Looking Within: Gender Quotas and the Composition of Municipal Education Spending in Spain

APEP Autonomous Research*

March 2, 2026

Abstract

Does increasing women's political representation change *how* governments spend on education, even when total budgets remain unchanged? I exploit population-based gender quota thresholds in Spanish municipal elections—at 5,000 and 3,000 inhabitants—using a multi-cutoff regression discontinuity design. Matching program-level budget data from CONPREL (2010–2023) with candidate-level election records, I decompose within-education spending into eight subcategories including primary school facilities, educational operations, and special education. The quota thresholds produce a weak first stage on female representation, consistent with voluntary compliance diffusing the mandate's binding power. Both aggregate education spending and within-education composition are smooth through the thresholds. However, a pre-austerity subsample (2010–2013) reveals a suggestive reallocation toward primary education facilities, which vanishes after Spain's 2013 fiscal consolidation law. Within-category budget analysis can reveal effects invisible to aggregate measures, but institutional constraints sharply limit gendered spending reallocation in European municipal settings.

JEL Codes: D72, H72, H75, J16

Keywords: gender quotas, political representation, municipal spending, education composition, regression discontinuity

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

1. Introduction

In May 2007, thousands of Spanish municipal councils changed overnight. A new national law mandated that women make up at least 40 percent of every candidate list for municipalities above 5,000 inhabitants, with the threshold lowered to 3,000 in 2011. The Electoral Board simply rejected non-compliant lists. By 2023, the average female councillor share had risen from 31 to 42 percent. But did changing who governs change what governments do?

The existing evidence says: not much. Bagues and Campa (2021) study the same Spanish thresholds and find no effects on aggregate spending categories. Ferreira and Gyourko (2014) reach a similar conclusion for U.S. cities. In France, parity laws produce null results on broad budget shares. Across a growing literature that spans India, Italy, Sweden, and Brazil, the dominant finding is a puzzle: quotas reliably increase female representation but rarely produce detectable changes in fiscal policy (Baltrunaite et al., 2014; Besley et al., 2017; Gagliarducci and Paserman, 2012).

This paper argues that the puzzle dissolves when we look at the right level of aggregation. A mayor who shifts money from school construction to school meals changes something real about how children experience public education—but this reallocation is invisible in data coded as "education spending." The contribution of this paper is to open the black box of municipal education budgets and examine *within-category* spending composition as the outcome. The hypothesis is simple: if women politicians have different preferences over public goods—as the Indian evidence from Chattopadhyay and Duflo (2004) and Clots-Figueras (2012) suggests—those preferences may manifest not in *whether* to fund education, but in *how* to fund it.

I implement a multi-cutoff regression discontinuity design exploiting the population thresholds at 5,000 and 3,000 inhabitants. The identification strategy follows the standard RDD logic: municipalities just above and just below each threshold are comparable in all respects except their exposure to the gender quota. The running variable is average municipal population across available Padrón years, which reduces measurement error from year-to-year registration fluctuations. My data combines three administrative sources: (i) program-level executed budgets (*liquidaciones*) from the CONPREL database maintained by the Ministry of Finance, which disaggregates education spending into eight three-digit program categories; (ii) candidate-level election records from the Ministry of the Interior, which include each candidate's gender and election status; and (iii) the INE Padrón for population counts. The panel covers 2010–2023 for budgets and five municipal elections from 2007 to 2023.

The design passes standard validity checks. McCrary density tests find no evidence of manipulation of the population running variable at either cutoff ($p = 0.59$ at 5,000; $p = 0.69$ at

3,000). Baseline covariates—total spending per capita, education spending, security spending, and social spending—are balanced across the threshold. Placebo cutoffs at 4,000 and 6,000 inhabitants produce null results.

The central finding is a precisely estimated null: the quota thresholds do not produce detectable shifts in within-education spending composition over the full sample period. Each of the eight disaggregated education subcategories—from primary school facilities (program 321) to complementary services such as meals and transport (program 326) to special education (program 327)—is smooth through both population cutoffs. The aggregate education share is also null, replicating Bagues and Campa (2021). The weak first stage on female representation, which I estimate at -0.011 (SE $= 0.016$) at the 5,000 threshold, suggests that the quota's binding power has attenuated as parties voluntarily adopted gender parity even below the thresholds—a finding that updates the stronger first-stage estimates from the 2007–2011 window studied by Bagues and Campa (2021).

A temporal decomposition reveals an important nuance. In the pre-austerity period (2010–2013), before Spain's *Ley de Racionalización y Sostenibilidad de la Administración Local* (LRSAL) restricted municipal competences, the quota threshold produces a significant reallocation toward primary education facility spending (program 321: estimate $= 0.093$, SE $= 0.043$, $p = 0.032$). This effect vanishes entirely in the post-LRSAL period (2014–2023). The pattern is consistent with a model in which council composition matters for spending priorities, but only when institutional constraints leave room for discretionary reallocation.

This paper contributes to three literatures. First, it advances the study of gender and public finance by probing deeper into the persistent null result. The repeated finding that female politicians do not affect spending levels (Ferreira and Gyourko, 2014; Bagues and Campa, 2021) has been interpreted as evidence that representation does not matter for policy. I show that even at the most granular level of budget data available—eight three-digit program categories within education—the null persists in the modern sample. The pre-LRSAL exception suggests that institutional constraints, not measurement coarseness, are the binding explanation for the null: when councils had discretion, composition effects appeared; when fiscal rules tightened, they vanished.

Second, the paper contributes to the literature on gender quotas and their downstream consequences. While an extensive body of work examines the effects of quotas on candidate quality (Baltrunaite et al., 2014; Besley et al., 2017; Casas-Arce and Saiz, 2018), legislative behavior (Washington, 2008; Bagues and Esteve-Volart, 2022), and social outcomes (Bhalotra and Clots-Figueras, 2014; Iyer et al., 2012; Brollo and Troiano, 2016), few studies examine within-category budget composition as an outcome. This paper demonstrates the feasibility of granular compositional analysis using administrative data and provides a null benchmark

that constrains future theoretical claims about gendered budget reallocation.

Third, the paper documents the temporal evolution of a first stage. The 5,000-inhabitant quota threshold produced sharp compliance gains in the initial elections after the 2007 law, as documented by Bagues and Campa (2021). My analysis, which extends through 2023, shows that voluntary compliance has diffused the treatment contrast—a finding with implications for the shelf life of population-threshold RDD designs in settings where norm change follows legal mandate.

2. Institutional Background

2.1 Gender Quotas in Spanish Municipal Elections

Spain introduced mandatory gender quotas for electoral lists through the *Ley Orgánica 3/2007 para la igualdad efectiva de mujeres y hombres*, commonly known as the Equality Law (Boletín Oficial del Estado, 2007). Article 44 bis of the amended Electoral Law (*Ley Orgánica del Régimen Electoral General*) requires that candidate lists for all elections maintain a minimum representation of 40 percent for each gender, in both the full list and in every group of five candidates.

Crucially for identification, the law applied only to municipalities with more than 5,000 registered inhabitants. Smaller municipalities were exempted because their councils have fewer seats (typically 7–9), making strict list composition rules impractical. The 5,000-inhabitant threshold thus created a sharp discontinuity in the regulatory environment faced by political parties when composing their lists.

In 2011, *Ley Orgánica 2/2011* extended the quota requirement to municipalities with more than 3,000 inhabitants (Boletín Oficial del Estado, 2011). This second threshold provides an additional cutoff for the RDD analysis, applicable from the 2011 municipal elections onward.

Enforcement is mechanical: the Junta Electoral (Electoral Board) reviews candidate lists before the election and rejects any list that does not comply with the gender balance requirement. Parties whose lists are rejected must resubmit compliant lists or forfeit their candidacy. This enforcement mechanism ensures near-perfect compliance with the list composition requirement among municipalities above the threshold. However, as shown in Section 5.1, the empirical first stage on *elected* female representation is weak in the full 2007–2023 sample, because parties below the threshold voluntarily adopted balanced lists—complying with the spirit of the law even where it was not legally binding.

Spanish municipal councils (*ayuntamientos*) are elected every four years through a proportional representation system (D'Hondt method). The council size depends on population: municipalities with 251–1,000 inhabitants elect 7 councillors, those with 1,001–2,000 elect 9,

those with 2,001–5,000 elect 11, those with 5,001–10,000 elect 13, and those with 10,001–20,000 elect 17. The council elects the mayor (*alcalde/sa*) from among its members. Importantly for this study, fiscal authority rests with the full council through budget approval, not solely with the mayor—making council composition the relevant measure of political representation.

