Building credibility: How gender budget transparency boosts commitment to fiscal policy and women's empowerment in Indian states*

CAMARA Cheick†

August 19, 2024

Abstract

This study delves into the effects of adopting gender budgeting on fiscal policies within Indian states over the period spanning from 1991 to 2020. Employing a Difference-in-Differences (DiD) framework alongside alternative estimators, the research tackles concerns regarding endogeneity to ensure the robustness of its findings. The findings reveal that states embracing gender budgeting tend to allocate a larger share of their expenditures to sectors that promote gender inclusivity, notably education and infrastructure. Intriguingly, the adoption of gender budgeting appears to bolster the credibility of states in prioritizing pro-gender policies, facilitated notably by central government transfers, including Centrally Sponsored Schemes (CSS), which serve as vital transmission channels. However, while gender budgeting presents a promising approach to tackling gender disparities and bolstering social outcomes, its implementation may encroach upon fiscal autonomy, necessitating careful deliberation regarding resource allocation and intergovernmental relationships. The study's policy implications stress the significance of contextual factors and transparency in ensuring the effective and credible implementation of gender budgeting reforms, particularly within the diverse economic landscape of India. In summary, this research offers valuable insights into the nuanced impact of gender budgeting adoption on fiscal dynamics, highlighting the imperative for further exploration to navigate complexities and optimize policy outcomes.

Keywords: Regional Government, Regional Policy, Public Economics, Public Finance, Political Economy.
JEL Codes: H1; H7; J16; R5.

*I am grateful to Emilie CALDEIRA and Bertrand LAPORTE for their advices and guidance, as well as to seminar participants at CERDI Ph.D. seminars for useful comments and suggestions. The views expressed in this document are those of the author and do not necessarily reflect those of CERDI.
first draft do not quote.

†Université Clermont Auvergne, CNRS, IRD, CERDI, F-63000 Clermont-Ferrand, France.
Email: cheick_amadou.camara@doctorant.uca.fr. Website: camaracheick.com.

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. It has become very popular since its first adoption by Australia in 1974. In recognition of gender inequities, more than 80 countries have applied a gender perspective to their budget process, initiatives referred to as gender-responsive budgeting (GRB) in 2021 (Rubin and Bartle (2023)). 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. Gender budgeting allows fiscal authorities to ensure that tax and spending policies and/or public financial management instruments[1] address gender inequality and the advancement of women in areas such as education, health, and economic empowerment as explained by Budlender et al. (1998) Budlender and Hewitt (2003) Stotsky (2006). The approach is sometimes called gender-sensitive or gender-responsive budgeting. If well-designed, such budgeting can improve the efficiency and equity of the overall budget process. Fiscal authorities at any level of government can assess the needs of boys and girls and men and women; identify key outcomes or goals; plan, allocate, and distribute public funds, and monitor and evaluate achievements. While gender budgeting initiatives around the world have tended to focus on expenditure policies, revenue policies also influence gender equality and could therefore be considered part of gender budgeting. The next graph summarizes the gender budgeting framework.

[1] with the inclusion of a section on assessing gender responsive public financial management in the PEFA framework

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

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

In India, the adoption process began in 2005 with different adoption waves and following an adoption by the central government in 2000. The Union initiative was institutionalized through the development of a system of classification of budgetary transactions, and the formation of groups or cells in each Ministry of the government to lead efforts to identify gender-related goals and ways to achieve these goals through the budget (Stotsky and Zaman (2016)). Drawing upon the central government framework, states began to adopt gender budgeting, starting with Odisha in 2005. Since then, most Indian states have had some form of gender budgeting in place. Gender budgeting at the State level in India offers a suitable empirical framework for assessment of its effects because several states have adopted and sustained gender budgeting efforts (Chakraborty (2016) and Stotsky and Zaman (2016)). State-level gender budgeting in India has also used the national-level analytical matrices and templates of the National Institute of Public Finance and Policy (NIPFP) [2]. This framework implies an ex-post analysis of the budget through a gender lens and a Gender

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

Budget statements that summarize theState(or national) effort implemented to reduce the gender gap and/or reach the objectives. Gender budgeting as previously explained requires assessments and transparency for the objectives and the results of the public policies either to reduce the gender gap or to reach Sustainable Development Goals which must be a common objective for the Indian States

Several recent studies like Joshi (2013); Parvati et al. (2012) and Demery and Gaddis (2009) provide an overview and assessment of the success of gender budgeting efforts at the State level in India. However, none of these studies has undertaken formal econometric evaluation. These studies indicated that State governments have implemented gender budgeting using a variety of approaches. Some of these approaches draw on a State policy for gender-related goals and include gender budget statements in the State budget. The substantive focus of gender budgeting initiatives has varied across states but has generally focused on the identification of critical goals to improve girls' and women's education, health, and welfare, and to build State infrastructure. In addition to these, Buchmann et al. (2008); Kleven and Landais (2017); Mbodji (2023) and Cannonier and Mocan (2018) found that an increase in schooling, triggered by the program, has an impact on women's attitudes toward matters that impact women's health and on attitudes regarding violence against women. An increase in education reduces the number of desired children by women and increases their propensity to use modern contraception and to be tested for AIDS. While education makes women more intolerant of practices that conflict with their well-being, increased education has no impact on men's attitudes toward women's well-being. Okojie (1994); Sen and Östlin (2008) also found a relationship between health spending and status and gender inequalities. Khera (2016) and Das et al. (2015a) examine the impact of gender-related policies on relative rates of labor force participation. Both studies found that government policies that increase female education, social spending, and labor market flexibility raise women's labor force participation. Khera (2016) also finds that higher investment in infrastructure is beneficial. I do not examine this issue in my study because the empirical specification is designed to capture the way gender budgeting adoption affects fiscal spending composition, and because relevant labor market data are not available in our data set.

Although gender budgeting efforts, in the international context, now date back several decades, and despite its growing importance there has been little effort to assess the results of gender budgeting adoption quantitatively. Most studies except Stotsky and Zaman (2016) focused on gender budgeting adoption determinants or its effects, but from a qualitative perspective like Chakraborty (2016); Stotsky and Zaman (2016); Quinn (2016); Jung (2022) Puig-Barrachina et al. (2017) and Elomäki and Ylöstalo (2021)).

Stotsky and Zaman (2016) assesses the effects of gender budgeting on gender equity,

and fiscal spending (measured as the share of State GDP). The possible reallocation effects due to the gender budgeting adoption and the staggered adoption framework have not been treated or considered. The issue of public spending allocation at the national and local levels is an important matter. Gadenne (2017) found suggest that local governments use the increase in taxes thanks to the Brazilian local tax capacity program to provide more education infrastructure than they do when faced with an increase in transfer revenues of the same amount. Besley and Smart (2007) with his principal-agent models of public finance in which asymmetries of information allow politicians to capture more rents if we assume that citizens are better informed about increases in taxes than increases in transfers. The lack of information and transparency, allows politicians to capture rents and to have a discretionary use of public money. Gender budgeting adoption through the transparency and disclosure of information about revenues and expenditures reduces the information asymmetry between citizens and politicians. So, it's important to assess how gender budgeting adoption can affect public spending allocation.

To summarize, this paper holds on the assumption that Gender Budgeting adoption by improving transparency and reducing the information's asymmetry (Morrissey (2018); Alonso-Albarran et al. (2021) and Sarraf (2003)) through reporting on budget allocations and impact assessments, GB provides a platform for citizen engagement in gender policies will affect the public spending composition (Besley and Smart (2007), Mejía Acosta (2013), Moschovis (2010)). This study thus adds an important dimension to research on gender budgeting by focusing on these reallocation mechanisms and by considering the staggered adoption framework. This paper aims to assess the overall and potential heterogenous effects of gender budgeting adoption on public spending allocation, to know if the timing and adoption context matter or if the effects are the same for all the adoption waves. The next graph shows an Indian map where the different adoption waves are represented.

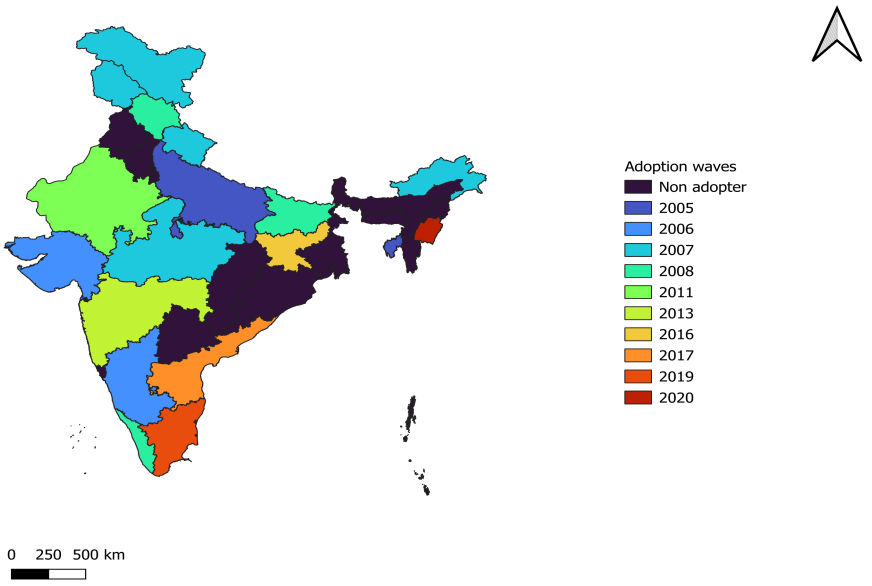


Figure 2: GB adoption waves (made by the author)

Using state-level panel data, with staggered Differences in Difference and alternative estimation tools, The paper shows that gender budgeting adoption positively affects the share of public spending devolved to gender inequalities reduction according to state governments statements. The results remain positive and significant over time regardless of the estimation tool. This positive effect also affect positively women situation at individual level.

The rest of the article is structured as follows: section 2 presents a brief overview of the Indian federal system and gender budgeting adoption. Section 3 presents the theoretical framework upon which the empirical analysis of the paper relies. The methodology, with the data, identification strategy, and the analysis of the parallel trend assumption are presented in section 4. Section 5 presents the first results of gender budgeting adoption with the Difference in Difference (DiD) staggered estimators of Callaway and Sant'Anna (2021). The results of the additional robustness analyses with alternative DiD estimators, entropy balancing, and other robustness checks are also presented in this section. Section 6 analyzes the effect at a microeconomic level, while section 7 tries to assess the potential transmission channel. Finally, section 8 concludes.

2 Indian federal system and Gender Budgeting adoption

As explained by Shair-Rosenfield et al. (2021), India is a federal republic composed of 28 States (some with asymmetry in shared rule) plus 5 standard union territories, 2 non-standard union territories, and one special autonomous region (the State of Jammu and Kashmir) which had lost its special autonomy status in 2019. India became independent from the British in 1947, the first constitution was enacted in 1950 and, as of 2019, there have been 102 amendments. Indian elites have struggled to balance internal diversity with a cosmopolitan identity and centralization with decentralization. On the whole, they have leaned to the side of centralization (Swenden (2015) and Jennings (1953)). Unusual for a federation, the constitution is very detailed in prescribing the internal organization of each constituent unit. Also, the national parliament can unilaterally change State boundaries as well as the conditions governing State representation in the Senate. And, except for the State of Jammu and Kashmir, which (until 2019) had its own Constitution, the rest of the States have no separate constitutions of their own.

From 1957, each State had a directly elected parliament, named the legislative council, and a centrally appointed governor. The governor holds executive power. He in turn appoints a council of ministers, with the chief minister at the head. The chief minister is the head of the local majority at the State parliament. The governor, on the advice of the chief minister, appoints members of the council of ministers– the cabinet that advises the chief minister. Despite the power vested in the chief minister, which includes advising the governor and chairing the legislature and the council of ministers, the governor is more than a figurehead. They have considerable authority, including the right to dismiss State governments' discretion over a state's contingency fund, the ability to enter into and execute contracts, the ability to grant pardons and suspend, remit, or commute sentences of convicted persons, and in some cases, can appoint a small subset of State legislators. States have exclusive authority over local police, education, welfare (pensions, unemployment, and disability), the economy (trade and commerce within the state, money-lending, and lenders), and culture (theatres, sports, religious societies, libraries, museums, and other similar institutions controlled or financed by the State; ancient and historical monuments and records).

