and disadvantaged tribes invest more in health and early education and favor "women-friendly" laws, such as amendments to the Hindu Succession Act, which was designed to give women the same inheritance rights as men. They also favor redistributive policies, such as land reforms. In contrast, female legislators from higher castes do not have any impact on "women-friendly" laws.

In addition, India is a very large country with very large states. Indeed, some Indian states like Rajhastan are greater and more populous than countries like Finland, Norway, or Ivory Coast. So, it could be interesting to check the potential effect at the local and individual level. It's also important to notice that gender budgeting seems to become bottom-up approach. That means it is not the allocation of resources in the budget at national and or state levels that has to see but the resources that flow to and are available to women at the field level i.e. the women in the villages, cities and towns of the country that need to be monitored (Sharma and Garg (2014)). To assess the microeconomic effects of gender budgeting on health-related outcomes, I adopt a twofold approach.

First, I focus on maternal and child health by examining the quality of care and nutrition provided to children after birth. These indicators reflect not only household investment in child well-being but also broader access to basic health services. To capture aspects of reproductive health, I include pregnancy outcomes—specifically, whether the pregnancy resulted in a loss—as a proxy for the quality of maternal health and access to prenatal care. These dimensions are critical to understanding how gender-responsive fiscal policies can improve women's health trajectories and reduce early-life vulnerability.

Second, I analyze access to health insurance schemes, whether public or community-based, as an indicator of individual protection against health-related financial risks. Health insurance coverage offers a concrete measure of how well individuals are shielded from the economic consequences of illness or injury. By linking coverage to health status and vulnerability (as Pan et al. (2016)), I aim to evaluate whether gender budgeting contributes to broader and more equitable health protection systems.

So, it could be an interesting outcome to assess the microeconomic effect (on women) of gender budgeting adoption.

5.1) Data and empirical strategy

Data at individual level come from the *Data Health Survey* (DHS), which have been conducted in Indian states since the 1990's. The DHS household surveys typically interview a representative sample of between 10,000 to 20,000 women (aged 15-49) and men (aged 15-59).

To assess the microeconomic effects of gender budgeting adoption, we use the three last waves of *Data health survey*. This choice is due to the availability of data from the respondents. I also merge the DHS repeated cross sections dataset with the previous dataset with macroeconomic indicators at states level. This process leaves me with a dataset combining macro and micro indicators for a sample of around 75,000 women in 31 Indian States/UT The use of many waves allows to consider a potential time effect among states and check the effect of the time since the first adoption wave.

The next table summarizes the main variables used for our probit regression analysis on the microeconomics effects of gender budgeting adoption.

Table 13: Summary statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Outcomes :					
post natal check-up	0.398	0.568	0	1	3363153
fortified food (baby)	0.111	0.37	0	1	254719
Got insurance scheme from:					
States	0.088	0.283	0	1	454067
Community	0.001	0.026	0	1	454067
National	0.198	0.399	0	1	454067
roof materials	37.065	18.292	11	97	38540
number of children	2.36	1.345	0	14	45467
backward class or casts	2.585	1.089	1	8	43729
religion	2.628	10.637	1	96	45467
log(nightlights)	11.354	0.474	10.278	12.728	37669
urbanization	34.182	13.18	9.83	71.400	35257
women in parliement	48.546	1.637	44.47	52.49	37657
log(population)	3.344	1.731	-0.635	5.476	35257
dose	4.76	5.421	0	15	45467
partner education	2.624	1.588	0	8	74021

The dependent variables are some binary variable coded as 1 (if the respondent answers yes regardless of high or low intensity) and 0 (otherwise). The variable of interest is the time (in years) since the first implementation of gender budgeting (to measure the intensity of the treatment). I called it "dose". Given the qualitative nature of the dependent variable, the preferred estimation method for estimating equation (1) is the probit model. To address the lack of reliable State-level GDP data disaggregated at the local level. I constructed a measure of nightlights intensity at the district level. This proxy allows me to capture local economic activity with finer spatial granularity. Nightlights data offer a consistent and comparable measure of economic intensity across districts, helping to overcome limitations in official economic statistics.

Compared to the linear probability method and the logit model, the probit model is the most effective and efficient in estimating the qualitative model. Unlike the linear model, the coefficients from the Probit model estimations are not directly interpretable. They are interpreted in terms of marginal effects. The sign and significance of the parameters provide an indication of the impact of explanatory variables on the probability of observing the dependent variable's occurrence.