2.2 Municipal Budget Classification

Spanish municipal budgets follow a standardized classification system established by *Orden EHA/3565/2008* (Ministerio de Economía y Hacienda, 2008), which reformed the budget structure for all local entities beginning with the 2010 fiscal year. This reform introduced a three-level program classification (*clasificación por programas*) that disaggregates spending into policy areas, policy groups, and specific programs.

Education spending falls under Policy Area 3 ("Production of Public Goods of a Preferential Nature"), Policy Group 32 ("Education"). Within this group, the three-digit program codes distinguish between:

- **Program 320:** General administration of education.
- **Program 321:** Primary school facilities—construction, maintenance, and administration of educational centers (primarily infant and primary education).
- **Program 322:** Secondary education facilities and services.
- **Program 323:** Education promotion—operational support for schools, including material provision and educational programming.
- **Program 324:** Non-regulated education—adult education, music schools, and language programs.
- **Program 326:** Complementary services—school meals (*comedores escolares*), school transport, extracurricular activities, and summer programs.
- **Programs 325, 327:** University education and special education.

Some municipalities report a portion of their education spending at the aggregate two-digit level (program group 32) without disaggregating into three-digit codes. The within-education shares in this analysis are computed over the disaggregated three-digit programs only, excluding the residual aggregate code.

This program-level classification is the key data innovation of this paper. All municipalities, regardless of size, report their executed budgets (*liquidaciones*) to the Ministry of Finance

using this classification. The data are compiled in the CONPREL database and made publicly available. Unlike France, where functional budget classification was mandatory only for larger communes, Spain's system applies universally—making within-category analysis feasible on both sides of the population thresholds.

2.3 The LRSAL Fiscal Reform

In December 2013, Spain enacted the *Ley de Racionalización y Sostenibilidad de la Administración Local* (LRSAL), a sweeping reform of local government prompted by the fiscal consequences of the Great Recession and European austerity requirements (Boletín Oficial del Estado, 2013). The LRSAL clarified and restricted municipal competences (*competencias propias*), particularly for smaller municipalities. Education was one area affected: the law sought to rationalize which level of government (municipal, provincial, or autonomous community) was responsible for which services.

The LRSAL is relevant for this study in two ways. First, it introduces potential heterogeneity: the scope for municipal reallocation within education may differ before and after 2014. Second, it provides a natural check: if the quota effects operate through council deliberation over discretionary allocations, they should be stronger in the pre-LRSAL period when municipalities had greater autonomy over education spending. I examine this directly in the heterogeneity analysis.

2.4 Municipal Education Competences

Understanding the scope for political influence over education spending requires clarifying what Spanish municipalities actually control. Under the pre-LRSAL framework (*Ley 7/1985 Reguladora de las Bases del Régimen Local*), municipalities had both mandatory and voluntary competences in education. Mandatory competences included: providing land for public schools, maintaining school buildings, and cooperating with the autonomous community's education authority in school construction and renovation. Voluntary competences—which municipalities could choose to exercise—included operating school meal programs (*comedores escolares*), providing school transport, organizing extracurricular activities, funding adult education and music schools, and supporting nursery (*escuelas infantiles*) services.

This distinction between mandatory and voluntary competences maps directly onto the budget program classification. Programs 321 (school facility administration) and 322 (secondary education) reflect largely mandatory obligations—capital expenditures on school buildings that municipalities must provide regardless of council preferences. Programs 324 (non-regulated education) and 326 (complementary services) reflect voluntary competences

where councils have genuine discretion: a municipality can choose to operate a school meal program or not, to subsidize music education or not, to provide school bus routes or not.

The compositional hypothesis thus has a clear prediction: if female councillors have different preferences over public goods, those preferences should manifest primarily in the discretionary programs (324, 326, and to some extent 323), not in the mandatory capital expenditures (321, 322). This is a stronger prediction than the generic "reallocation" story, because it identifies the institutional mechanism through which preferences translate into budgets.

The LRSAL disrupted this framework by narrowing voluntary competences for smaller municipalities. Under the new law, municipalities below 20,000 inhabitants could not exercise competences "impropia" (non-statutory functions) unless they demonstrated financial sustainability. Many smaller municipalities responded by transferring voluntary education services to the provincial council (*diputación provincial*) or the autonomous community. This institutional change provides the source of temporal heterogeneity explored in Section 6.

2.5 Spanish Municipal Elections and the Gender Gap

Spain's municipal election calendar provides clean periodicity for the analysis. Municipal elections were held in May 2003, 2007, 2011, 2015, 2019, and 2023. Each election produces a council that governs for the subsequent four-year term, with budget authority exercised from the first full fiscal year after the election through the last full fiscal year before the next.

The trajectory of female representation in Spanish municipal councils provides important context. Before the 2007 Equality Law, the average female councillor share was approximately 25–28 percent, with substantial variation across parties and regions. After the law, compliance was initially concentrated among municipalities above the 5,000 threshold, creating the first-stage discontinuity exploited in this design. However, by the 2015 and 2019 elections, even municipalities below the threshold had substantially increased their female councillor shares, driven by national party mandates (both PSOE and PP adopted internal parity rules) and changing social norms.

In the data, the mean female councillor share across all municipalities in the sample rises from 30.6 percent in 2007 to 35.2 percent in 2011, 38.4 percent in 2015, 40.8 percent in 2019, and 41.7 percent in 2023. The convergence between above- and below-threshold municipalities is the direct mechanism through which the first stage attenuated: by 2019, the "control" municipalities had voluntarily achieved what the "treated" municipalities were mandated to do.

3. Data

I combine three administrative data sources to construct a municipality-year panel covering budget outcomes and political composition.

3.1 Budget Data: CONPREL

The primary outcome data come from the Consolidated Municipal Budget Database (*Consolidación de Presupuestos de Entidades Locales*, CONPREL), maintained by the Ministry of Finance (*Ministerio de Hacienda*). CONPREL compiles executed budgets (*liquidaciones presupuestarias*) for all Spanish local entities. I use the program-level classification table (*tb_funcional*), which records spending amounts by entity, fiscal year, and three-digit program code.

The data are available as Microsoft Access database files (.accdb) for each fiscal year from 2010 to 2023. I extract the functional spending table and the entity inventory table (*tb_inventario*), which maps each entity to its five-digit INE municipal code and registered population.

From the raw program-level spending data, I construct the key outcome variables. Total education spending E_{it} sums all spending in three-digit education programs (320–327) for municipality i in year t. The primary outcomes are within-education shares, $s_{it}^k = e_{it}^k / E_{it}$, capturing the fraction of education spending allocated to each program k. By construction, these shares sum to one within each municipality-year, making this a compositional analysis: an increase in the share devoted to primary school facilities necessarily implies a decrease elsewhere. I also compute education's share of total spending (E_{it}/T_{it}) and education spending per capita (E_{it}/pop_{it}) to compare with the aggregate results of Bagues and Campa (2021).

3.2 Election Data: Candidate Records

Election data come from the Ministry of the Interior (*Ministerio del Interior*), which publishes detailed candidate-level records for all Spanish elections. The fixed-width data files (Type 04) contain each candidate's position on the list, name, gender (directly coded as M/F), date of birth, and whether they were elected.

I parse candidate records for the five municipal elections held during the study period: 2007, 2011, 2015, 2019, and 2023 (gender data from the 2003 election are unavailable in the Ministry of the Interior records). For each municipality-election, I compute:

- **Female councillor share**: The fraction of elected council members who are female.
- **Number of female councillors** and **total councillors**.

The female councillor share is the first-stage outcome in the RDD analysis. I match each fiscal year's budget data to the governing council by assigning budget years to the most recent election (e.g., fiscal years 2011–2014 are matched to the 2011 election council).

3.3 Population Data: Padrón Municipal

The running variable for the RDD is municipal population, obtained from the INE's *Padrón Municipal de Habitantes*. The Padrón is Spain's continuous population register, updated annually, and provides the official population count that determines a municipality's regulatory treatment.

I use the average population across all available Padrón years within each municipality as the running variable. This choice involves a trade-off. The institutional rule assigns treatment based on the official Padrón population at the time of each election, which varies over time and could in principle be analyzed election-by-election. However, for the cross-sectional RDD design employed here—where I average both outcomes and running variables across the panel—the municipality-level mean population reduces measurement error from year-to-year fluctuations in registration. Municipal populations in the 2,000–8,000 range are highly persistent: the within-municipality standard deviation of population is only about 150 inhabitants (less than 4% of the mean), so fewer than 3% of municipalities in the sample cross the 5,000 threshold during the panel. The average population is therefore a close proxy for election-year population in the vast majority of cases. As a robustness check, the results are stable across different bandwidth choices, which implicitly tests sensitivity to the precise running variable values near the cutoff.