The National Capital Territory of Delhi (NCT) was controlled by a chief commissioner appointed by the central government from 1947 until 1951. From 1952 to 1956, the NCT was authorized to create a directly elected 48-member unicameral assembly and a chief minister of Delhi selected by the assembly. In 1956 the States Reorganization Act eliminated the assembly and the chief minister of Delhi and replaced both with a union minister appointed by the president of India. This was the beginning of a long period of President's Rule. In September

1966 President's Rule was softened with the Delhi Administration Act, which restored a weakened version of the prior representational structure. The new dual structure consisted of an advisory Delhi Council, with 56 directly elected members, 5 appointed members, and a centrally appointed lieutenant governor. The Constitution Act of 1991 created a directly elected legislative assembly and an executive council of ministers elected among assembly members and chaired by a chief minister of Delhi. While in the states the chief minister is appointed by a centrally appointed governor, the chief minister of Delhi is appointed by the president. The position of lieutenant governor continues to exist and now resembles the position of State governor.

The next graph represents the repartition of States and Union Territoiries in 2021.

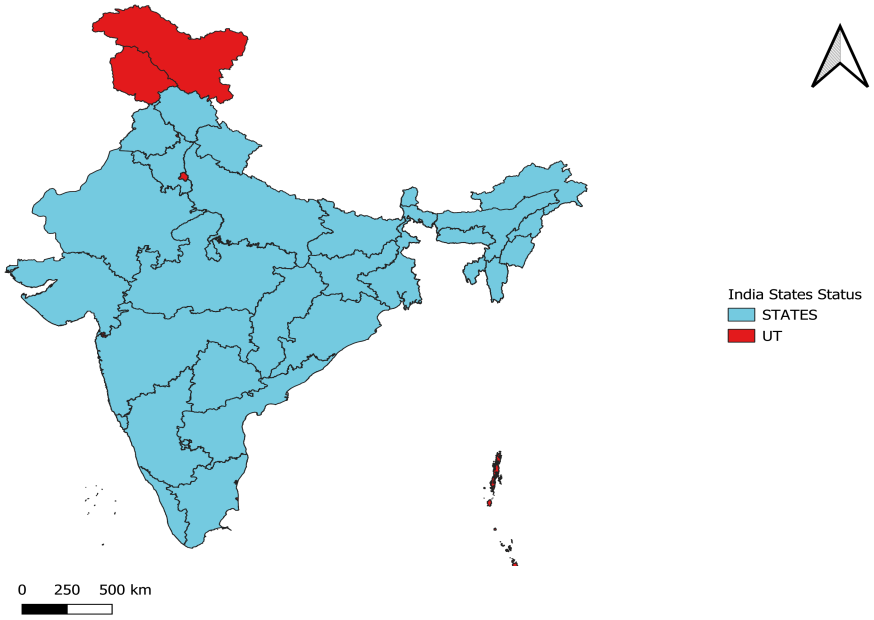


Figure 3: Indian States vs UT

Gender Budgeting is a tool for gender mainstreaming, applying a gender lens to the entire policy process. It involves gender-sensitive formulation, resource allocation, and continuous monitoring to address vulnerabilities faced by women throughout their life cycles. Gender budgeting was first adopted at the Union level in India in 2000 (Stotsky and Zaman (2016)) and then the states followed. 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. Drawing upon the central government framework, states began to adopt gender budgeting, starting with Odisha in 2005. Since then, another

27 states have adopted gender budgeting efforts as well. States are largely responsible for education, health, infrastructure, and economic services according to Rao (2009). So, in general, they focused their initiatives on social services sectors such as women and child development, education, health, nutrition, welfare, and infrastructure. Gender budgeting efforts across these states include the development of a State policy for gender equality and adoption of gender budgeting statements in budget documents; identifying nodal departments for gender budgeting and formation of gender budgeting cells; implementation of policies and programs related to gender-related objectives; and data collection and analysis.

3 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 $\delta \in$, 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 weigh 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 allocate more fiscal space to sectors that contribute to reducing gender inequalities.

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 results obtained.

4 Empirical methodology

4.1 Data

The data used in this paper come from the Reserve Bank of India datasets. These datasets provide an interesting and wide range of detailed data about the Indian States and Union Territories (UT).

First of all, to assess the potential reallocation effect of gender budgeting on public spending, I compute a "pro gender" variable. It's the addition of all public spending that is related to gender equality according to the literature and is a part of priority areas defined by States and UT in their gender budgeting statements. This pro gender variables, as the other expenditures variables are expressed as a percentage of States/UT total expenditures. To check the potential reallocation effects. The "pro-gender"variable will be like this:

$$Pro\_gender_{it} = \frac{(Education_{it} + Health_{it} + Water_{it} + Infrastructures_{it})}{Total\_expenditure_{it}} * 100 \quad (3)$$

The infrastructure spending contains the roads, bridges, and physical capital expenditures. Parikh et al. (2015) provided evidence that infrastructure investment is critical for well-being of slum dwellers and women in particular. Das et al. (2015b) also found that higher investment in infrastructure is beneficial for women empowerment and gender equality in India. Indeed, it found that adequate infrastructure and education levels predict higher female entry and that there are strong agglomeration effects in both manufacturing and services sectors. In similar fashion Calvo et al. (1994) found that, in Zambia, women spent over 800 h per year gathering and transporting firewood, while their male counterparts spent no more than 50 h per year on these activities. More generally, available data suggest that, on average, women in rural Sub-Saharan Africa spent between 0.9 and 2.2 h per day on transporting water and firewood (see Weiss (1999)).

The next table summarizes the "pro-gender"variable and its components.

Table 1: outcomes Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.
pro_gender	5.193	2.858	0.338	16.794
infrastructures	2.311	1.474	0.053	12.235
education	0.73	0.852	0	5.517
health	0.582	0.626	0	5.849
water	1.534	1.461	-0.004	7.805
N		857		

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 2: Repartition of treatment

Treated	125
Untreated	444

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

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

These states are also less dependent on central government transfers and are more autonomous in their political choices. The fiscal autonomy variable is a ratio between states' own local revenues and their total revenues (transfers and grants included).

The most urbanized states can generate some scale economies, or sometimes some congestion effects which make less effective and less efficient public spending and policies related to health and education issues. To avoid this, more urbanization may call for more expenditure centralization by attracting the rural population towards big centers, favoring a certain concentration of public expenditures(Sacchi and Salotti (2016)).

GDP per capita affects the sensibility to women rights and gender equality. Indeed, Doepke et al. (2012) found a strong connection between women's rights and economic

development.

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 3: Controls summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
gender_budgeting	0.22	0.414	0	1	1325
state autonomy	48.938	25.713	5.466	100	887
log(GDP per capita)	10.287	1.061	7.886	12.832	942
Urbanization (%)	33.568	19.098	7.98	99.900	870
trend	16.815	9.352	1	33	1325
women in parliement (%)	48.53	1.878	42.39	54.87	902
fiscal rule	0.381	0.486	0	1	1325
Agriculture (%GDP)	26.575	20.109	0.052	130.834	930
log(population size)	20.847	0.125	20.608	21.025	843

4.2 Stylized facts

The graphs 4 and 5 highlight a comparison between the average “pro gender” expenditures share for the treated (1) and untreated units (0). The adopters seem to allocate more of their funds to "pro-gender"items than the non-adopters. This result seems to suggest that gender budgeting adopters spend more on "pro-gender"items than others. However, this correlation means nothing in terms of causes and consequences because a correlation does not necessarily imply causality.

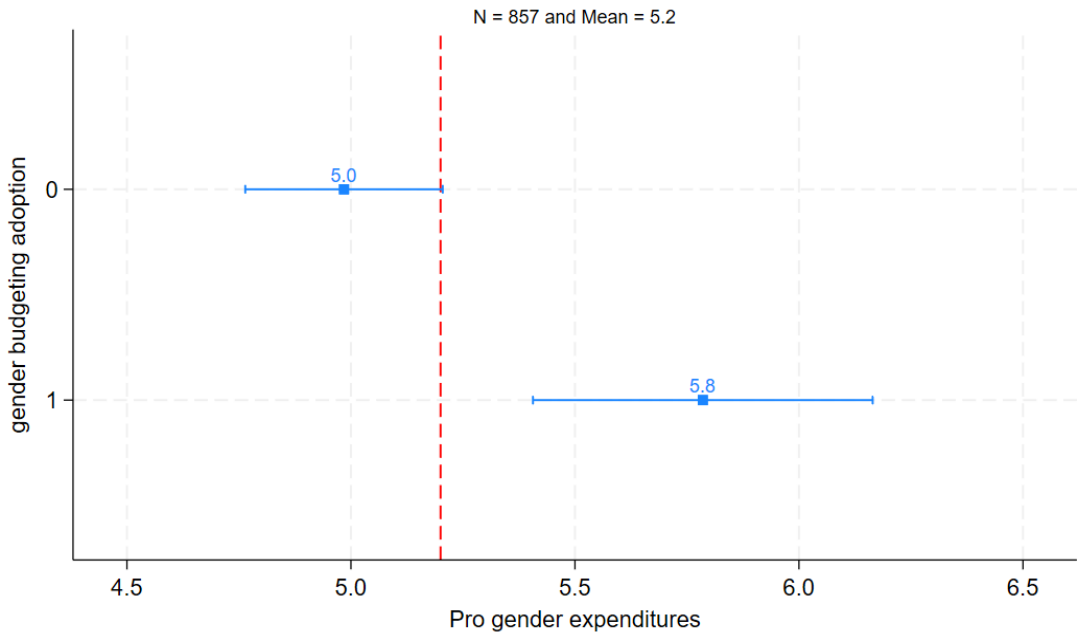
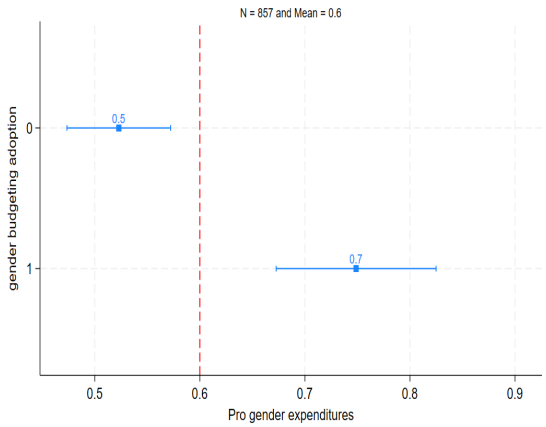


Figure 4: Pro gender spending by treatment status

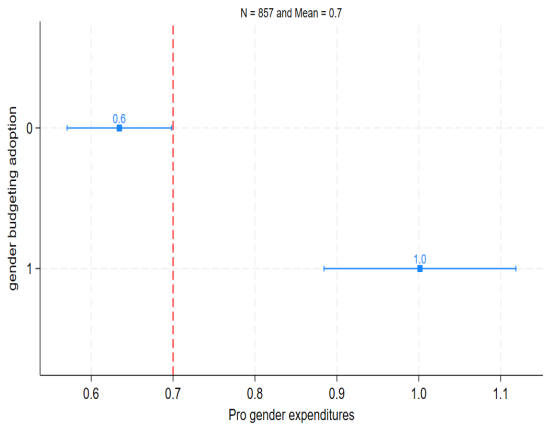
Always to analyze the potential difference in the average share of total expenditures between treated and untreated units, I computed the same statistics and made the same graph for the different components of "pro gender" expenditures.

The graph 5 summarizes the average public expenditures for untreated (0) and treated units (1) by comparison to the global average of the sample (the vertical red line). The difference between both means suggests that the treated units seem to spend more for the three items (health, infrastructure, and education) than the untreated ones. In addition, their average expenditures in these items are more important than the average of the global sample. However, the situation is different for the water and sanitation services. Indeed, for this item, the treated units spend less on average on this item than the untreated ones. So, maybe the reallocation of public money is made to the detriment of water and sanitation services. However, I have to be cautious about this conclusion and the interpretations and need further investigations.