The probit estimation equation is presented below:

$$\Pr(Y_i = 1 \mid X_i) = \Phi(X_i'\beta)$$

$$Y_i^* = X_i'\beta + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0,1)$$

$$Y_i = \begin{cases} 1 & \text{if } Y_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

5.2) Microeconomic results

The results show that greater exposure to the gender-responsive budgeting (GRB) program is associated with a lower probability of pregnancy loss and no significant change in access to fortified food. However, exposure is linked to a decrease in the likelihood of receiving postnatal check-ups. On the insurance side, GRB exposure significantly increases the probability of being covered by state and national insurance schemes, while having no significant effect on community-based insurance. These patterns suggest that GRB may influence maternal health and financial protection through distinct institutional and behavioral channels. The reduction in pregnancy loss could stem from targeted investments in antenatal care services, improved monitoring, or more effective outreach enabled by gender-sensitive resource allocation. The lack of impact on fortified food may reflect supply limitations or lower prioritization of nutrition-specific interventions within GRB frameworks. The negative association with postnatal care might signal a gap in budget continuity across the maternal care cycle, or persistent structural barriers—such as mobility, time constraints, or lack of follow-up mechanisms—that limit post-delivery service uptake. The significant increase in state and national insurance coverage points to a stronger role for formal institutions in extending financial protection. These effects likely arise from gender-responsive reforms that streamline administrative processes, improve targeting of women, or earmark funds to subsidize enrollment. In this context, insurance coverage should be understood as part of a broader process of financial inclusion. Financial inclusion goes beyond banking access; it encompasses affordable and reliable financial services—credit, savings, and especially insurance—that help households manage risk and reduce vulnerability (van Hees et al. (2019)). By facilitating access to formal insurance schemes, GRB strengthens women's financial resilience (Habib et al. (2016)) and autonomy, empowering them to better cope with health shocks and reducing their dependence on informal safety nets. The absence of significant effects on community-based insurance may reflect the limited reach of GRB interventions in informal or decentralized structures that operate outside of state-led budgeting channels. These findings underscore the institutional nature of GRB's influence, enhancing financial inclusion through state systems while leaving room for further integration with community-level mechanisms.

The result's tables are available below, while table 20 in appendix summarize the expected effects.

Dependent variable:	`fortified food`	`pregnancy loss`	`post natal check-up`
	(1)	(2)	(3)
dose	0.006	-0.009**	-0.018***
	(0.004)	(0.004)	(0.003)
Observations	12,959	12,959	12,959

Table 14: Regression results (Standard errors in parentheses).

Insurance from:	States	Community	Country
dose	0.019***	0.065	0.034***
	(0.006)	(0.059)	(0.004)
Observations	12,959	12,959	12,959

Table 15: Results for insurance subscription

Note: *p<0.1; **p<0.05; ***p<0.01

6) Conclusion

Through this work, I have evaluated the effects of gender budgeting adoption on the efficiency of health public spending in Indian states over the period 1997–2020. Using difference-in-differences estimator and entropy balancing—an approach that combines matching with linear regression to mitigate endogeneity—I find that states adopting gender budgeting achieve higher efficiency scores than those that do not. One likely transmission channel identified is the strengthening of local administrative capacities, as outlined in Table 10. At the macro level, gender budgeting imposes a framework of continuous evaluation and accountability in fiscal and budgetary processes. This institutionalization of monitoring and goal-setting not only improves the overall fiscal framework but also enhances the coherence and effectiveness of public policy design and implementation. In doing so, gender budgeting emerges as more than just a tool to reduce gender disparities—it becomes a catalyst for better governance. Improved spending efficiency is particularly valuable in a context of fiscal constraints, as it enables subnational governments to allocate limited resources more effectively and improve service delivery quality. The effectiveness of this framework, however, appears to vary by timing of adoption, pointing to possible interactions with institutional readiness or political conditions.

At the micro level, complementary analysis of household and individual data shows that exposure to gender budgeting is associated with tangible changes in social outcomes. My findings suggest a reduction in pregnancy loss and improvements in insurance coverage at the state and national levels, although postnatal care appears to remain insufficiently addressed. These results point to a broader set of benefits tied to the adoption of gender-sensitive fiscal practices, including better health system engagement and improved financial inclusion for women. In particular, access to formal insurance schemes underscores how gender budgeting can shape financial protection mechanisms by reinforcing state-led enrollment and coverage processes.