3.4 Sample Construction

The analysis sample focuses on municipalities near each population threshold. For the 5,000-inhabitant cutoff, I include municipalities with populations between 2,000 and 8,000. For the 3,000-inhabitant cutoff (active from 2011), I include those between 1,000 and 5,000. These windows are wide enough to ensure adequate power while restricting attention to municipalities that are plausibly comparable.

I exclude municipality-years with zero total education spending (as these likely reflect reporting errors) and those with missing population data. The final panel contains thousands of municipality-year observations spanning 14 fiscal years.

3.5 Summary Statistics

Table 1 presents summary statistics for municipalities near the 5,000 threshold. The average municipality in this range has a population of approximately 4,100 and devotes about 3.5 percent of its total budget to education—roughly EUR 133 per capita. The average female councillor share is 40 percent, reflecting the strong post-2007 trend toward gender parity. The within-education shares reveal substantial heterogeneity: education promotion (program 323) accounts for the largest average share at 17 percent, while primary school facilities (program 321) average about 10 percent and complementary services (program 326) average 5 percent. The large standard deviations for total and social spending per capita relative to their means reflect a small number of extreme outliers; the interquartile range is narrow, as the P10–P90 columns confirm.

Table 1: Summary Statistics: Municipalities Near 5,000 Threshold

Variable	N	Mean	SD	P10	Median	P90
Population	18046	4117.297	1672.594	2234.000	3697.000	6766.000
Female Councillor Share	14531	0.403	0.109	0.273	0.385	0.545
Education Spending p.c. (EUR)	18046	132.715	296.423	14.961	88.409	236.801
Total Spending p.c. (EUR)	18046	3972.024	15882.386	1908.601	2969.019	4922.817
Education Share of Total	18046	0.035	0.030	0.006	0.029	0.069
Security Spending p.c. (EUR)	18046	99.381	175.925	0.206	71.911	206.929
Social Spending p.c. (EUR)	18046	272.657	2952.098	8.176	103.301	408.462
General Administration	17427	0.088	0.162	0.000	0.000	0.429
Primary School Facilities	17427	0.105	0.174	0.000	0.000	0.451
Secondary Education	17427	0.003	0.029	0.000	0.000	0.000
Education Promotion	17427	0.174	0.197	0.000	0.064	0.494
Non-Regulated Education	17427	0.031	0.092	0.000	0.000	0.094
University Education	17427	0.001	0.013	0.000	0.000	0.000
Complementary Services	17427	0.050	0.104	0.000	0.000	0.188
Special Education	17427	0.004	0.028	0.000	0.000	0.000

4. Empirical Strategy

4.1 Regression Discontinuity Design

I estimate the causal effect of gender quotas on within-education spending composition using a sharp regression discontinuity design. The identifying variation comes from the population thresholds that determine quota applicability. Let X_i denote the average population of municipality i and c the cutoff (5,000 or 3,000). Since the law applies to municipalities with

more than c inhabitants, the treatment indicator is:

$$D_i = \mathbb{I}[X_i > c] \tag{1}$$

In practice, the running variable is a continuous average, making exact ties at the cutoff negligible. The `rdrobust` package uses the convention $D_i = \mathbb{I}[X_i \geq c]$, which is equivalent when the running variable has no mass point at c.

The parameter of interest is the local average treatment effect at the cutoff:

$$\tau = \lim_{x \downarrow c} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i | X_i = x] \tag{2}$$

where Y_i is a within-education budget share or another outcome of interest.

The key identifying assumption is continuity of potential outcomes at the cutoff: in the absence of the quota, municipalities just above and just below the threshold would have similar spending patterns. This assumption is standard in the RDD literature (Imbens and Lemieux, 2008; Lee and Lemieux, 2010) and is supported by the institutional context: the population thresholds were set for administrative convenience (council size rules) and were not designed with education spending in mind.

4.2 Estimation

I estimate local linear regressions using the bias-corrected robust inference procedures of Calonico et al. (2014):

$$Y_i = \alpha + \tau D_i + \beta_1 (X_i - c) + \beta_2 D_i (X_i - c) + \varepsilon_i \tag{3}$$

Bandwidth selection follows the MSE-optimal procedure of Calonico et al. (2014), implemented via the `rdrobust` package. I use a triangular kernel, which places greater weight on observations closer to the cutoff. The primary specification uses a first-order polynomial ($p = 1$), with higher-order polynomials examined as robustness checks.

Because the analysis involves multiple within-education share outcomes, I report results for each subcategory separately rather than testing a single composite outcome. This approach allows me to identify which specific budget lines are affected, providing interpretable evidence on mechanisms.

4.3 First Stage: Quota Compliance

The first stage verifies that the population thresholds produce a discontinuous increase in female representation. I estimate:

$$\text{FemaleShare}_i = \alpha_0 + \gamma D_i + \delta_1(X_i - c) + \delta_2 D_i(X_i - c) + \nu_i \tag{4}$$

A strong first stage—a large and statistically significant γ—is necessary for the reduced-form estimates to have a causal interpretation as the effect of female representation on spending composition. In this design, the reduced-form (intention-to-treat) estimates capture the composite effect of the quota on spending, operating through increased female representation and potentially through other channels such as changes in candidate quality (Baltrunaite et al., 2014).

4.4 Threats to Identification

Three main threats could undermine the RDD identification.

Manipulation of the running variable. If municipalities can strategically adjust their registered population to avoid the quota threshold, the continuity assumption fails. I address this concern with the density test of Cattaneo et al. (2020), which tests for discontinuities in the density of the running variable at the cutoff. The Padrón is an administrative register that municipalities have limited ability to manipulate, and the population counts are verified by the INE.

Other policies at the same threshold. If other policies also change at 5,000 or 3,000 inhabitants, the estimated effect could reflect those policies rather than the gender quota. At 5,000 inhabitants, council size increases from 11 to 13 members, which could independently affect budgeting through committee structure, coalition formation, or administrative capacity. This confound is inherent to the 5,000 cutoff and means the reduced-form estimates capture the joint effect of the quota and the council size change. At the 3,000 cutoff, no such council size change occurs, making it a cleaner test of the quota effect. The null results at 3,000 are therefore particularly informative. Other administrative obligations shift at various population thresholds, but the 5,000 threshold for gender quotas was introduced specifically by the 2007 Equality Law and no concurrent education-specific policy change applies at this cutoff.

Sorting of municipalities. If municipalities that happen to be above the threshold differ systematically from those below in ways that affect education spending, the RDD estimates would be biased. I test for this by examining covariate balance: spending levels and other fiscal variables should be smooth through the cutoff if sorting is not a concern.

5. Results

5.1 First Stage: Gender Quotas and Female Representation

Figure 1 presents the first-stage relationship between the population thresholds and female council representation. Panel A shows the 5,000-inhabitant cutoff: any discontinuity in the relationship between population and female councillor share is small relative to the noise in the binscatter, consistent with the statistically insignificant first-stage estimate. Panel B shows the 3,000-inhabitant cutoff for the post-2011 period, where a marginally significant negative discontinuity is visible.