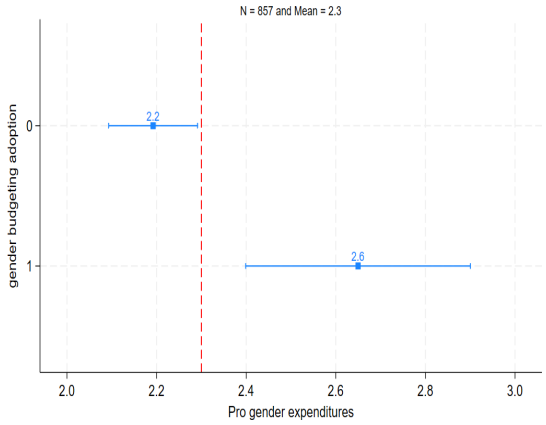
All these elements tend to suggest that States that have adopted gender budgeting spend significantly more than the others. However, I can conclude that this difference is due to the gender budgeting adoption. The difference can be due to a simple correlation between the variables or to the fact that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



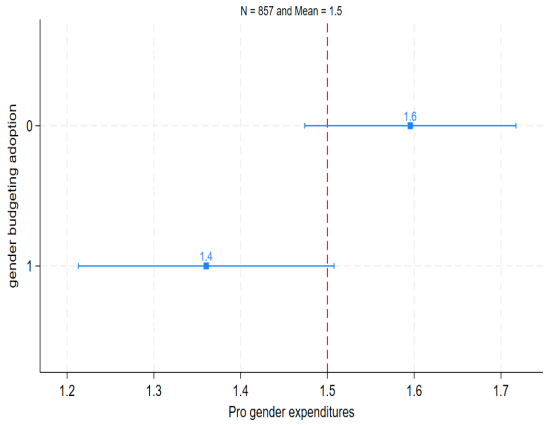
(a) Average expenditures share for health



(b) Average expenditures share for education



(c) Average expenditures share for infrastructures



(d) Average expenditures share for Water and sanitation services

Figure 5: Stylized facts

The last graph summarizes the evolution of "pro-gender" components (on average) through time.

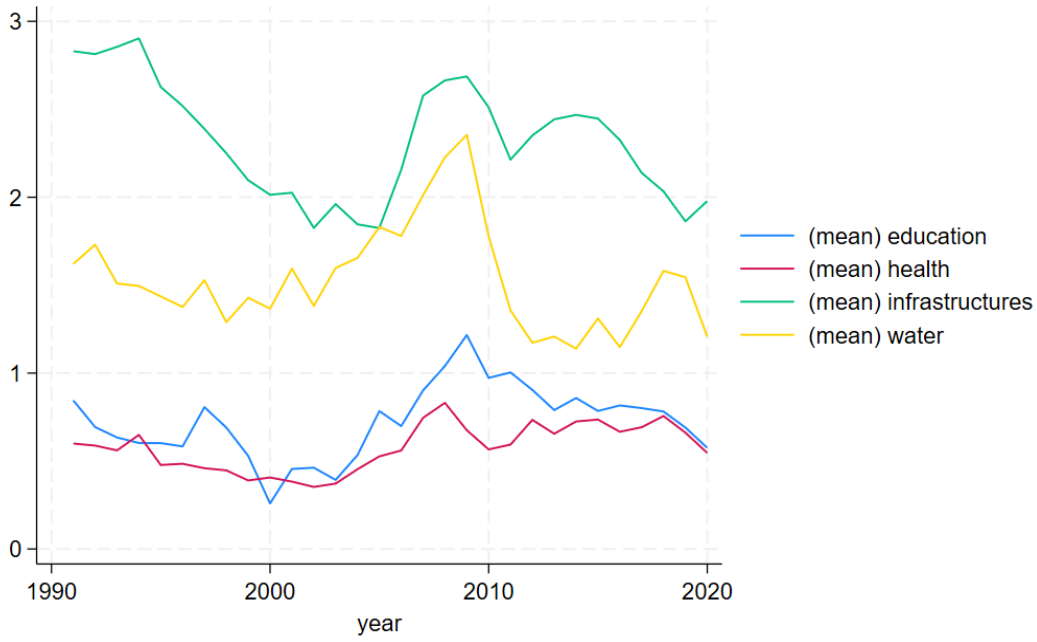


Figure 6: Pro gender spending components evolution

The graph 7 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 to gender equity concerns across the country.

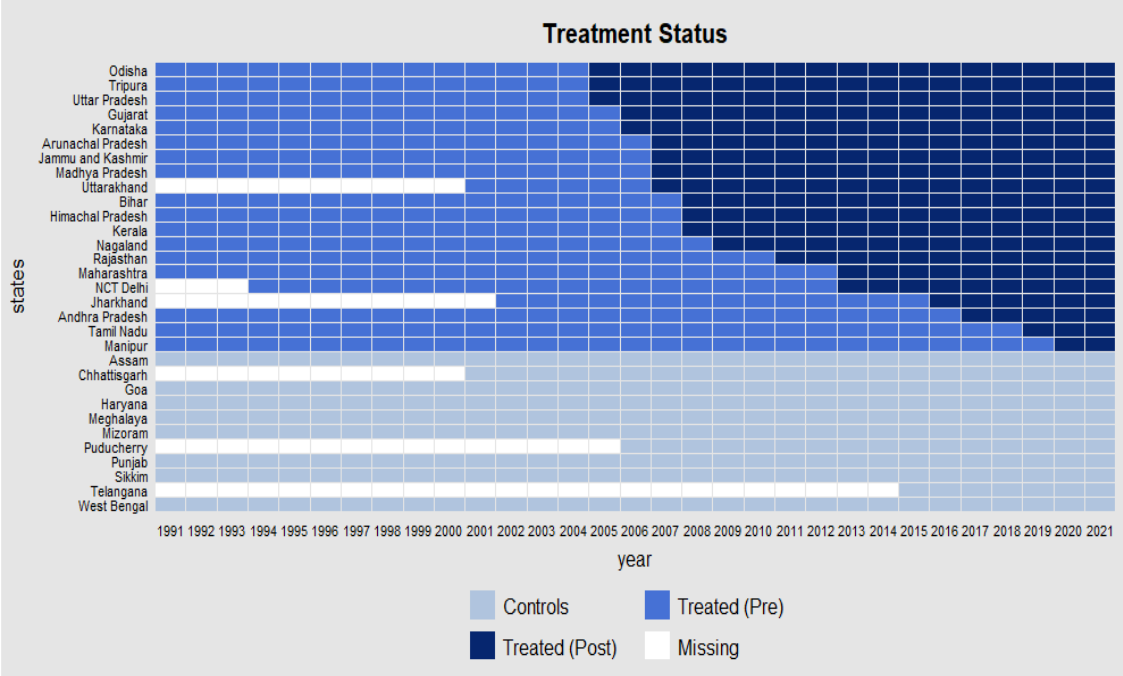


Figure 7: Gender budgeting adoption by states

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.

4.3 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_it is the outcome variable, i.e., public expenditures for State i in year t = 1, ..., T

$$Y_{it} = \alpha + \beta^W {D_{it}}^W + \rho X_{it} + \theta_i + \gamma_t + \epsilon_{it} \tag{4}$$

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.

4.4 Parallel trend assumption

The graph 8 highlights the evolution of spending allocated to "pro-gender"items for treated and untreated units through time. This concern about the parallel trend is crucial to confirm the internal validity of the Difference in Difference (DiD) Method. it provides some reassurance that untreated units could provide a reasonable counterfactual, particularly if they most closely resembled the treated ones. It also allows us to reduce the concerns about the fact that our results could be driven by pre-trend effects which would bias our results. This graph allows us to check if the pre-trend assumption holds a few years before the treatment was applied as used in Callaway and Sant'Anna (2021), and other DiD estimators (applied in robustness check) which suggests a weaker pre-trend assumption about the existence of parallel pre-trend between treated and untreated units.

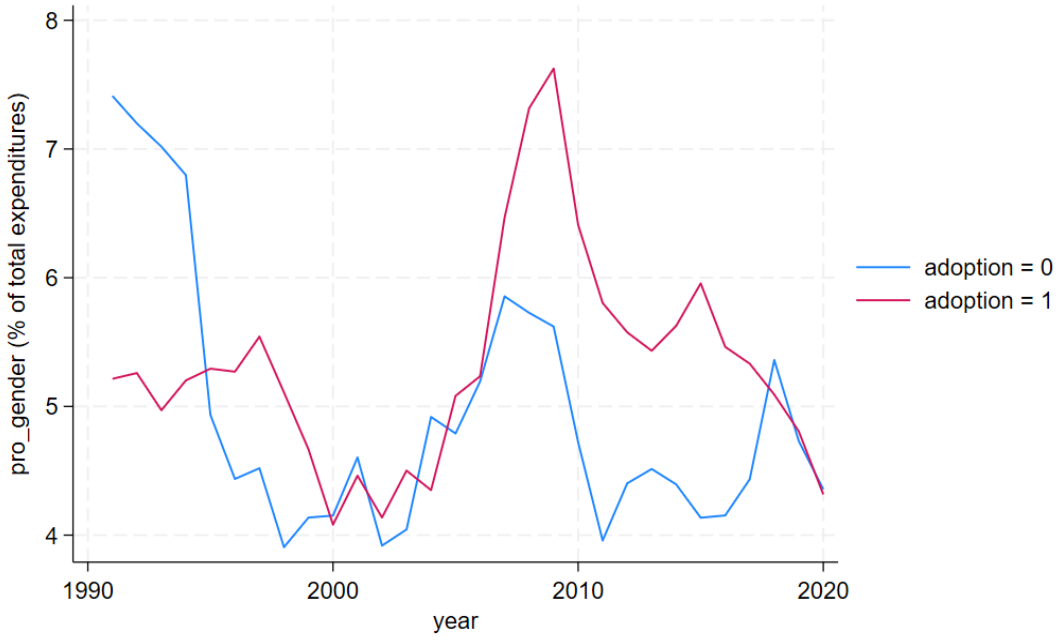


Figure 8: Evolution of "pro gender" spending share through the time

In addition, we 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 is relatively close for both (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	Adopters	non Adopters	Difference
Pro gender spending	5.04	4.911	
Education	0.71	0.54	
Health	0.50	0.55	
Infrastructures	2.26	2.11	
Water	1.52	1.68	
After adoption			
Outcomes	Adopters	non Adopters	Difference
Pro gender spending	5.72	4.76	***
Education	0.97	0.61	***
Health	0.73	0.53	***
Infrastructures	2.48	1.92	***
Water	1.49	1.68	*

Table 4: Outcome means before the treatment (by status)

The following graph (9) has been inspired by the work of Rambachan and Roth (2023) on a more credible approach to the parallel trend assumption. They propose some tools for robust inference in difference-in-differences and event-study designs where the parallel trends assumption may be violated. Instead of requiring that parallel trends hold exactly, they impose restrictions on how different the post-treatment violations of parallel trends can be from the pre-treatment differences in trends ("pre-trends"). They recommend that researchers use their methods to construct robust confidence intervals, under restrictions on the possible violations of parallel trends that are motivated by domain knowledge in their empirical setting. According to them, there are some key concerns about the pre-trend assumption. Despite the statistical or visual results, it's important to consider some macroeconomic shocks that can disturb the pre-trend evolution. Figure 9 shows robust confidence sets for the treatment effect, using different values of Mbar [3]. The figure shows that if I impose Mbar$<$1, meaning that I restrict the post-treatment violations of parallel trends to be no larger than the maximal pre-treatment violation of parallel trends, then I obtain a robust confidence set for the causal effect on the expenditures share. This is wider than the original (without covariates) confidence interval, which is only valid if parallel trends hold exactly, but rule out

[3]Mbar is a degree of smoothness, or how much I allow a violation of pre-trend assumption

a null effect on expenditures share.

The intuition for why the confidence sets are larger through time is that I have bound the violation of parallel trends across consecutive periods by Mbar times the max in the pre-treatment period. Thus, the identified set will be larger for later periods, since the treatment and control groups have more time to diverge. If I am willing to bound the magnitude of economic shocks by the max in the pre-treatment period, I will thus typically obtain wider confidence sets for parameters involving later periods. As suggested by Rambachan and Roth (2023) the table 9 available in the appendix summarizes the different bands of confidence interval according to the Mbar values.