In terms of policy implications, this analysis suggests that embedding clear objectives, designing targeted measures, and systematically evaluating their implementation—principles at the core of gender budgeting—can improve public spending quality across sectors. The moral imperative to address gender inequalities supports the institutionalization of these good practices, creating a replicable cycle of reform that can be extended to other policy areas such as urban development or environmental management. The durability of these gains, however, depends on strong political backing and the active involvement of civil society, both of which are essential for ensuring accountability. In contexts like India, where higher authorities can enforce compliance with gender budgeting norms, this oversight function reinforces the legitimacy and sustainability of the process. In contrast, in settings without centralized enforcement or strong civil society engagement, these mechanisms may be harder to replicate, limiting the transposability of my findings to the national or international level.

References

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

Adom, P. K., Agradi, M., and Vezzulli, A. (2021). Energy efficiency-economic growth nexus: what is the role of income inequality? *Journal of Cleaner Production*, 310:127382.

Afonso, A., Jalles, J. T., and Venâncio, A. (2021). Taxation and public spending efficiency: An international comparison. *Comparative Economic Studies*, 63:356–383.

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*.

Boetti, L., Piacenza, M., and Turati, G. (2012). Decentralization and local governments' performance: how does fiscal autonomy affect spending efficiency? *FinanzArchiv/Public Finance Analysis*, pages 269–302.

Borusyak, K., Jaravel, X., and Spiess, J. (2022). Revisiting event study designs: Robust and efficient estimation. *Available at SSRN 2826228*.

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

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

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

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

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.

Clots-Figueras, I. (2011). Women in politics: Evidence from the indian states. *Journal of public Economics*, 95(7-8):664–690.

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

De 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.

Erlangga, D., Suhrcke, M., Ali, S., and Bloor, K. (2019). The impact of public health insurance on health care utilisation, financial protection and health status in low-and middle-income countries: a systematic review. *PloS one*, 14(8):e0219731.

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.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Habib, S. S., Perveen, S., and Khuwaja, H. M. A. (2016). The role of micro health insurance in providing financial risk protection in developing countries-a systematic review. *BMC public health*, 16:1–24.

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.

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.

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,*.

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.*

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.

Pan, J., Lei, X., and Liu, G. G. (2016). Health insurance and health status: exploring the causal effect from a policy intervention. *Health economics*, 25(11):1389–1402.

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.

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.

Sharma, P. and Garg, P. (2014). Women empowerment through gender budgeting in india. *International Research Journal of Human Resources and Social Sciences*, 1(6).

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.

van Hees, S. G., O'Fallon, T., Hofker, M., Dekker, M., Polack, S., Banks, L. M., and Spaan, E. J. (2019). Leaving no one behind? social inclusion of health insurance in low-and middle-income countries: a systematic review. *International journal for equity in health*, 18:1–19.

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

7) Appendix

7.1) Diff-in-Diff

Table 16: Diff in Diff results by cohorts

Cohorts	eff score	eff score	eff score	eff score	eff score	eff score
2005	2.1241 ***					
	(0.7330)					
2006		0.2637				
		(0.5678)				
2007			2.7257			
			(1.7532)			
2009				1.5613		
				(0.7953)		
2014					0.4753***	
					(0.0419)	
2016						0.3527***
						(0.0672)
Observations	668	668	668	668	668	668