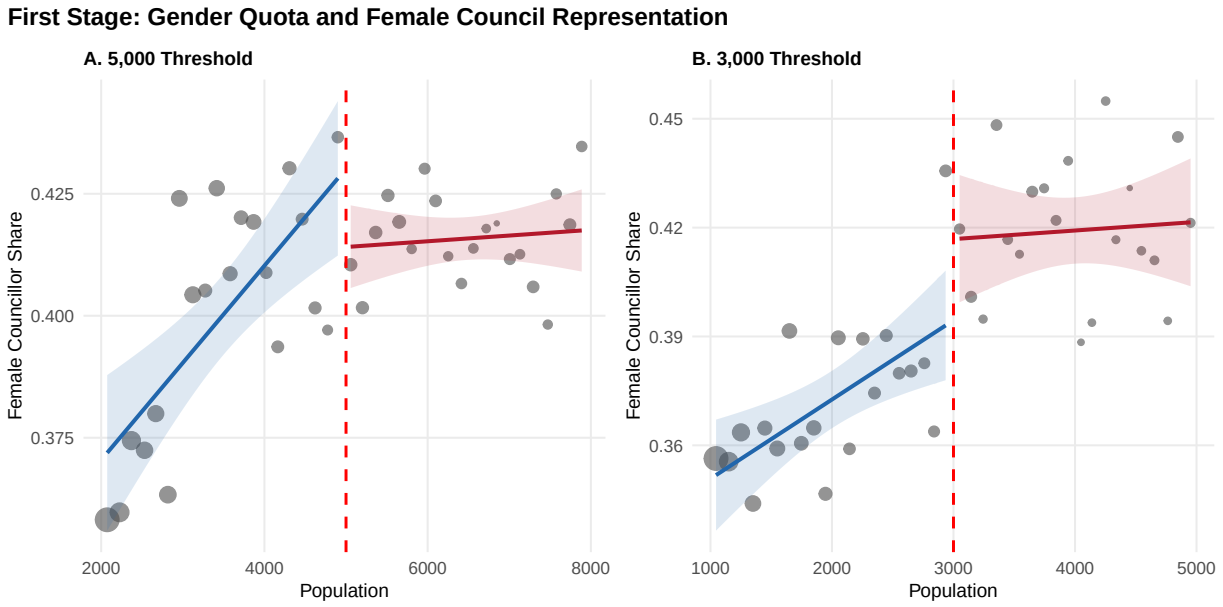

Figure 1: First Stage: Gender Quota and Female Council Representation

Notes: Each panel shows a binscatter of average female councillor share against municipal population, with local linear fits on each side of the cutoff. The vertical dashed line indicates the population threshold. Panel A: 5,000-inhabitant threshold, all election years. Panel B: 3,000-inhabitant threshold, 2011 onward.

Table 2 reports the formal RD estimates. At the 5,000 threshold, the estimate is -0.011 (SE $= 0.016$, $p = 0.49$), with an MSE-optimal bandwidth of 876 inhabitants and an effective sample of 260 municipalities. The point estimate is small and statistically insignificant. At the 3,000 threshold, the estimate is -0.038 (SE $= 0.021$, $p = 0.07$), marginally significant but with the wrong sign—suggesting that, if anything, municipalities *above* the threshold have slightly *fewer* female councillors.

This weak first stage contrasts with the sharp compliance effects documented by Bagues and

Campa (2021) for the 2007–2011 elections. The difference likely reflects temporal attenuation: as the gender parity norm diffused across the political spectrum, parties voluntarily adopted balanced lists even in municipalities below the threshold. By averaging across all elections from 2007 to 2023, the first stage captures both the initial sharp compliance effect and the subsequent convergence. This finding has implications for the design: the reduced-form RDD estimates should be interpreted as intention-to-treat effects of the quota *threshold*, not as the causal effect of female representation per se.

Table 2: First Stage: Effect of Gender Quota on Female Councillor Share

Cutoff	Estimate	SE	p-value	Bandwidth	N (left)	N (right)
5,000	-0.0109	0.0157	0.4864	876	127	133
3,000	-0.0380	0.0207	0.0662	360	124	93

5.2 Validity Tests

5.2.1 Density Test

Figure 2 and Table 3 present the results of the Cattaneo et al. (2020) density test. At both the 5,000 and 3,000 cutoffs, the test fails to reject the null hypothesis of no density discontinuity. There is no evidence that municipalities systematically sort around the thresholds.

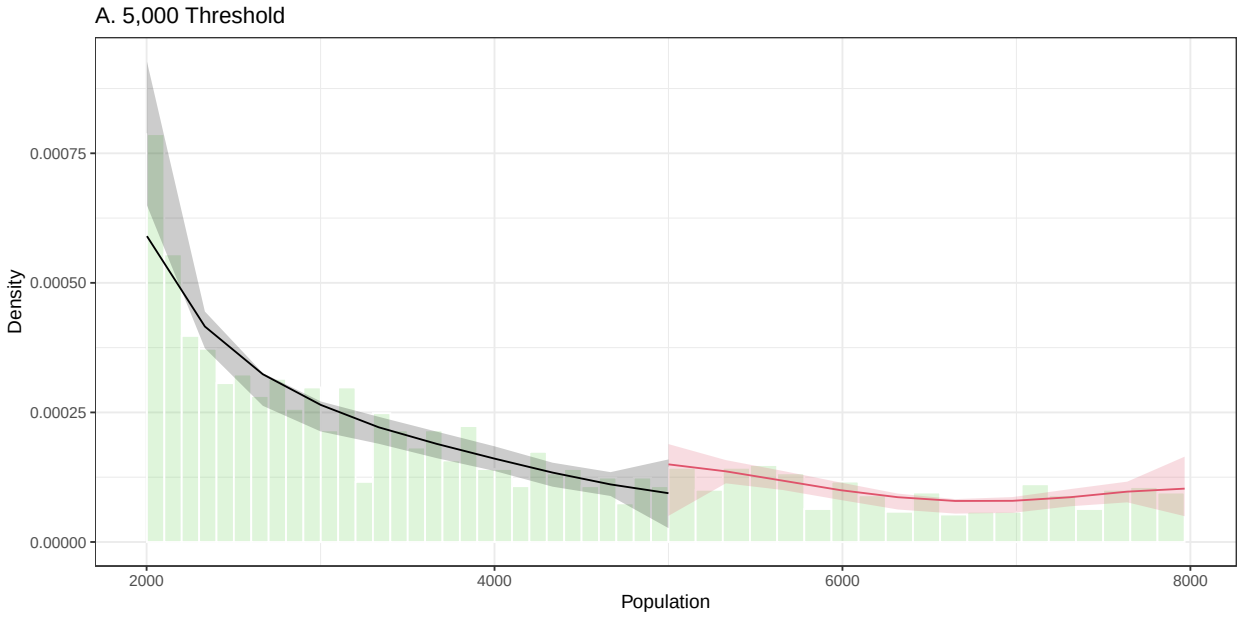

Figure 2: McCrary Density Test at Population Thresholds
Notes: Density estimates following Cattaneo et al. (2020). Panel A: 5,000-inhabitant threshold. Panel B: 3,000-inhabitant threshold.

Table 3: McCrary Density Tests at Population Thresholds

Cutoff	T-statistic	p-value
5,000	0.538	0.5905
3,000	-0.394	0.6933

5.2.2 Covariate Balance

Table 4 tests for discontinuities in baseline (2010) spending variables at the 5,000 threshold. While 2010 falls after the 2007 introduction of the quota, these fiscal variables—total spending, education spending, security spending, and social spending per capita—are predetermined with respect to the population threshold and should be smooth through the cutoff if municipalities did not sort around it. All four covariates show no significant discontinuity, supporting the identification assumption that municipalities on either side of the threshold are comparable in their fiscal characteristics.

Table 4: Covariate Balance at 5,000 Population Threshold (2010)

Variable	Estimate	SE	p-value	Bandwidth	N (left)	N (right)
Total Spending p.c. (EUR)	-145.53	430.36	0.735	348	60	71
Education Spending p.c. (EUR)	4.09	42.25	0.923	521	87	110
Security Spending p.c. (EUR)	-10.34	29.31	0.724	633	113	126
Social Spending p.c. (EUR)	46.48	51.38	0.366	342	60	69

5.3 Main Results: Within-Education Spending Composition

Table 5 presents the core results of the paper. Each row reports the RD estimate for a different within-education budget share at the 5,000-inhabitant threshold. The effective sample sizes (N left and N right) vary across outcomes because `rdrobust` selects a separate MSE-optimal bandwidth for each regression; these are the municipalities within the chosen bandwidth, not the full panel.[1]

The results reveal no significant compositional shift at the 5,000 threshold in the full sample. None of the eight disaggregated education program shares shows a statistically significant discontinuity at conventional levels. The largest point estimates are for primary

[1]The total sample contains 18,046 municipality-year observations (Table 1), which are averaged to municipality-level means for the cross-sectional RDD. The effective N in each regression is the number of municipalities within the outcome-specific optimal bandwidth. For the same reason, sub-period regressions (Table 9) may have wider bandwidths—and thus larger effective N—than full-sample regressions, because the sparser data in each sub-period induces the bandwidth selector to widen the window.