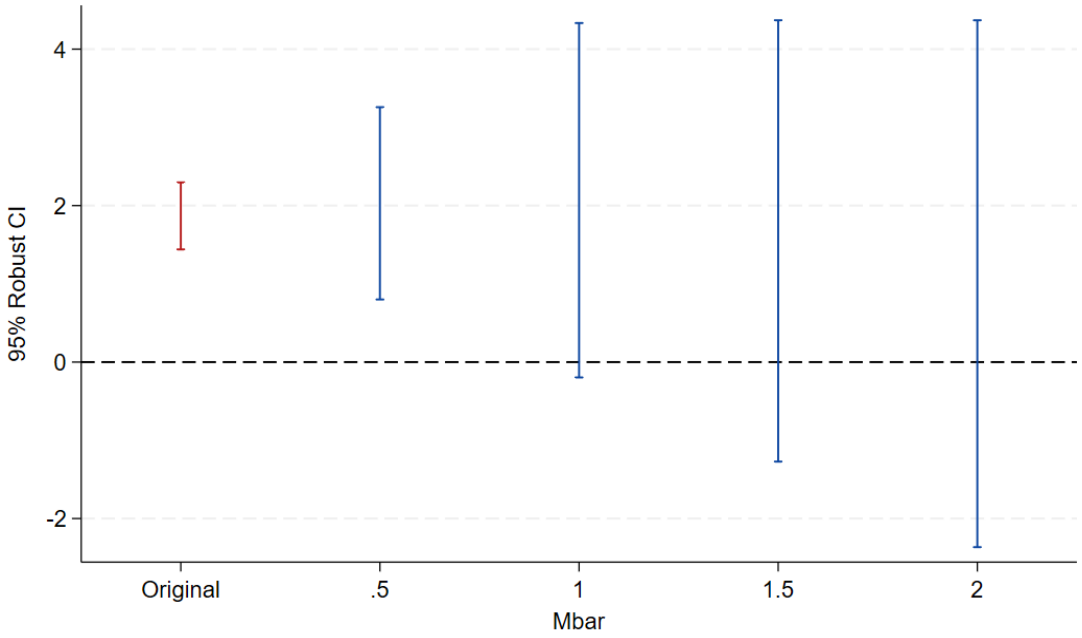


Figure 9: Parallel trend with honest DiD (Rambachan and Roth (2023))

4.5 Results

The results of the ATT are available in the next table. These results suggest a positive effect of gender budgeting adoption on the share of gender-friendly spending and some specific items (infrastructure and education) for the Indian states. This result seems coherent with the fact that education and infrastructure could be some important concerns for women and gender equality and are part of targeted items by State governments to achieve their objectives to reduce gender inequality. In addition, Montes et al. (2019) show that the potential positive effects of fiscal transparency can positively affect the efficiency and allocation of public spending. However, these effects are mediated by the level of democracy

in the country. In detail, in democratic countries, a higher degree of disclosure of fiscal information is correlated with a higher efficiency of government spending while, in non-democratic countries, fiscal transparency does not seem to provide any effect. The fact that India is a democratic country with some alternative change of government at the State level could be an explanation for the decision to reallocate some additional spending to specific sectors. Indeed, fiscal transparency can be a powerful device where politicians can be held accountable for their actions, while it could fail to provide positive results and avoid an under-provision of public goods where strong and effective vertical accountability is missing (De Simone et al. (2019) Gavazza and Lizzeri (2009) and Montes et al. (2019)). By imposing the disclosure of public financial information through the *gender budgeting statement* the implementation of gender budgeting led to a more transparent process which increased and improved the political rulers' accountability. So, they are urged to make some efforts and put in place some actions in the specific sectors targeted as gender friendly to reach their objectives.

Table 5: Diff in Diff results

	pro gender	education	health	infrastructures	water
ATT	1.706**	0.371***	0.219	1.040***	0.102
	(2.41)	(3.59)	(1.28)	(2.69)	(0.25)
Observations	668	668	668	668	668

t statistics in parentheses

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

In addition, I compute an event-study graph to assess the effects of gender budgeting adoption over time. Results suggest an overall positive effect of gender budgeting adoption. Indeed, the DiD event study clearly shows a positive effect of gender budgeting adoption on "pro-gender" public spending. States that have adopted gender budgeting seem to devote more space to "pro gender" spending, particularly on education and infrastructures (c.f tableB.2). This positive result also seems to be sustained through time by the adopters. So, the States that adopted gender budgeting seem to allocate a greater share of their expenditures to health and education than those that have not adopted it.

To be sure that this sustained effort is the same for all the adopters, we estimate the cohort effects of gender budgeting to check the potential heterogeneous effects of gender budgeting.

The following graphs summarise these event-study results

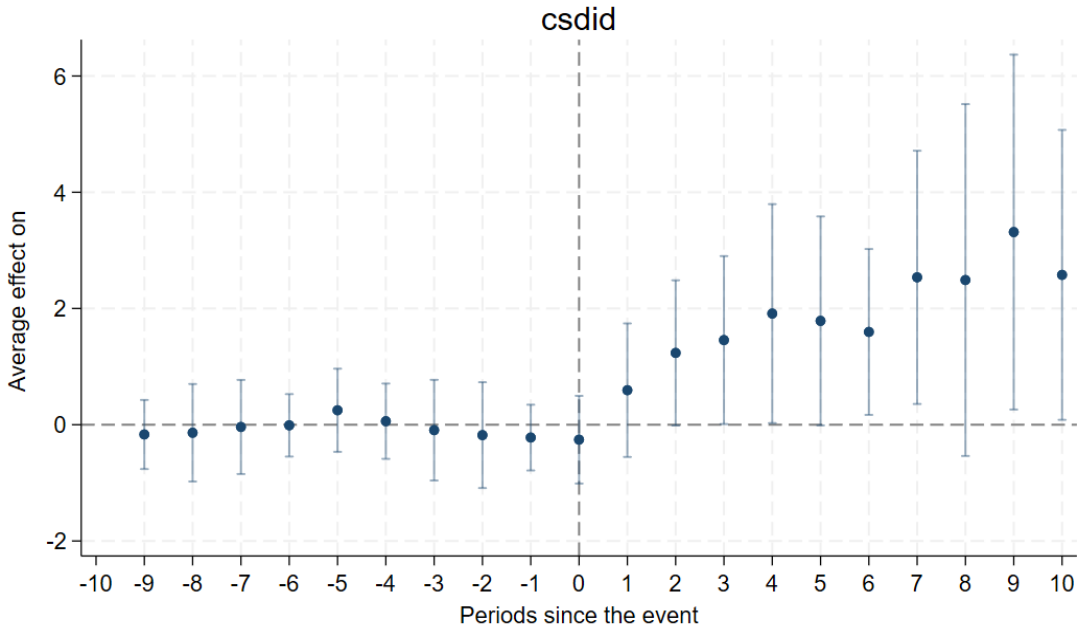


Figure 10: Event study results for "pro-gender"spending

5 Robustness Checks

5.1 Alternative DiD estimator

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

The point for him is not to conclude that other recent approaches – such De Chaisemartin and d'Haultfoeuille (2020), Callaway and Sant'Anna (2021), Borusyak et al. (2022), among

others – are not valuable and cannot improve over flexible TWFE methods. However, he is recommending not abandoning simple regression approaches because they can identify the treatment effects of interest very generally and can be made very flexible.

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

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

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

Table 6: Wooldridge Diff in Diff results

	pro gender	Education	Health	Infrastructures	Water
gender budgeting	1.151 *	0.387 ***	0.0921	0.912***	-0.229
	(1.68)	(5.02)	(0.51)	(3.47)	(-0.60)
Observations	668	668	668	668	668

t statistics in parentheses

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

The results seem to suggest a positive and significant effect of gender budgeting adoption on the share of public spending devolved to pro-gender, education, and infrastructure public spending.

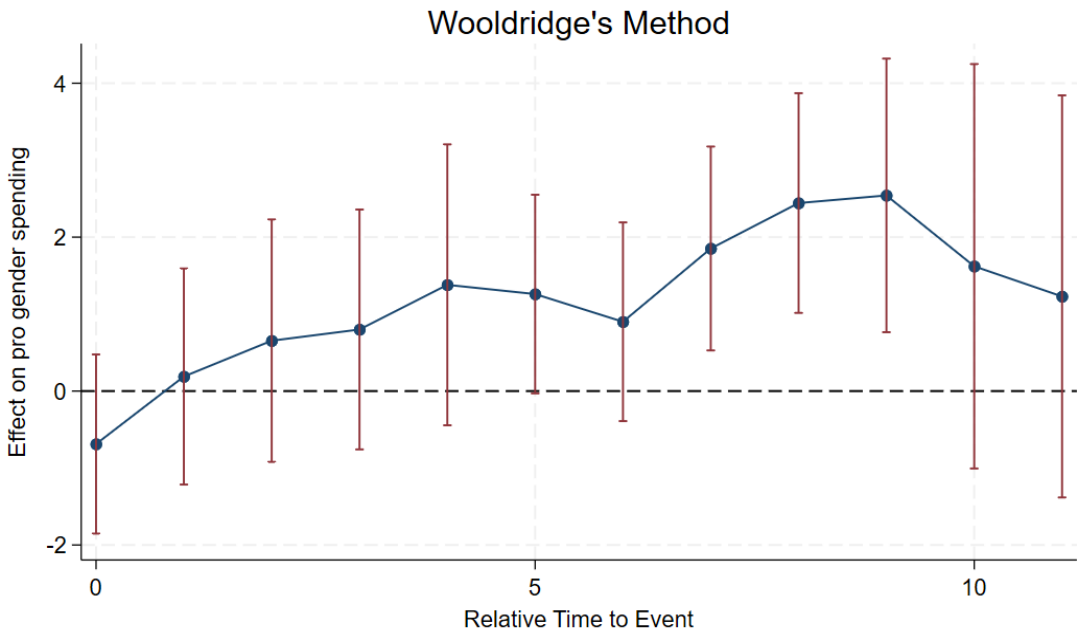


Figure 11: Results for Wooldridge estimator

There is a huge literature about the DiD estimators, but to check the robustness of our analysis, in addition to Wooldridge (2021) we will consider the estimators proposed by Borusyak et al. (2022) and De Chaisemartin and d'Haultfoeuille (2020).

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

dates depends on groups' realized treatments. Overall, this discussion suggests that whether the estimators in Borusyak et al. (2022) should be preferred to Callaway and Sant'Anna (2021) may depend on one's degree of confidence in the parallel trends assumption, on the type of violations of this assumption that seems more likely to arise in the application at hand, and on one's willingness to undertake a sensitivity analysis such as the one proposed by Rambachan and Roth (2023). The event study graph of the Borusyak et al. (2022) estimator is available in the graph B.3 in the appendix.

In addition to the previous DiD estimators, we use the estimator of De Chaisemartin and d'Haultfoeuille (2020). De Chaisemartin and d'Haultfoeuille (2020) propose an estimator that can be applied when treatment turns on and off and consider settings where units are treated at different times, but do not necessarily require that treatment is an absorbingState(Roth et al. (2023)). Their estimators intuitively compare changes in outcomes for units whose treatment status changed to other units whose treatment status remained constant over the same periods. This approach yields an interpretable causal effect under generalizations of the parallel trends assumption and an additional "no carryover" assumption that imposes that the potential outcomes depend only on the current treatment status and not on the full treatment history. In addition, De Chaisemartin and d'Haultfoeuille (2020) allows for both binary and continuous treatment. This particularity makes it very interesting to assess the effect of gender budgeting adoption (binary) and gender budgeting implementation experiences (continuous treatment). We define the continuous treatment by the time since the gender budgeting first adoption. The assumption here is that gender budgeting is a continuous process, so its adoption will affect the budgetary process over time. The first adopters will be more used to the process than the later ones or the non-adopters.

The results for both approaches are available in the appendix section in graph B.4 for the binary treatment and graph B.5 for the continuous one.

5.2 Entropy Balancing

5.2.1 Methodological concept

For the robustness check, I also use the entropy balancing method of Hainmueller (2012) like Baccini et al. (2018) who worked on fiscal decentralization and tax competition between local jurisdictions. Because many macroeconomic shocks have been able to change the expectations of the population,State rulers, or local administrations. The announcement of the gender budgeting adoption could also raise the expectations of the population in terms of public service quality. At the same time, theState rulers could have the incentive to improve the quality of their public spending to get the people's favor, even if they have not adopted gender

budgeting. The competition effect can affect the pre-trends and bias the results. For this reason, I use several DiD estimators which are less restrictive on the pre trends assumptions, but also the entropy balancing of Hainmueller (2012) which doesn't require the pre-trend assumption.

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

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

Entropy balancing is implemented in two steps. First, weights are computed that are assigned to units not subject to treatment. These weights are chosen to satisfy pre-specified balanced constraints involving sample moments of pre-treatment characteristics by remaining, at the same time, as close as possible to uniform base weights. In our analysis, the balance constraints require equal covariate means across the treatment and the control group, which ensures that the control group contains, on average, units not subject to treatment that are as similar as possible to units that received treatment. Second, the weights obtained in the first step are used in a regression analysis with the treatment indicator as an explanatory variable. This yields an estimate for the Average Treatment on Treated (ATT), that is, the conditional difference in means for the outcome variable between the treatment and control group. The advantage of entropy balancing over the other treatment effects methods is the fact that entropy balancing is not a parametric method. Indeed, this method does not need a specific empirical model for either the outcome variable or selection into treatment needs to be specified. Hence, potential types of misspecifications like those, for instance, regarding the functional form of the empirical model, which likely leads to biased estimates, are ruled out.