t statistics in parentheses

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

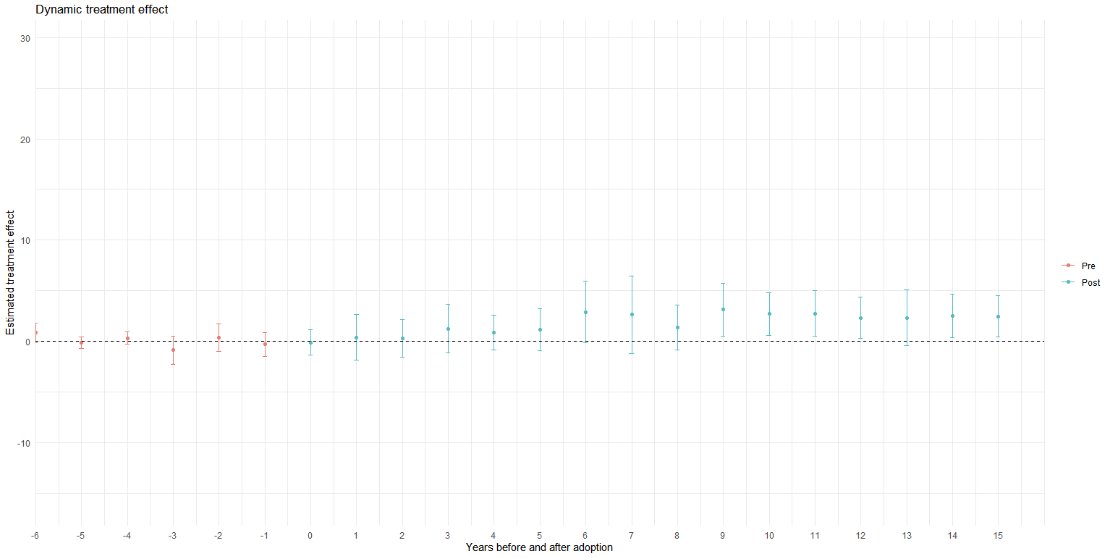
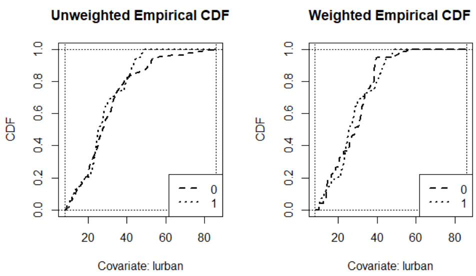
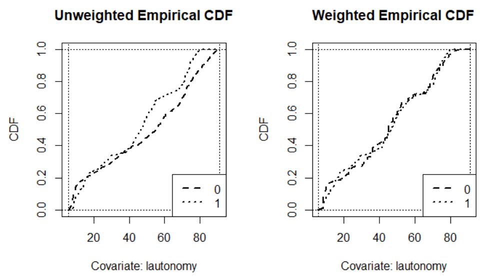


Figure 5: Diff-in-Diff event stduy

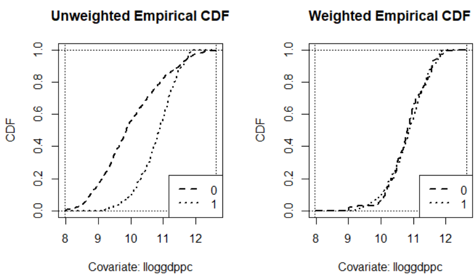
7.2) Entropy balancing



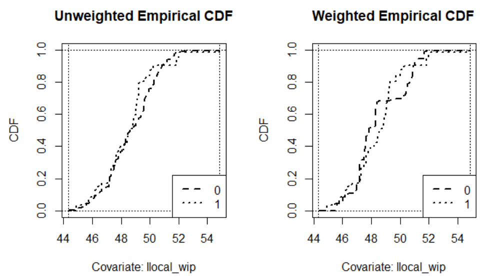
(a) Urbanization balance



(b) Autonomy balance



(c) GDPpc balance



(d) Wip Balance

Figure 6: Entropy banlancing graphs

7.3) Placebo test

Table 17: Results for the placebo test

	Dependent variable:
	efficiency score
placebo	0.122
	(0.436)
Observations	541
Note:t statistics in parentheses	*p<0.1; **p<0.05; ***p<0.01

7.4) Anticipation effects

Table 18: Results for the anticipation test

	Dependent variable:
	efficiency score
anticipation	−0.871
	(-1.122)
Observations	541
Note:t statistics in parentheses	*p<0.1; **p<0.05; ***p<0.01

7.5) Political fragmentation

Table 19: Entropy balancing results

Variables	efficiency score			
GRB*HHI	6.54	7.26**	10.00***	5.50*
	(1.20)	(2.33)	(2.85)	(1.83)
(GRB $\times$ HHI) ^{2}	-8.23	-11.6*	-12.00*	-10.3*
	(-0.848)	(-1.94)	(-1.95)	(-1.88)
parties FE	No	No	Yes	Yes
years FE	No	Yes	No	Yes
Covariates	Yes	Yes	Yes	Yes
Turning point	0.397	0.313	0.417	0.267
Observations	537			

t statistics in parentheses *p<0.1; **p<0.05; ***p<0.01

7.6) Microeconomic effects

Outcome	**Effect of Gender Budgeting Duration**
Postnatal check within 2 months	Negative
Pregnancy loss	Negative (fewer losses)
Fortified food access	No significant effect
Insurance subscription (state level)	Positive
Insurance subscription (community)	No significant effect
Insurance subscription (national)	Positive

Table 20: Summary of effects of time since gender budgeting adoption