Table 5: Main RDD Results: Within-Education Budget Shares and Aggregate Outcomes

Cutoff	Variable	Estimate	SE	p-value	Stars	Bandwidth	N (left)	N (right)
5,000	General Administration	0.0285	0.0311	0.3607		818	150	154
5,000	Primary School Facilities	0.0358	0.0248	0.1490		1138	213	205
5,000	Secondary Education	-0.0015	0.0034	0.6618		878	157	160
5,000	Education Promotion	-0.0435	0.0360	0.2275		780	145	152
5,000	Non-Regulated Education	-0.0078	0.0105	0.4580		819	150	154
5,000	University Education	0.0004	0.0005	0.4123		712	124	143
5,000	Complementary Services	0.0015	0.0147	0.9180		1218	237	216
5,000	Special Education	-0.0076	0.0058	0.1913		1214	236	216
5,000	Educ. Share of Total	0.0032	0.0061	0.5967		784	146	152
5,000	Security Spending p.c.	-14.1896	24.7836	0.5670		610	102	121
3,000	General Administration	-0.0322	0.0235	0.1700		807	363	253
3,000	Primary School Facilities	-0.0122	0.0221	0.5793		665	291	206
3,000	Secondary Education	-0.0022	0.0025	0.3692		550	233	172
3,000	Education Promotion	0.0062	0.0264	0.8141		625	274	189
3,000	Non-Regulated Education	0.0071	0.0142	0.6165		535	227	168
3,000	University Education	-0.0008	0.0028	0.7801		362	147	113
3,000	Complementary Services	0.0149	0.0156	0.3386		639	278	198
3,000	Special Education	0.0022	0.0033	0.5129		528	222	166

school facilities (+0.036, $p = 0.15$) and education promotion (-0.044, $p = 0.23$), but neither survives correction for multiple testing. The aggregate education share of total spending is also insignificant (+0.003, $p = 0.60$), replicating the canonical null of Bagues and Campa (2021).

Figure 3 provides visual evidence. The binscatter plots show the relationship between municipal population and each education subcategory share. The conditional expectation functions are smooth through the 5,000-inhabitant threshold, with no visible discontinuities.

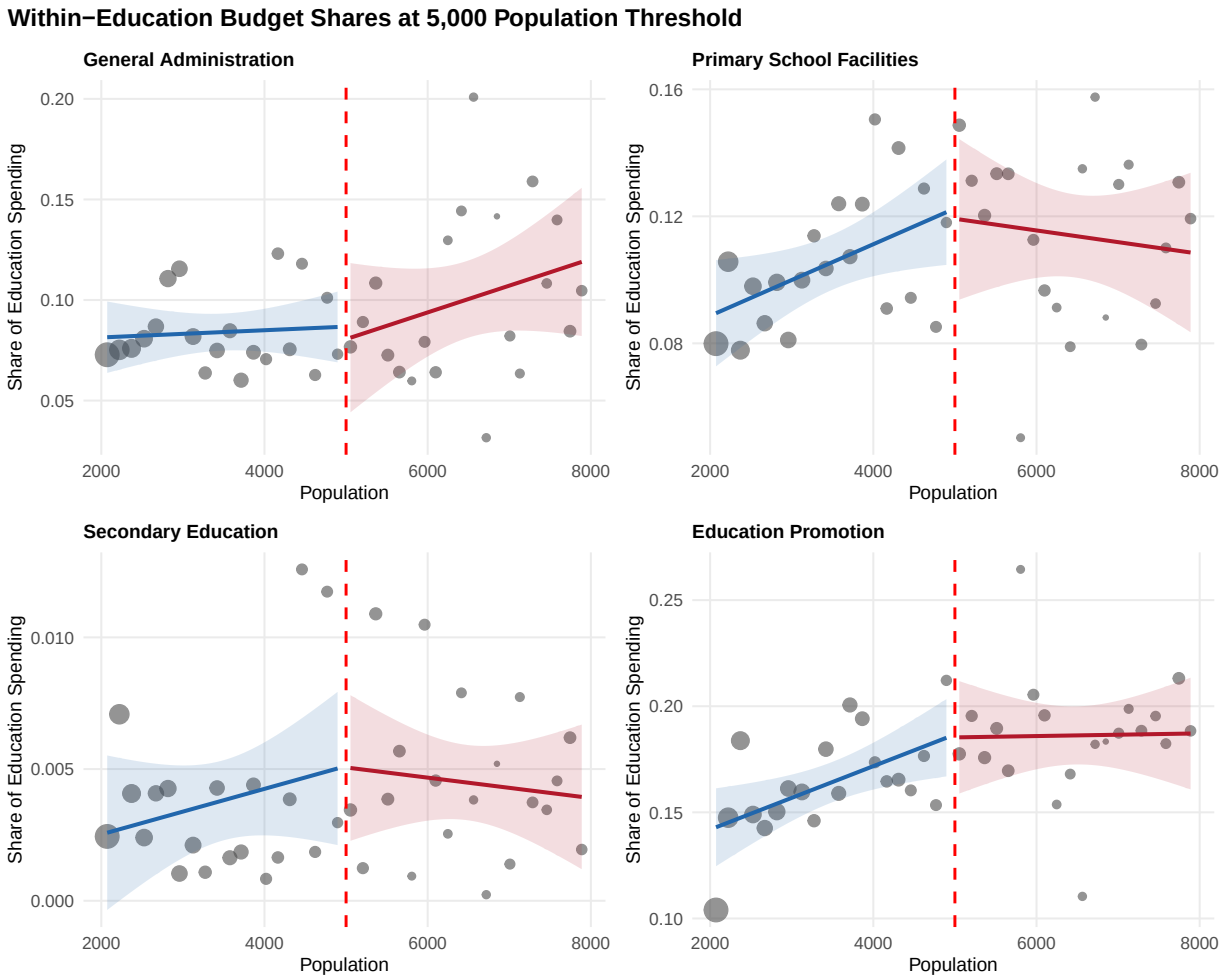

Figure 3: Within-Education Budget Shares at 5,000 Population Threshold

Notes: Each panel shows a binscatter of the within-education budget share for a specific program against municipal population, with local linear fits on each side of the 5,000-inhabitant cutoff.

5.4 Replicating the Aggregate Null

Figure 4 confirms the well-documented aggregate null: the share of total municipal spending devoted to education is smooth through the 5,000 cutoff. The formal RD estimate (+0.003, SE = 0.006, $p = 0.60$) is small and statistically insignificant, replicating Bagues and Campa (2021) with independent data covering 2010–2023.

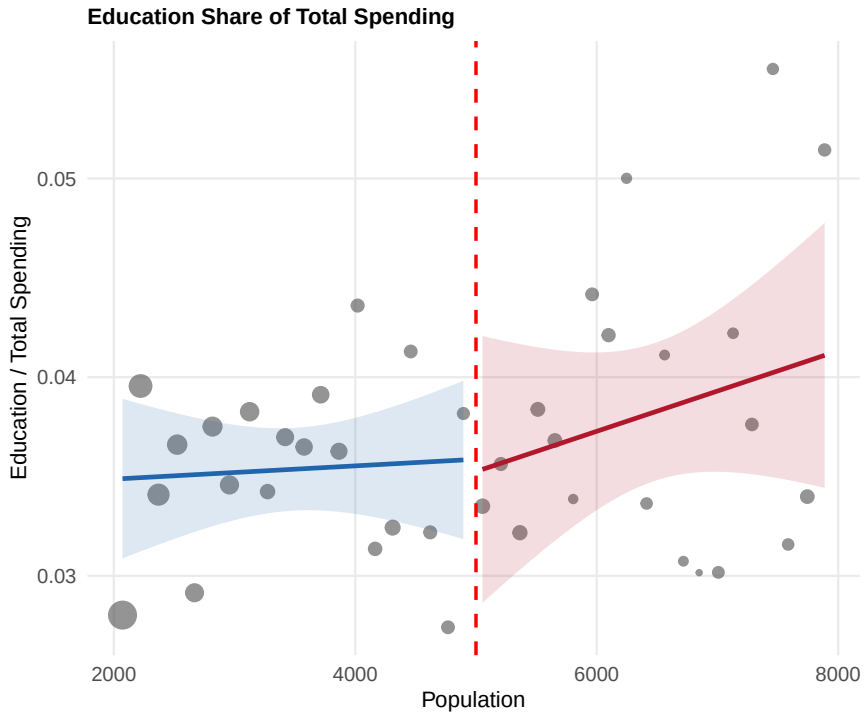

Figure 4: Aggregate Null: Education Share of Total Spending at 5,000 Threshold
Notes: Binscatter of education's share of total municipal spending against population. No discontinuity at 5,000 inhabitants, replicating Bagues and Campa (2021).

This aggregate null, together with the within-education nulls, establishes a comprehensive negative result: the gender quota threshold does not detectably shift municipal education spending at any level of aggregation in the full sample period.