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

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

$$\pi ATT(x) = E[Y(1)|T = 1, X = x] - E[Y(0)|T = 0, X = x] \tag{5}$$

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

As pointed out by Neuenkirch and Neumeier (2016), entropy balancing has several advantages over traditional matching methods. First, unlike the propensity score matching methods or the difference-in-differences estimator, entropy balancing is a non-parametric approach, thus requiring no specification of the functional form of the empirical model or the treatment assignment procedure, which may avoid specification errors or collinearity problems. Second, entropy balancing ensures a sufficient balance of pre-treatment characteristics between treatment and control groups, even in the presence of a small sample or a limited number of untreated units. This makes it possible to construct a suitable control group, representing a near-perfect counterfactual of the treated group. Finally, in the second step, the estimator exploits the longitudinal nature of the data by including individual and time effects to control for unobserved heterogeneity across units and biases due to changes over time, independent of treatment. Tübbicke (2022) and Zhao and Percival (2017) also show that entropy balancing is doubly robust concerning linear outcome regression and logistic propensity score regression. It reaches the asymptotic semiparametric variance bound when both regressions are correctly specified. They suggest that entropy balancing is a very appealing alternative to the conventional weighting estimators that estimate the propensity score by

maximum likelihood.

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

$$Y_{i_t} = \beta_1 T_{it} + \alpha_1 log(GDP\_pc)_{it} + \alpha_2 log(density)_{it} + \alpha_3 X_{it} + \mu_i + \psi_t + \epsilon_{it} \tag{6}$$

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

5.2.2 Correlation issue

Table 12 shows a simple comparison of pre-weighting sample means of all matching covariates between treated (Column [2]) and control (Column [1]) states, which represent the potential synthetic group. Column [5] shows significant differences between the two groups for all pre-treatment variables, as some p-values are below the threshold of 5%. Such differences could bias the true treatment effect due to a potential selection problem. Therefore, in Panel B (Column [1]), I compute a synthetic control group by re-weighting the control units, using the pre-treatment covariates from the benchmark specification. This approach allows us to make the means of the pre-treatment covariates of the synthetic group as comparable as possible to those of the treated units. As can be seen in Column [5] of Panel B, the weighting eliminated any significant pre-treatment difference between the means of the treated and synthetic covariates. Thus, I can consider the synthetic group as a perfect counterfactual of the treated group.

Table 7: Balance Statistics

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

5.2.3 Results

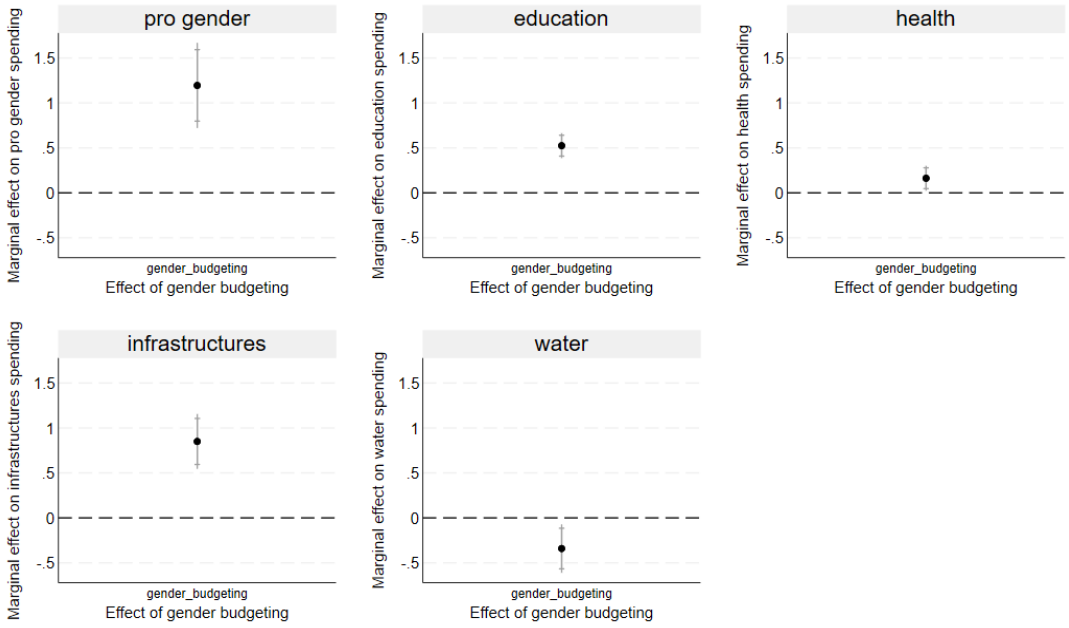


Figure 12: Entropy balancing Results

5.3 Dose Response Function

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

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

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

adoption waves, will get more experience in the gender budgeting adoption than the members of the next cohorts. The results are available in the table C.1 in the Appendix section.

5.4 Placebo Test

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

5.5 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)) By construction, De Chaisemartin and d'Haultfoeuille (2020) use a placebo to estimate the pre-trend coefficients to assess the evolution of outcome if they were not treated. So, the fact that these coefficients are not significant during the three first periods before the adoption could mean an absence of anticipation effects.

However, we 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 by using Callaway and Sant'Anna (2021) are presented in the appendix section at the table E.1.

The results show a non-significant effect for our alternative adoption waves. 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 education spending allocation.

5.6 Narrowing the control window

Finally, the effect captured in this work may suffer from some problems. Indeed, gender budgeting adoption can lead to a change in States' environments. In this sense, the effect captured may not be due to gender budgeting but to changes in institutional, political, social, or economic conditions after its adoption. Also, any other characteristic that may determine gender budgeting adoption may be a source of endogeneity. To circumvent these problems, we employ a similar approach as Neuenkirch and Neumeier (2015), Apeti (2023), and Apeti and Edoh (2023) by removing all observations before and after the initial year of adoption. Thus, we expect that this narrow time window characterizing our new treatment variable should provide a more robust estimate of its effect on public expenditures since the (generally slow-changing) institutional, political, social, and economic environment is more likely to be stable over a narrow time window. In total, we explore the robustness of our findings with two modifications to the sample period. In addition to the first adoption period, we consider (i) a window of five years around it, and (ii) a window of three years around it.

Using entropy balancing with this narrow time window, table **??** provides results that reinforce our previous findings. Thus, we can conclude that it seems unlikely that the estimated effects of gender budgeting are due to a coincidental change in the institutional, political, social, and economic environment in the gender budgeting adopters States' or to any other characteristics that may predict its adoption.

6 Microeconomic Effects

6.1 On the microeconomic effects of gender budgeting adoption

Beyond the effects that gender budgeting adoption can have on fiscal policy strategies and how it led to an increase of fiscal space dedicated to gender issues, I tried to assess how this can affect women daily lives. 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 very 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 measure the effect of this policy reform of women empowerment and/or gender equality, I used intimate partner violence (IPV) as a measure of gender equality evolution and women empowerment. Indeed, I assumed that an improvement of women empowerment will lead to a decrease of the likelihood to accept or agree with IPV. Schuler and Nazneen (2018) findings suggest that women's empowerment has evolved in several ways that may be contributing to reductions in IPV: in its magnitude (for example, many women are earning more income than they previously did), in women's perceived exit options from abusive marriages, in the propensity of community members to intervene when IPV occurs, and in the normative status of empowerment (it is less likely to be seen as transgressive of gender norms). Dalal (2011) shows that economic empowerment is not the sole protective factor. Economic empowerment, together with higher education and modified cultural norms against women, may protect women from IPV. By focusing on the education and integrating ender issues into fiscal policy strategies and reflexion gender budgeting can help to reduce IPV. So, it could be an interesting outcome to assess the microeconomic effect (on women) of gender budgeting adoption.

6.2 Data and empirical strategy

Data on IPV 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). By collecting answers about IPV among others from representative samples of the population, the DHS Program provides representative estimates of IPV tolerance rates among Indian states.

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 about IPV tolerance 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 on the IPV tolerance 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 8: Summary statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
s936f	0.248	0.432	0	1	38483
s936g	0.357	0.479	0	1	38540
s945	0.825	0.38	0	1	34872
v745a	0.954	1.192	0	3	74094
v745b	0.772	1.145	0	3	74094
s929	0.473	0.499	0	1	74094
loggdppc	11.354	0.474	10.278	12.728	376696
urbanization	34.182	13.18	9.83	71.400	352557
local_wip	48.546	1.637	44.47	52.49	376577
logpopulation	3.344	1.731	-0.635	5.476	352557
dose	4.76	5.421	0	15	454067
v729	2.624	1.588	0	8	74021
awfacte	100	0	100	100	454067

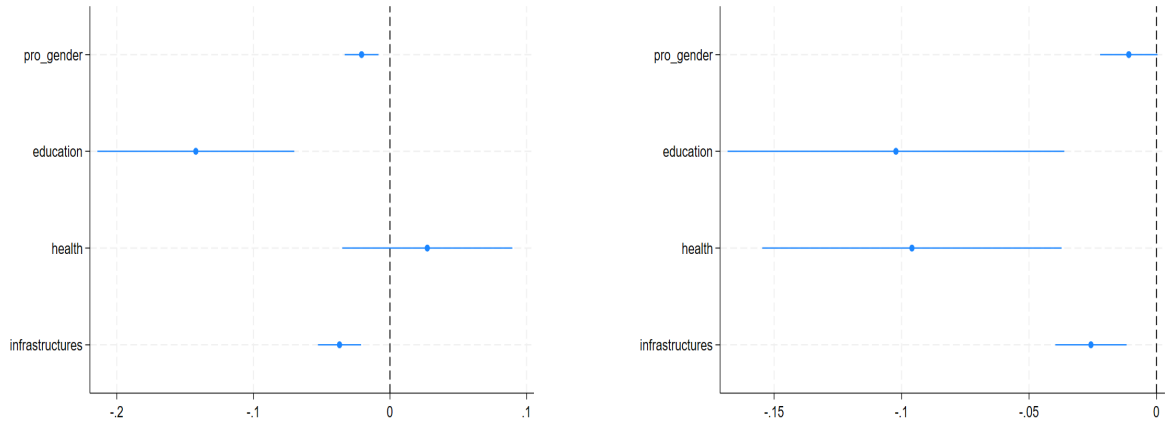
The dependent variable is a binary variable coded as 1 (if the respondent considers as normal to beat a wife for specific reasons regardless of high or low intensity) and 0 (otherwise). The variable of interest is the time (in year) since the first implementation of gender budgeting (to measure the intensity of the treatment). Given the qualitative nature of the dependent variable, the preferred estimation method for estimating equation (1) is the probit model. Compared to the linear probability method and the logit model, the probit model is the most effective and efficient in estimating the qualitative model. For the variable of interest, I used the share of so called "pro gender" spending and its components to ensure that this increase of the fiscal space dedicated to these items can effectively improve the women situation and help to their empowerment. 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 relevance

of the identification strategy is verified through sensitivity analysis.

6.3 Results

The results suggest a negative effect of pro gender and education public spending increase on the likelihood for a woman to consider as acceptable some Intimate Partner Violence (IPV) for unfaithful and disrespect. Women are also less likely to agree with the fact that their partners have sexual relationship with other women. The women who live in states that have implemented gender budgeting seem to be more aware of their rights. The results seem to be more important with the increase of public spending education spending.

The next graphs summarize the results for both outcomes.



(a) Marginal effect on IPV for disrespect

(b) Marginal effect on IPV for disrespect

Figure 13: Marginal effects of pro gender public spending on IPV

The tables G.1 and G.2 available in appendix also show the results from a numerical point of view.