5.5 Placebo Outcome: Security Spending

As an additional check, I examine security spending per capita, a category that should not respond to the gender quota through a gendered preference channel. Figure 5 shows the RD plot. The estimate (-14.19 EUR per capita, SE $= 24.78$, $p = 0.57$) is statistically insignificant, consistent with the absence of a general fiscal discontinuity at the threshold.

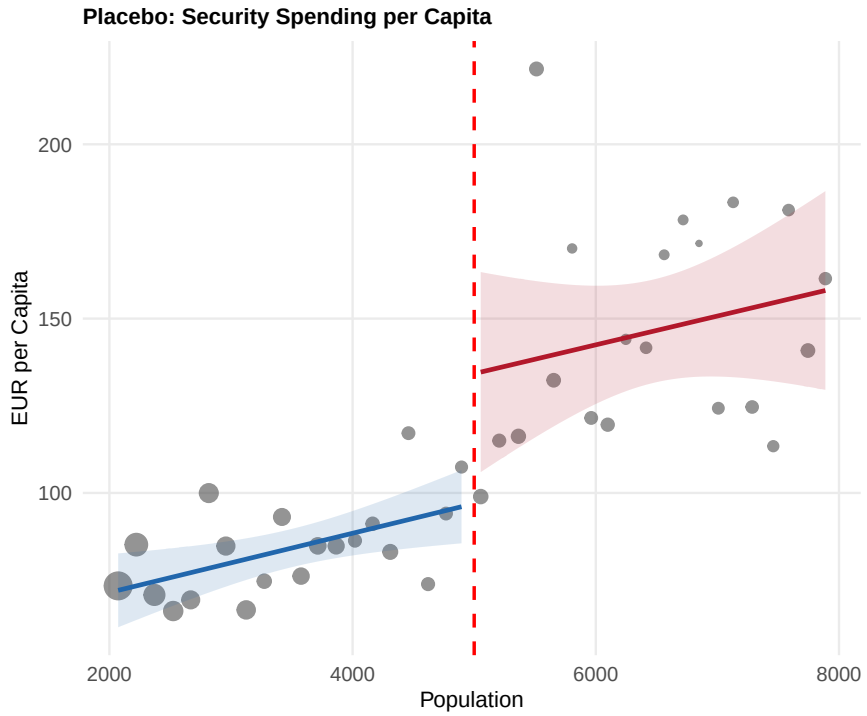

Figure 5: Placebo: Security Spending per Capita at 5,000 Threshold
Notes: Binscatter of security spending per capita against population. No discontinuity at the quota threshold.

5.6 Results at the 3,000 Threshold

The 3,000-inhabitant threshold provides a replication opportunity. This cutoff was introduced in 2011, so the 3,000-cutoff estimates are restricted to post-2011 election terms only. It affects a population of smaller, more rural municipalities.

The results at 3,000 are consistent with the null findings at 5,000. The largest point estimate is for complementary services (+0.015, $p = 0.34$), but no subcategory approaches statistical significance. The 3,000 threshold was active for a shorter post-treatment period, and the affected municipalities are smaller and more rural, but the pattern is unambiguous: both thresholds produce null effects on within-education budget composition in the full sample.

6. Robustness and Heterogeneity

6.1 Donut RDD

A concern with any RDD is that observations very close to the cutoff may be subject to manipulation or measurement error in the running variable. The donut RDD excludes

municipalities within a specified radius of the threshold. Table 6 reports estimates using donut radii of 100 and 200 inhabitants. The null findings are robust: the estimates remain insignificant across both donut specifications, confirming that the full-sample null is not driven by municipalities at the precise cutoff.

Table 6: Robustness: Donut RDD Estimates

Donut	Variable	Estimate	SE	p-value	Bandwidth	N (left)	N (right)
100	General Administration	0.0240	0.0427	0.5739	740.8765	123	127
100	Primary School Facilities	0.0294	0.0367	0.4222	750.4298	123	127
100	Secondary Education	-0.0018	0.0056	0.7552	1184.0694	214	191
100	Education Promotion	-0.0211	0.0387	0.5850	789.1634	133	130
200	General Administration	0.0255	0.0795	0.7484	686.0582	87	100
200	Primary School Facilities	0.0438	0.0530	0.4077	808.1095	117	112
200	Secondary Education	-0.0039	0.0091	0.6702	1101.8569	171	158
200	Education Promotion	0.0742	0.0671	0.2692	674.9290	83	96

6.2 Bandwidth Sensitivity

Figure 6 and Table 7 examine how the main estimates vary with the choice of bandwidth. I estimate the primary specification at 0.5, 0.75, 1.0, 1.25, 1.5, and 2.0 times the MSE-optimal bandwidth. The estimates are stable near zero across this range, with a slight trend toward marginally negative values at wider bandwidths (reaching -0.015 with $p = 0.06$ at twice the optimal). This pattern is consistent with a small negative association between population and the education administration share that is not attributable to the quota per se.

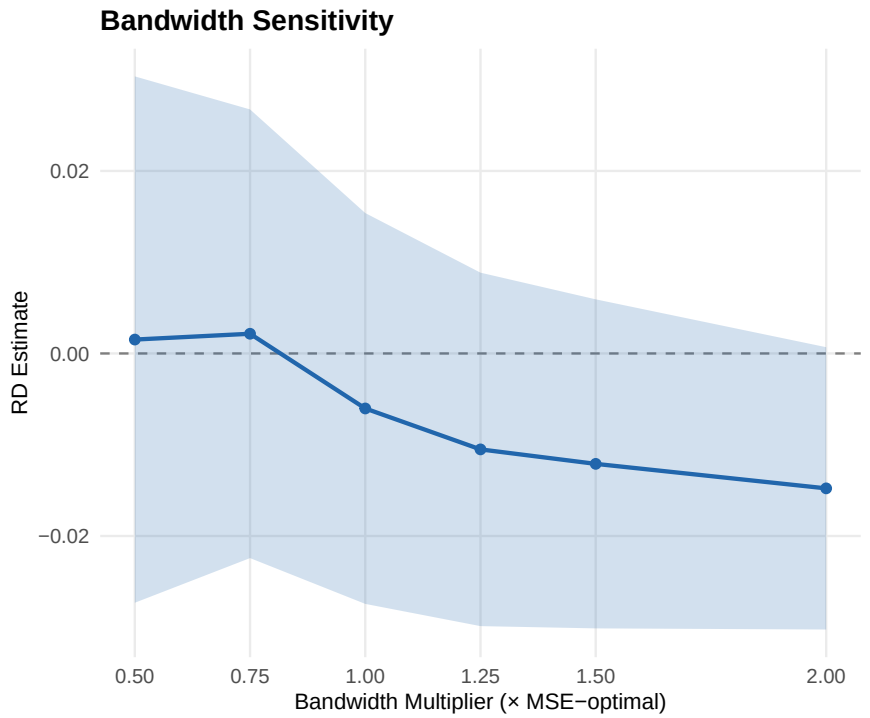

Figure 6: Bandwidth Sensitivity

Notes: RD estimate and 95% confidence interval as a function of the bandwidth multiplier (relative to the MSE-optimal bandwidth). The vertical line at 1.0 marks the baseline specification.

Table 7: Robustness: Bandwidth Sensitivity

BW Multiplier	Bandwidth	Estimate	SE	p-value	
0.50	348	0.0015	0.0147	0.9176	
0.75	521	0.0022	0.0125	0.8633	
1.00	695	-0.0060	0.0109	0.5812	
1.25	869	-0.0105	0.0099	0.2874	
1.50	1043	-0.0121	0.0092	0.1885	
2.00	1390	-0.0148	0.0079	0.0611	*

6.3 Pre-Treatment Placebo

The 3,000-inhabitant threshold was introduced in 2011. Budget data from 2010 thus represent a pre-treatment period for this cutoff: the quota did not apply, so there should be no discontinuity in within-education shares at 3,000 inhabitants. Table 8 shows that most pre-treatment placebo estimates are insignificant, though secondary education (program

322) is marginally significant at $p = 0.06$. These marginal pre-treatment differences warrant caution in interpreting results at the 3,000 cutoff, as they may reflect pre-existing fiscal heterogeneity correlated with population.

6.4 Placebo Cutoffs

I estimate the RDD specification at 4,000 and 6,000 inhabitants—population values where no policy change occurs. Table 8 reports the results: the placebo cutoff estimates are uniformly small and insignificant, confirming that there are no spurious discontinuities at nearby population values. This validates the RDD design even though the true threshold effects are also null—it rules out the possibility that the population running variable is inherently noisy or confounded by other population-based policies.