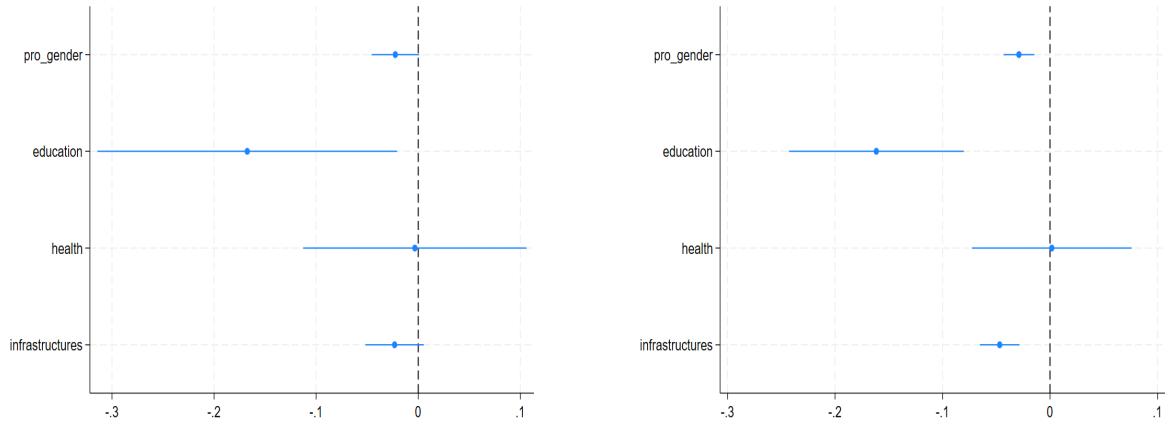
6.4 Sensitivity Analysis

To assess the potential factors that can affect women's empowerment, I conducted a sensitivity analysis to see if living areas or educational attainment could influence my results. Indeed, living areas could significantly impact the effectiveness of gender budgeting. In rural regions, implementing and monitoring gender budgeting can pose challenges due to limited infrastructure, lower administrative capacity, and potential lack of awareness or political will. Additionally, funds designated for gender-focused initiatives may disproportionately benefit urban areas where resources and administrative capabilities are more concentrated. This

potential urban bias can lead to an unequal distribution of benefits, undermining efforts to address gender disparities in rural communities. Consequently, the intended outcomes of gender budgeting—such as improved access to education, healthcare, and economic opportunities for women—may not be fully realized in rural areas, exacerbating existing inequalities (Bhana (2010)). So, it could be interesting to check if this factor among others can affect my results.

To examine these effects, I recompute the same equation, first using only observations from urban areas and then from rural areas. This allows me to determine if the results remain significant in both contexts and to compare the magnitude of these effects between urban and rural areas. By isolating urban and rural observations, I can assess whether the implementation and impact of gender budgeting differ across these living areas, providing insight into any disparities and helping to tailor more effective gender-focused policies for each context.

The next graphs (14a and 14b) summarize the results for IPV because of unfaithful from women and show that the effects remain positive and significant in both areas and are mainly driven by education public spending increase.



(a) Marginal effect on IPV in urban areas

(b) Marginal effect on IPV in rural areas

Figure 14: Marginal effects of pro gender public spending on IPV because of unfaithful

The tables available in appendix also show the results from a numerical point of view.

The results indicate that both overall pro-gender spending and education significantly impact domestic violence rates. However, spending on infrastructure, particularly roads and bridges, is notably more significant and exerts a stronger effect in rural areas. This heightened impact can be attributed to several factors. Firstly, improved roads and bridges enhance accessibility to essential services, such as healthcare and law enforcement, which are crucial in addressing and preventing domestic violence. Secondly, better transportation

networks facilitate economic opportunities, reducing financial stress and associated violence. Lastly, enhanced infrastructure promotes social connectivity and support networks, which are vital in rural areas where isolation can exacerbate domestic violence situations. Therefore, infrastructure improvements in rural regions could play a critical role in mitigating domestic violence.

Education spending is also significant and exerts an important effect in rural areas. This heightened impact can be attributed to several factors. Firstly, improved education enhances awareness and understanding of domestic violence, equipping individuals with the knowledge to address and prevent it. Secondly, better educational opportunities lead to economic empowerment, reducing financial stress and associated violence. In urban areas, education fosters diverse social connectivity and support networks, while in rural areas, it mitigates isolation that can exacerbate domestic violence situations. Therefore, education improvements play a critical role in mitigating domestic violence across both urban and rural regions.

In the next section, I have tried to check the potential transmission channels that can explain my results

7 Transmission channels

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

Table 9: Transmission channel

	CSS(% of revenues)	CSS (% of expenditures)	Credibility index
Before adoption	4.08	4.18	4.13
After adoption	5.82	6.05	6.51
Non Gender Budgeting	4.54	4.60	5.33

I try to estimate the potential transmission channels by using the same process as

Neuenkirch and Neumeier (2016) I compute the means of the two variables for (a) the treatment group during times when gender budgeting is in place, (b) the treatment group focusing only on years before gender budgeting implementation, and (c) our synthetic control group obtained via entropy balancing. The results are outlined in table 9 just above. The descriptive statistics indicate some differences between the control group obtained via entropy balancing and states which apply gender budgeting. When comparing the control group to the treatment group before gender budgeting was applied, however, I find that the latter is characterized by a notably greater share of CSS for both measures.

Indeed, before the treatment, the treated units received less CSS in the percentage of revenues (4.08% versus 4.54%) and expenditures (4.18% versus 4.60%) than the non-adopter ones, but the situation became different after the adoption for revenues (5.82 vs 4.54) and expenditures (6.05 vs 4.60).

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

I also compute a kind of credibility index. To do it, I compute the difference between the share of "pro-gender"expenditures expected in the budget announcement and the share of "pro-gender"expenditures in the States financial account. I have assumed that this bias index will allow us to apprehend the performance of subnational administration. The differences between forecast and realization are possible and usual, but a systematic and important difference may mean a lower level of performance in its administration. I compute it as an absolute value. *The absolute value refers to the fact that I multiply the negative value by -1 to get only positive values* to consider the distance (bias) between the forecast and the realization. I 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. The credibility index is computed as:

$$Bias_{it} = \left(\frac{Pro\_gender_{it}}{Total\_expenditure_{it}} - \frac{Pro\_gender\_expected_{it}}{Total\_expenditure\_expected_{it}}\right) * 100 \tag{7}$$

Where $Pro\_gender\_expected$ represents the expenditures previously considered as "pro-gender"in the budget announcement. They are expressed as a percentage of Total expenditures also expected in the budget announcement ($Total\_Expenditures\_expected$).

However, to make it easy to interpret and more intuitive, Itransform it under the form:

$$credibility_{it} = \frac{1}{Bias_{it}} \tag{8}$$

The results available in table 9 also suggest that Gender Budgeting adopters seem to become more credible than non-adopters.

8 Conclusion

Through this work, I have tried to evaluate the effects that gender budgeting adoption can have on Indian state's fiscal policies, and more precisely on so called "pro gender" spending for a panel of Indian States over the period 1991-2020. To answer this question, I use different DiD estimators and another novel method — entropy balancing — combining a matching approach with linear regression, thus mitigating endogeneity issues that may lead to skepticism about our conclusions. In addition to DiD and entropy balancing, I use a Dose-Response Model and perform some alteration and anticipation tests to ensure that our results are robust.

Our results suggest that Indian states which adopt gender budgeting tend to devolve a greater share of their expenditures to "pro-gender" ones. The effects are more significant for education and infrastructure spending (Samarakoon and Parinduri (2015)), while they are ambiguous for water and sanitation spending. The adopters' states seem to have been credible in their wish to prioritize "pro-gender" spending. The potential transmission channel identified in this paper is through an increase in the transfers received from central governments and more precisely Centrally Sponsored Schemes (CSS) as denoted in the table 9.

Gender budgeting adoption by imposing a continuous assessment affects the overall fiscal framework. Gender budgeting is a useful policy tool that has some effects on fiscal policy because fiscal policy and public expenditures are the means used to reach the objectives to reduce gender inequalities. Indeed, gender budgeting can play an important role in tackling gender inequalities and increasing social outcomes like Chakraborty (2016) and Stotsky and Zaman (2016) shown. However, it can also affect fiscal policy and expenditure decisions and priorities. In addition, Gender budgeting adoption seems to make adopters more credible in the prioritization of "pro-gender" policies. In a context like the Indian one, with an existing democracy and strong civil society at both (national and local level) gender budgeting adoption through transparency and gender budgeting statement can make the government more credible in their wish to tackle gender inequalities.

In the context of Indian states, the implementation of gender budgeting, particularly

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

In conclusion, the positive effects of implementing Gender Budgeting also manifest at the individual level by improving the situation of women and reducing the acceptance of domestic violence against women. The adoption of Gender Budgeting, while adhering to macroeconomic commitments, translates into a relative improvement in women's conditions through targeted fiscal policies and public expenditures. Therefore, the idea of a mere change in name without structural reform should be rejected, as only a genuine adoption of Gender Budgeting can yield these significant benefits at both macroeconomic and microeconomic levels. Additionally, the implementation of this reform highlights the importance of transparency in public policy-making to ensure their complete and effective realization, ultimately benefiting local populations. So, the adopters need to consider their underlying policy context (Polzer et al. (2023)). They must also consider the importance of transparency through policy statements to make the reform and the government that applies it fully effective and credible.

References

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

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.

Alonso-Albarran, V., Preston, G., Soler, A., Tchelishvili, N., and Weerathunga, S. (2021). *Gender budgeting in G20 countries*. International Monetary Fund.

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.

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

Baccini, L., Li, Q., Mirkina, I., and Johnson, K. (2018). Regional competition, business politicians, and subnational fiscal policy. *Business and Politics*, 20(3):410–437.

Besley, T. and Smart, M. (2007). Fiscal restraints and voter welfare. *Journal of public Economics*, 91(3-4):755–773.

Bhana, D. (2010). "here in the rural areas they don't say that men and women are equal!" contesting gender inequalities in the early years. *Agenda*, 24(84):9–18.

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).

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.

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

Budlender, D. and Hewitt, G. (2003). *Engendering Budgets: A practitioners' guide to understanding and implementing gender-responsive budgets*. Commonwealth Secretariat.

Budlender, D., Sharp, R., and Allen, K. (1998). *How to do a gender-sensitive budget analysis: Contemporary research and practice*. Commonwealth Secretariat.

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

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

Calvo, C. M. et al. (1994). *Case study on intermediate means of transport: bicycles and rural women in Uganda*. Citeseer.

Cannonier, C. and Mocan, N. (2018). The impact of education on women's preferences for gender equality: Evidence from sierra leone. *Journal of Demographic Economics*, 84(1):3–40.

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

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

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

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

Dalal, K. (2011). Does economic empowerment protect women from intimate partner violence? *Journal of injury and violence research*, 3(1):35.

Das, M. S., Jain-Chandra, M. S., Kochhar, M. K., and Kumar, N. (2015a). *Women workers in India: why so few among so many?* International Monetary Fund.

Das, S., S., J.-C., K., K., and N., K. (2015b). Women workers in india: Why so few among so many.

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*, 46(7):1365–1379.

Demery, L. and Gaddis, I. (2009). Social spending, poverty and gender equality in kenya-a benefit incidence analysis. *Eschborn: German Technical Cooperation*, pages 1–38.

Doepke, M., Tertilt, M., and Voena, A. (2012). The economics and politics of women's rights. *Annu. Rev. Econ.*, 4(1):339–372.

Elomäki, A. and Ylöstalo, H. (2021). Gender budgeting in the crossroad of gender policy and public financial management: The finnish case. *Public Money & Management*, 41(7):516–526.

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

Gadenne, L. (2017). Tax me, but spend wisely? sources of public finance and government accountability. *American Economic Journal: Applied Economics*, pages 274–314.

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.

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

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political analysis*, 20(1):25–46.

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

Jennings, I. (1953). Some characteristics of the indian constitution: being lectures given in the university of madras during march 1952 under the sir alladi krishnaswami aiyer shashtiabdapoorthi endowment. *(No Title)*.

Joshi, S. B. (2013). Assessing gender equality investments: A multi-state perspective. *UN Women and National Foundation for India, New Delhi*.

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

Khera, P. (2016). *Macroeconomic impacts of gender inequality and informality in India*. International Monetary Fund.

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

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.

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

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.

Mejía Acosta, A. (2013). The impact and effectiveness of accountability and transparency initiatives: The governance of natural resources. *Development Policy Review*, 31:s89–s105.

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.

Morrissey, S. (2018). *Gender Budgeting: A Useful Approach for Aotearoa New Zealand*. Number 18/02. New Zealand Treasury Working Paper.

Moschovis, G. (2010). Public spending allocation, fiscal performance and corruption. *Economic Papers: A journal of applied economics and policy*, 29(1):64–79.

Neuenkirch, M. and Neumeier, F. (2015). The impact of un and us economic sanctions on gdp growth. *European Journal of Political Economy*, 40:110–125.