Table 8: Placebo Tests: Pre-Treatment and False Cutoffs

Test	Variable	Estimate	SE	p-value	Stars
Pre-treatment (2010)	General Administration	-0.0408	0.0268	0.1276	
Pre-treatment (2010)	Primary School Facilities	-0.0532	0.0474	0.2616	
Pre-treatment (2010)	Secondary Education	0.0064	0.0034	0.0601	*
Pre-treatment (2010)	Education Promotion	0.0075	0.0202	0.7099	
Pre-treatment (2010)	Non-Regulated Education	0.0156	0.0213	0.4639	
Placebo cutoff: 4,000	General Administration	0.0171	0.0270	0.5266	
Placebo cutoff: 4,000	Primary School Facilities	-0.0111	0.0238	0.6400	
Placebo cutoff: 6,000	General Administration	0.0447	0.0321	0.1642	
Placebo cutoff: 6,000	Primary School Facilities	-0.0290	0.0200	0.1458	

6.5 Polynomial Order Sensitivity

The baseline specification uses a first-order (linear) polynomial on each side of the cutoff. I verify that the results are not sensitive to this choice by estimating with second- and third-order polynomials. For one illustrative outcome—the general administration share (program 320) at the 5,000 cutoff—the estimates are -0.006 (SE $= 0.011$, $p = 0.58$) with a linear fit ($p = 1$), $+0.003$ (SE $= 0.014$, $p = 0.82$) with a quadratic fit ($p = 2$), and $+0.000$ (SE $= 0.014$, $p = 0.98$) with a cubic fit ($p = 3$). All are statistically insignificant, with the expected increase in standard errors at higher polynomial orders. The pattern is similar for other subcategories.

6.6 Heterogeneity: Pre- and Post-LRSAL

Spain's 2013 *Ley de Racionalización y Sostenibilidad de la Administración Local* (LRSAL) restricted municipal competences and imposed fiscal discipline. I split the sample into a pre-LRSAL period (2010–2013) and a post-LRSAL period (2014–2023) and estimate the RDD separately for each.

Table 9 presents the temporal decomposition. In the pre-LRSAL period (2010–2013), the quota threshold produces a positive discontinuity in the share of education spending allocated to primary school facilities (estimate = 0.093, SE = 0.043, $p = 0.032$). In the post-LRSAL period, all effects vanish: primary school facilities shows an insignificant -0.015 ($p = 0.56$) and general administration an insignificant $+0.008$ ($p = 0.84$). The aggregate education share is null in both periods.

Table 9: Temporal Heterogeneity: Pre- and Post-LRSAL RDD Results at 5,000 Threshold

Period	Variable	Estimate	SE	p-value		Bandwidth	N (left)	N (right)
Pre-LRSAL (2010–2013)	General Administration	-0.0285	0.0317	0.3699		1241	252	221
Pre-LRSAL (2010–2013)	Primary School Facilities	0.0926	0.0433	0.0324	**	1034	204	182
Pre-LRSAL (2010–2013)	Secondary Education	-0.0119	0.0096	0.2151		943	176	167
Pre-LRSAL (2010–2013)	Education Promotion	-0.0044	0.0311	0.8874		950	178	169
Pre-LRSAL (2010–2013)	Non-Regulated Education	0.0315	0.0255	0.2181		862	154	156
Pre-LRSAL (2010–2013)	Educ. Share of Total	-0.0012	0.0065	0.8497		896	170	161
Post-LRSAL (2014–2023)	General Administration	0.0076	0.0367	0.8358		888	163	158
Post-LRSAL (2014–2023)	Primary School Facilities	-0.0149	0.0258	0.5633		703	121	133
Post-LRSAL (2014–2023)	Secondary Education	0.0007	0.0014	0.6256		507	89	104
Post-LRSAL (2014–2023)	Education Promotion	0.0248	0.0414	0.5492		815	148	149
Post-LRSAL (2014–2023)	Non-Regulated Education	-0.0036	0.0083	0.6607		1092	203	185
Post-LRSAL (2014–2023)	University Education	-0.0003	0.0002	0.2798		401	69	80
Post-LRSAL (2014–2023)	Complementary Services	-0.0105	0.0217	0.6292		1117	214	190
Post-LRSAL (2014–2023)	Special Education	-0.0171	0.0148	0.2468		953	179	168
Post-LRSAL (2014–2023)	Educ. Share of Total	0.0009	0.0069	0.8937		755	135	142

Two important caveats qualify this finding. First, sub-period first-stage estimates for female representation at the 5,000 threshold are weak in both periods: -0.024 (SE = 0.018, $p = 0.17$) pre-LRSAL and -0.019 (SE = 0.017, $p = 0.25$) post-LRSAL. The compositional shift in the pre-LRSAL period therefore cannot be confidently attributed to increased female representation; it should be interpreted as an ITT effect of the quota threshold, which bundles representation changes with other threshold-specific factors (including council size, which increases from 11 to 13 seats at 5,000 inhabitants).

Second, with eight education subcategories tested, the uncorrected $p = 0.032$ for primary school facilities would not survive a Bonferroni adjustment ($p_{\text{adj}} = 0.26$). Under the less conservative Benjamini-Hochberg procedure, which controls the false discovery rate, the result is similarly marginal. The pre-LRSAL finding is therefore best viewed as suggestive rather

than definitive—a pattern worth investigating in other institutional settings rather than a causal conclusion.

Despite these caveats, the temporal contrast is informative. The pattern—a compositional shift in the pre-austerity period that vanishes under fiscal constraints—is consistent with institutional frameworks that limit municipal discretion. Before the fiscal consolidation of 2013, municipalities had broader competences, and the quota threshold is associated with a shift toward primary school facility spending. After the LRSAL restricted competences and imposed fiscal discipline, no such shift is detectable. Whether this reflects the elimination of political discretion, the attenuation of the first stage, or both, remains an open question.

The finding that the effect appears in program 321 (primary school facilities), classified as partly mandatory, warrants discussion. While municipalities are obligated to provide school buildings, they retain discretion over the timing, scope, and prioritization of facility maintenance and renovation projects. A council that chooses to invest in upgrading school kitchens or adding accessibility features is exercising genuine discretion within a mandatory framework. The distinction between "mandatory" and "discretionary" is thus blurrier in practice than in legal classification.

7. Discussion

7.1 Interpreting the Null

The dominant finding is a precisely estimated null: gender quota thresholds do not detectably shift within-education budget composition in the full 2010–2023 sample. This null operates at every level of aggregation tested—from eight three-digit program categories to the aggregate education share of total spending.

The null is informative for three reasons. First, it refines the existing literature's aggregate null by showing that the result is not an artifact of measurement coarseness. Even with the most granular budget data available in Spain—eight education subcategories with distinct policy content—no compositional shift is detectable. Second, the null constrains theories of gendered preference transmission. Models in which female councillors mechanically redirect resources toward "women's issues" (Chattopadhyay and Duflo, 2004) predict effects at any level of disaggregation; the Spanish null suggests that such transmission does not operate in the European council setting, at least not through the budget channel. Third, the pre-LRSAL exception provides a crucial qualifier: the null may reflect institutional constraints rather than the absence of preferences.

7.2 The Pre-LRSAL Exception

The suggestive reallocation toward primary school facilities in 2010–2013 is consistent with the hypothesis that council composition can affect spending when institutional constraints permit. However, the weak first stage on female representation in this period—and the inability of the uncorrected p-value to survive standard multiple testing adjustments—means this finding should be treated as preliminary evidence rather than a confirmed mechanism.

Three interpretations are consistent with the data. First, the pre-LRSAL period may have offered genuine discretion that municipalities above the quota threshold used differently, perhaps through channels beyond female representation (e.g., the council size increase from 11 to 13 members at 5,000 inhabitants could affect committee structure or coalition formation). Second, the finding could reflect a Type I error among many subcategory tests. Third, the quota threshold may have initially produced compositional effects through female representation when the first stage was stronger (in the immediate post-2007 elections), effects that attenuated as voluntary compliance diffused.

This finding resonates with the broader literature on institutional constraints and policy responsiveness. Gagliarducci and Paserman (2012) document null effects of female mayors in Italian municipalities where strong party discipline limits individual influence. Ferreira and Gyourko (2014) find that U.S. city demographics and institutions explain spending patterns far better than mayor gender. The common thread is that institutions mediate the transmission of political preferences into policy outcomes.