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

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.

Parikh, P., Fu, K., Parikh, H., McRobie, A., and George, G. (2015). Infrastructure provision, gender, and poverty in indian slums. *World Development*, 66:468–486.

Parvati, P., Jhamb, B., Shrivastava, S., and Mobeen ur Rehman, K. (2012). Recognising gender biases, rethinking budgets: Review of gender responsive budgeting in the union government and select states.

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

Puig-Barrachina, V., Ruiz, M. E., García-Calvente, M. d. M., Malmusi, D., Sánchez, E., Camprubí, L., Muntaner, C., Cortès-Franch, I., Artazcoz, L., and Borrell, C. (2017). How to resist austerity: The case of the gender budgeting strategy in andalusia. *Gender, Work & Organization*, 24(1):34–55.

Quinn, S. (2016). Europe: A survey of gender budgeting efforts.

Rambachan, A. and Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, page rdad018.

Rao, M. G. (2009). Fiscal federalism in india—trends and reform. *Decentralization policies in Asian development*, 107.

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

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3):305–22.

Roth, J., Sant'Anna, P. H., Bilinski, A., and Poe, J. (2023). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*.

Rubin, M. M. and Bartle, J. R. (2023). Gender-responsive budgeting: a budget reform to address gender inequity. *Public Administration*, 101(2):391–405.

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

Salim, A. (2016). Gender responsive budgeting: State and civil society initiatives at national and sub-national levels in indonesia. *Gender Responsive and Participatory Budgeting: Imperatives for Equitable Public Expenditure*, pages 179–193.

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

Sarraf, M. F. (2003). *Gender-responsive government budgeting*. International Monetary Fund.

Schuler, S. R. and Nazneen, S. (2018). Does intimate partner violence decline as women's empowerment becomes normative? perspectives of bangladeshi women. *World development*, 101:284–292.

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

Shair-Rosenfield, S., Schakel, A. H., Niedzwiecki, S., Marks, G., Hooghe, L., and Chapman-Osterkatz, S. (2021). Language difference and regional authority. *Regional & Federal Studies*, 31(1):73–97.

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

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

Stotsky, J. G. (2006). Gender budgeting.

Stotsky, M. J. G. and Zaman, M. A. (2016). *The influence of gender budgeting in Indian states on gender inequality and fiscal spending*. International Monetary Fund.

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

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

Swenden, W. (2015). India and the management of ethnic diversity: The unfinished business of accommodation. *From the Margins to the Mainstream: Institutionalising Minorities in South Asia*.

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.

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

Weiss, J. (1999). *Infrastructure and economic development*. African Development Bank.

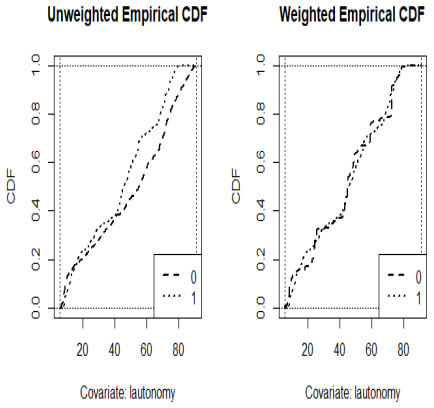
Wooldridge, J. M. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Available at SSRN 3906345*.

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

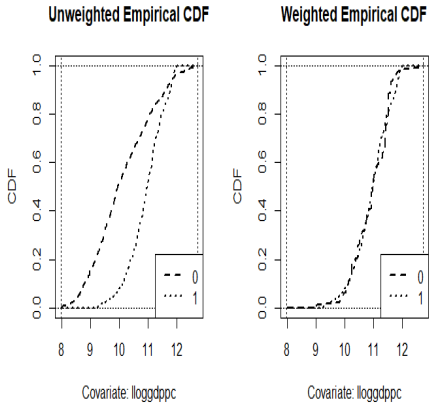
Appendices

Appendix A

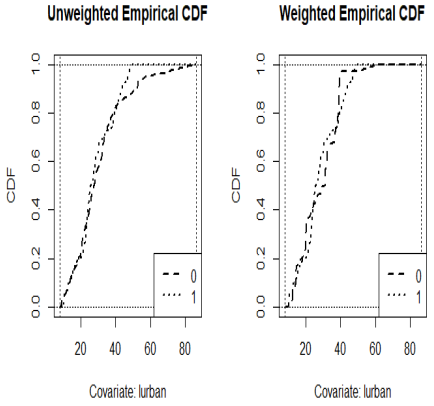
A Entropy balancing



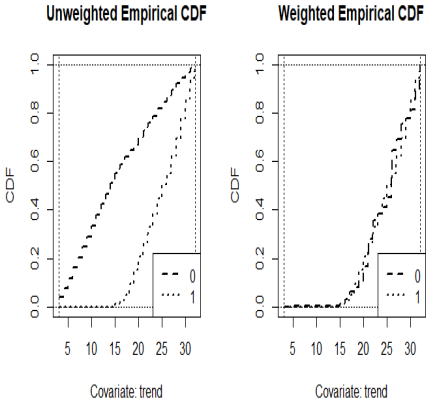
(a) Autonomy reweighted



(b) GDPpc reweighted



(c) Urbanization reweighted



(d) Trend reweighted

Figure A.1: Entropy balancing results

B Staggered DiD

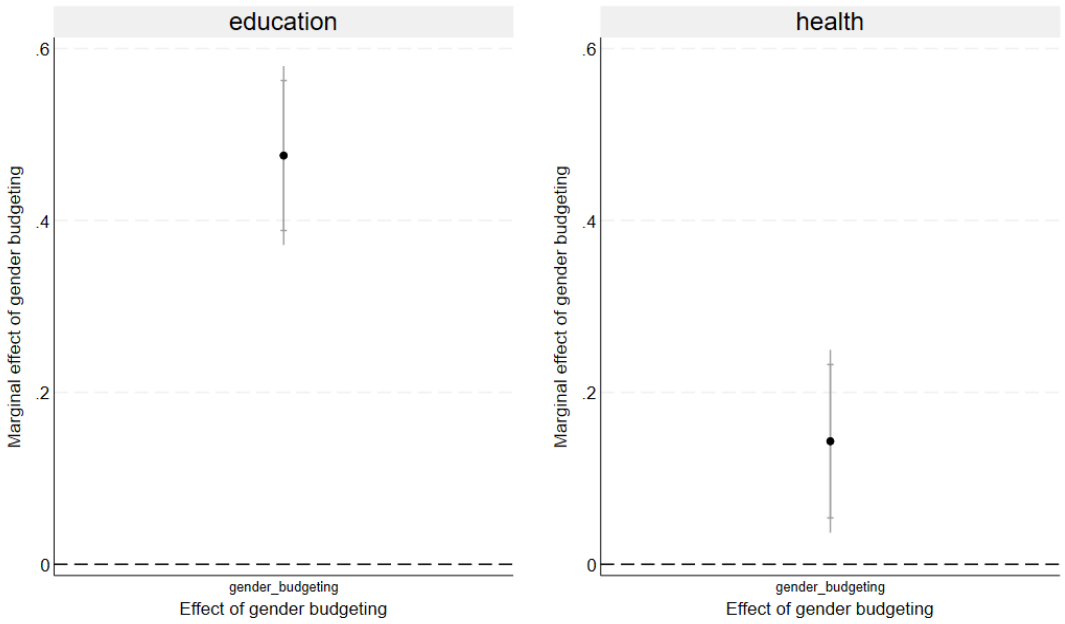
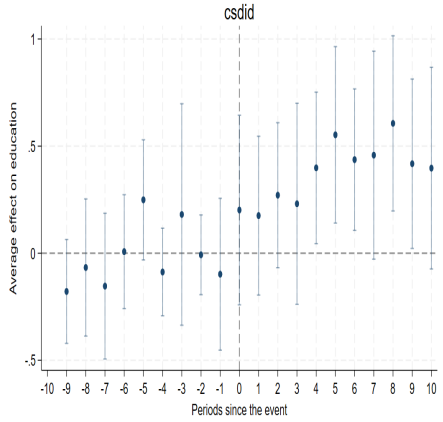
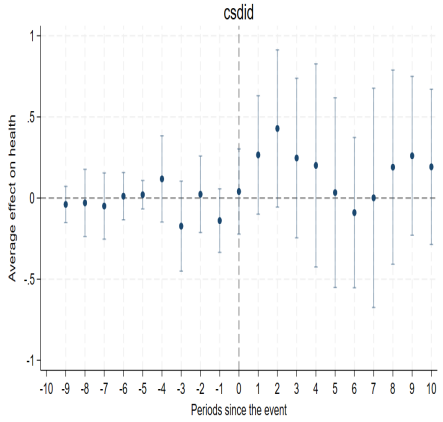


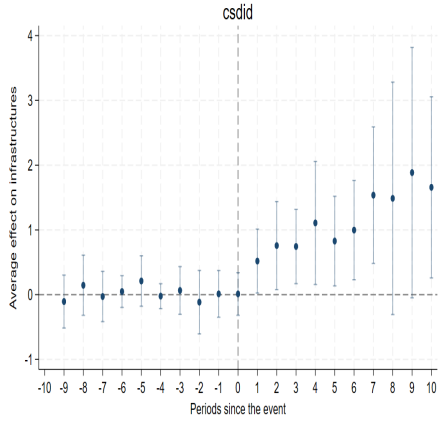
Figure B.1: Visual results of Callaway and Sant'Anna



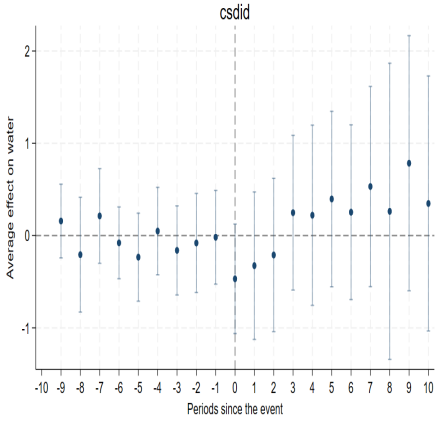
(a) Event study for education



(b) Event study health



(c) Event study infrastructures



(d) Event study water and sanitation

Figure B.2: Event study for "pro gender" components (Callaway and Sant'Anna (2021))

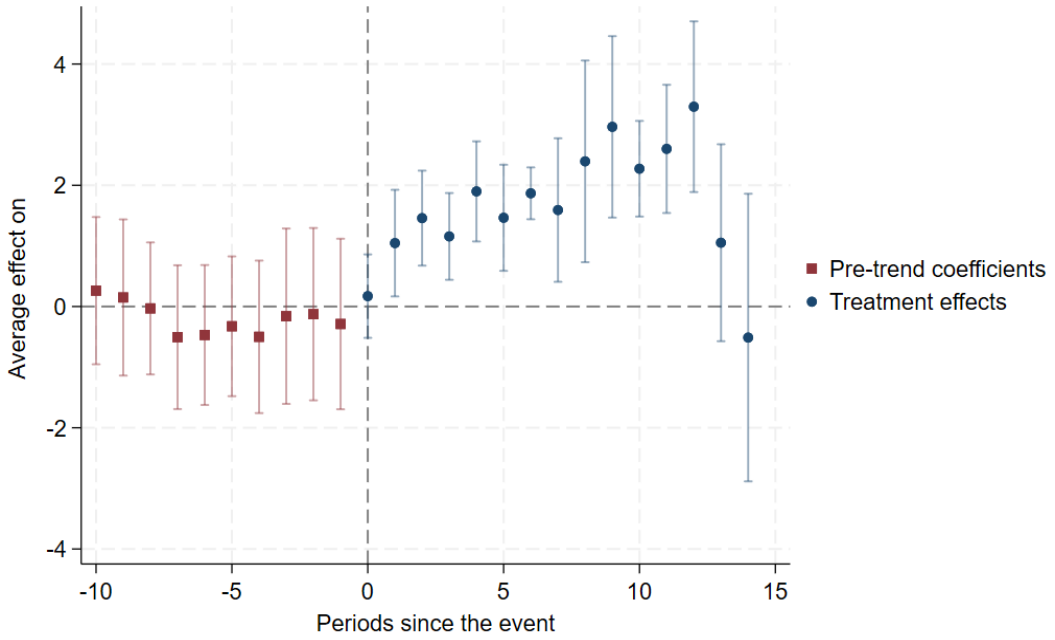


Figure B.3: Event study results for "pro-gender"spending

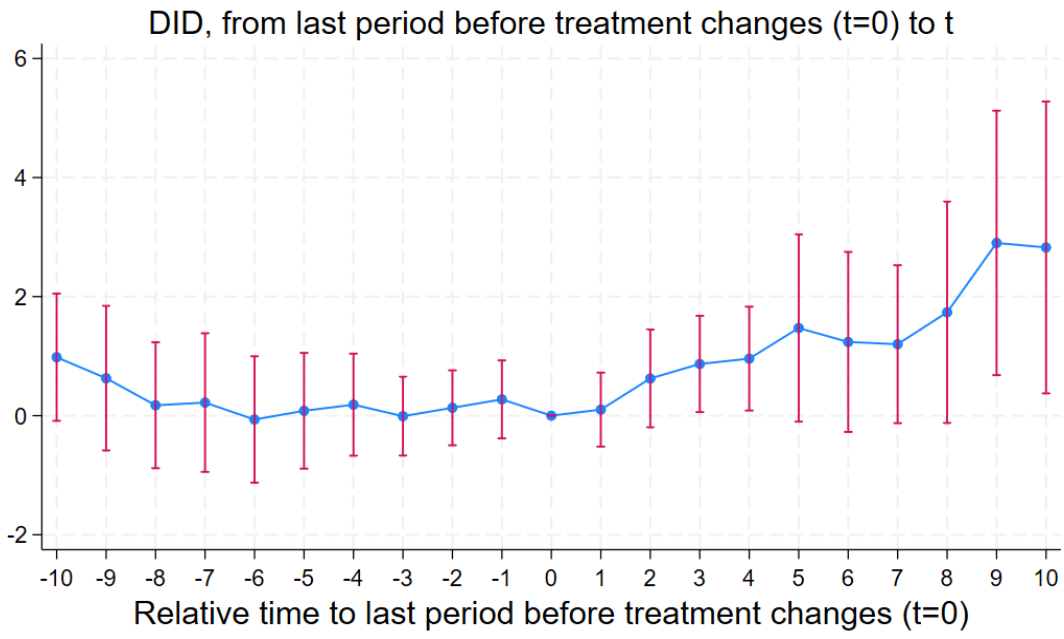


Figure B.4: Event study results for "pro-gender"spending

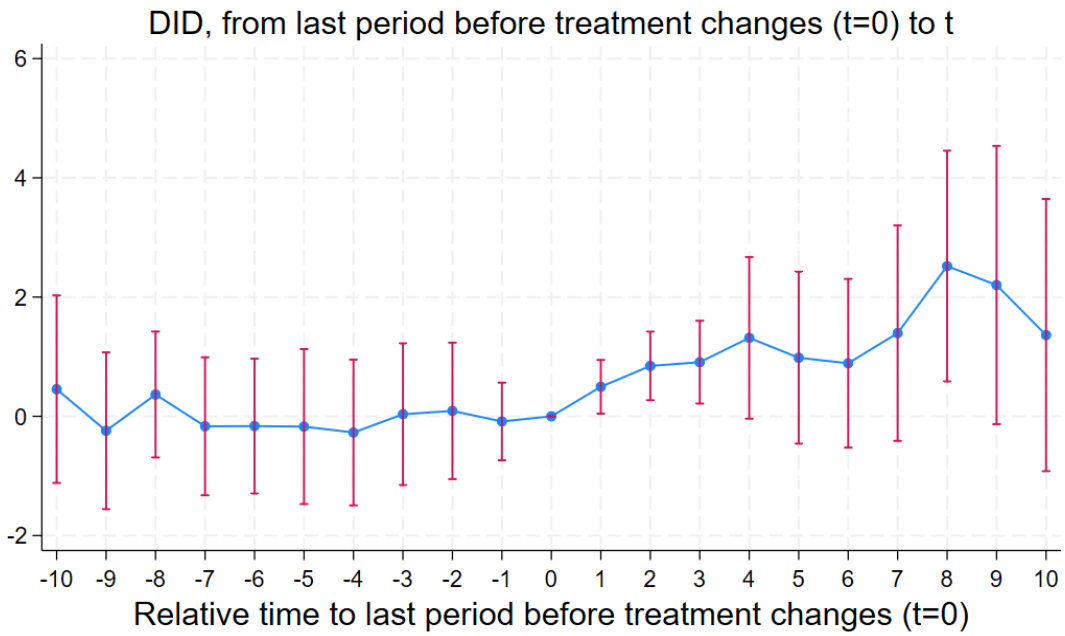


Figure B.5: Event study results for "pro-gender"spending with continuous treatment

Appendix B

C Dose response model

Table C.1: Dose Response Model results

	(1) pro_gender	(2) education	(3) health	(4) infrastructures	(5) water
gender_budgeting	2.077***	0.612***	0.510**	0.731***	0.237
	(0.556)	(0.200)	(0.190)	(0.244)	(0.319)
N	668	668	668	668	668
R^2	0.245	0.148	0.279	0.225	0.139

Standard errors in parentheses

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

D Placebo Test

Table D.1: Placebo test Results for components

	(1) pro_gender	(2) education	(3) health	(4) infrastructures	(5) water
placebo	0.106	0.0735	0.0224	0.0805	-0.0665
	(0.169)	(0.0553)	(0.0427)	(0.106)	(0.100)
N	668	668	668	668	668
R^2	0.435	0.242	0.242	0.242	0.207

Standard errors in parentheses

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

E Anticipation effects

Table E.1: Anticipation effects results

	pro gender	Education	Health	Infrastructures	Water
ATT	0.640	0.293*	-0.117	0.444	0.0239
	(0.94)	(1.91)	(-0.49)	(1.55)	(0.06)
Observations	668	668	668	668	668

t statistics in parentheses

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

F Narrowing the control window

Table F.1: Results for narrowing -3 ; +3

	Pro gender	Education	Health	Infrastructures	Water
gender_budgeting	1.456***	0.298***	0.184**	1.443***	-0.410
	(3.60)	(3.83)	(2.27)	(5.93)	(-1.32)
Observations	166	166	166	166	166

t statistics in parentheses

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

Table F.2: Results for narrowing -5 ; +5

	Pro gender	Education	Health	Infrastructures	Water
gender_budgeting	1.551***	0.325***	0.153*	1.465***	-0.327
	(4.02)	(4.27)	(1.90)	(6.09)	(-1.13)
Observations	180	180	180	180	180

t statistics in parentheses

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

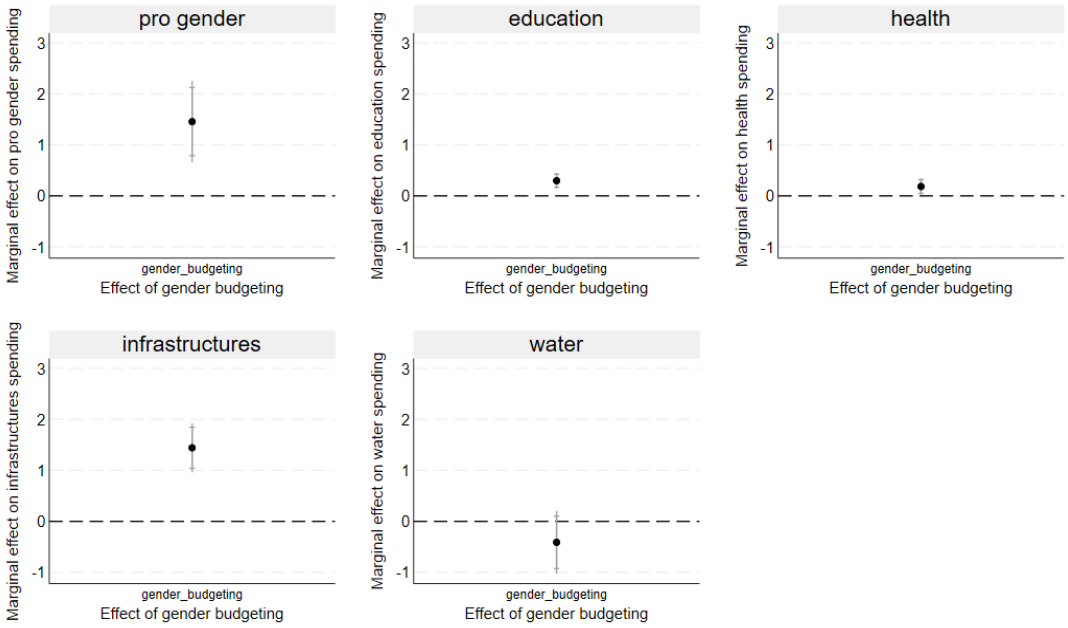


Figure F.1: Results for a 3 years around period

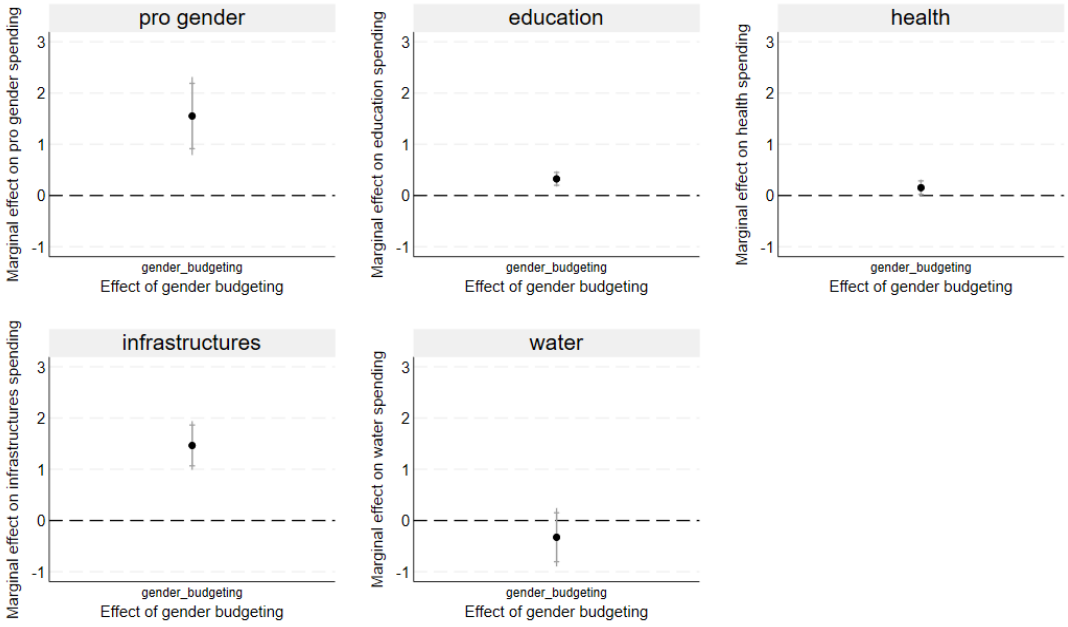


Figure F.2: Results for 5 years around period

Appendix C

G Microeconomic effects

Table G.1: Marginal effect on IPV due to unfaithful

	pro gender	education	health	infrastructure
justifies domestic violence: wife unfaithful				
pro_gender	-0.0208***			
	(-3.29)			
education		-0.142***		
		(-3.86)		
health			0.0273	
			(0.86)	
infrastructures				-0.0369***
				(-4.58)
N	49294	49294	49294	49294

t statistics in parentheses

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

Table G.2: Marginal effect on IPV due to disrespect

	pro gender	education	health	infrastructure
justifies domestic violence: wife disrespect				
pro_gender	-0.0109*			
	(-1.89)			
education		-0.102***		
		(-3.03)		
health			-0.0959***	
			(-3.20)	
infrastructures				-0.0257***
				(-3.60)
N	49351	49351	49351	49351

t statistics in parentheses

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

Table G.3: Marginal effect on IPV due to unfaithful in urban areas

	pro gender	education	health	infrastructure
pro_gender	-0.0225*			
	(-1.89)			
education		-0.167**		
		(-2.24)		
health			-0.00329	
			(-0.06)	
infrastructures				-0.0231
				(-1.58)
N	11625	11625	11625	11625

t statistics in parentheses

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

Table G.4: Marginal effect on IPV due to unfaithful in rural areas

	pro gender	education	health	infrastructure
pro_gender	-0.0289***			
	(-3.95)			
education		-0.162***		
		(-3.89)		
health			0.00167	
			(0.04)	
infrastructures				-0.0468***
				(-4.98)
N	38037	38037	38037	38037
			(-1.58)	

t statistics in parentheses

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