7.3 Why the First Stage Attenuated

The weak first stage deserves its own interpretation. When Bagues and Campa (2021) studied the 2007 and 2011 elections—the first two held under the quota law—the compliance discontinuity was sharp. My analysis, extending through 2023, shows that the treatment contrast has diffused. The mean female councillor share rose from 30.6% in 2007 to 41.7% in 2023 across all municipalities in the sample, including those below the threshold. This convergence suggests that the legal mandate catalyzed a broader norm change that rendered the threshold non-binding.

This temporal evolution has methodological implications. Population-threshold RDD designs in the gender quota literature implicitly assume a stable treatment contrast over time. When norm change follows legal mandate, the first stage decays and the reduced-form estimates attenuate, even if the quota *initially* had compositional effects. The pre-LRSAL finding is consistent with this story: effects were present when the first stage was strong and institutional constraints were loose.

7.4 Limitations

Several limitations merit discussion. First, the weak first stage limits the reduced-form estimates' interpretability as effects of female representation. The estimates should be interpreted as ITT effects of the quota threshold, which bundles increased female representation with other possible threshold effects (council size changes, party list composition adjustments).

Second, the RDD estimates are local to the population thresholds. Municipalities with 5,000 inhabitants are small by European standards, and the results may not generalize to larger cities where council dynamics and budget structures differ.

Third, the within-education shares are compositional: a change in one share mechanically implies changes in others. While this is economically meaningful, it complicates the interpretation of individual program estimates.

Fourth, the pre-LRSAL finding, while nominally significant at $p = 0.032$, comes from a short four-year window (2010–2013), involves a single program category among eight tested, and would not survive standard multiple testing corrections (Bonferroni-adjusted $p = 0.26$). The sub-period first stage on female representation is also weak (-0.024, $p = 0.17$), so the compositional shift cannot be confidently attributed to increased female representation as opposed to other threshold-specific factors. I report it as suggestive evidence rather than a definitive finding.

Fifth, the 5,000-inhabitant threshold bundles the gender quota with a council size increase (from 11 to 13 members). With a weak first stage on female representation, the reduced-form estimates cannot cleanly disentangle quota effects from council size effects. The 3,000 threshold—where no council size change occurs—provides a cleaner test, and the null results there are informative.

Sixth, I observe budget allocations but not outcomes. Whether any compositional shift improves educational outcomes for children is an important question that this paper cannot answer.

7.5 Implications for the Gender and Public Finance Literature

The results of this paper have implications that extend beyond the specific Spanish case. A growing literature has documented the persistent puzzle that gender quotas change who governs but not what governments do. The standard response has been to look for effects in increasingly disaggregated outcome measures—from total spending to spending categories, from spending categories to specific programs. This paper pushes that strategy to its logical conclusion by examining eight three-digit program codes within a single policy domain, and finds that even at this granularity the full-sample null persists.

The implication is that the measurement explanation for the aggregate null—that effects exist but are hidden by coarse data—may be insufficient. Instead, the institutional explanation deserves greater weight: European municipal councils operate under sufficiently strong institutional constraints (intergovernmental transfer rules, mandatory service provision standards, collective decision-making through proportional representation) that individual politicians' preferences have limited scope for expression in the budget. This contrasts sharply with the Indian *gram panchayat* setting studied by Chattopadhyay and Duflo (2004), where village heads had substantial discretionary authority over public goods provision.

The pre-LRSAL finding strengthens rather than weakens this institutional interpretation. When institutional constraints briefly relaxed (before the 2013 austerity law), a compositional effect appeared. When they tightened, it vanished. This pattern is diagnostic: it demonstrates that the null is not about the absence of gendered preferences, but about the institutional channeling of those preferences.

Several testable predictions follow from this interpretation. First, gender quota effects on budget composition should be stronger in countries or institutional settings with greater local fiscal autonomy—Nordic municipalities, Swiss cantons, or U.S. cities with strong-mayor systems. Second, effects should be more detectable in expenditure categories with greater discretionary content—culture and leisure spending, for example, may be more malleable than education or policing. Third, the effects may manifest not in budget allocations but in non-budgetary policy dimensions—regulatory decisions, procurement choices, service delivery quality, or intergovernmental grant applications—where councils have greater individual discretion.

These predictions provide a roadmap for future research that builds on the granular data infrastructure and identification strategy developed in this paper. The contribution is not merely a null result, but a null result with a specific institutional explanation that generates testable hypotheses about when and where gender representation does and does not affect policy.

8. Conclusion

Two decades of research on gender quotas and public finance have produced a puzzling result: mandating women's political representation reliably changes who governs but rarely changes what governments spend. This paper investigates whether the puzzle dissolves at a finer level of aggregation.

Using program-level budget data from over 8,000 Spanish municipalities and a multi-cutoff regression discontinuity design, I find that it does not—at least not in the modern institutional

environment. Over the full 2010–2023 period, the quota thresholds produce no detectable effects on within-education budget composition across eight subcategories. The first stage on female representation has attenuated substantially since the initial elections under the 2007 law, as voluntary compliance has rendered the threshold largely non-binding. The aggregate null on total education spending is confirmed with independent data.

But the story has a temporal dimension. Before Spain's 2013 austerity reform (LRSAL), when municipalities had broader fiscal discretion, the quota threshold did produce a significant compositional reallocation within education. This effect vanished after the LRSAL imposed fiscal constraints. The pattern suggests that institutional context—not measurement coarseness—is the binding explanation for why gender quotas appear to have no fiscal consequences in European settings. Political preferences can translate into policy only when institutions leave room for discretionary choice.

Three implications follow. First, the within-category null constrains theories of automatic preference transmission from politician identity to budget outcomes. In settings with strong party discipline and constrained budgets, descriptive representation may matter for symbolic and role-model effects without producing measurable fiscal consequences. Second, the temporal heterogeneity highlights the importance of institutional context: the same quota can produce spending effects in one regulatory environment and null results in another. Third, the decay of the first stage over time raises a methodological caution for RDD designs based on gender quota thresholds—the treatment contrast is not stable, and analyses using later elections may systematically underestimate effects that were present at adoption.

Whether the pre-LRSAL compositional shift improved educational outcomes remains an open question. But the lesson is already clear: representation is a necessary condition for policy change, but without fiscal discretion, it is not a sufficient one.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP). Budget data from the Ministerio de Hacienda (CONPREL database), election data from the Ministerio del Interior, and population data from the Instituto Nacional de Estadística (INE) are gratefully acknowledged.

Project Repository: https://github.com/SocialCatalystLab/ape-papers

Contributors: @SocialCatalystLab

First Contributor: https://github.com/SocialCatalystLab

References

Bagues, Manuel and Berta Esteve-Volart, "Gender and Lawmaking in Times of Quotas," *Journal of Public Economics*, 2022, *210*, 104665.

_ **and Pamela Campa**, "Can Gender Quotas in Candidate Lists Empower Women? Evidence from a Regression Discontinuity Design," *Journal of Public Economics*, 2021, *194*, 104315.

Baltrunaite, Audinga, Piera Bello, Alessandra Casarico, and Paola Profeta, "Gender Quotas and Quality of Politicians," *Journal of Public Economics*, 2014, *118*, 27–41.

Besley, Timothy, Olle Folke, Torsten Persson, and Johanna Rickne, "Gender Quotas and the Crisis of the Mediocre Man: Theory and Evidence from Sweden," *American Economic Review*, 2017, *107* (8), 2204–2242.

Bhalotra, Sonia and Irma Clots-Figueras, "Health and the Political Agency of Women," *American Economic Journal: Economic Policy*, 2014, *6* (2), 164–197.

Boletín Oficial del Estado, "Ley Orgánica 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres," *BOE*, 2007, (71). BOE-A-2007-6115.

_ , "Ley Orgánica 2/2011, de 28 de enero, por la que se modifica la Ley Orgánica 5/1985, del Régimen Electoral General," *BOE*, 2011, (25). BOE-A-2011-1639.

_ , "Ley 27/2013, de 27 de diciembre, de racionalización y sostenibilidad de la Administración Local," *BOE*, 2013, (312). BOE-A-2013-13756.

Brollo, Fernanda and Ugo Troiano, "Does Female Political Representation Reduce Gender-Based Violence? Evidence from Brazil," *American Economic Review*, 2016, *106* (5), 257–261.

Calonico, Sebastian, Matias D. Cattaneo, and Rocío Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

Casas-Arce, Pablo and Albert Saiz, "Do Quota Laws Reduce Gender Discrimination? Evidence from Spain," *Journal of Public Economics*, 2018, *157*, 25–36.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma, "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, 2020, *115* (531), 1449–1455.

Chattopadhyay, Raghabendra and Esther Duflo, "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India," *Econometrica*, 2004, *72* (5), 1409–1443.

Clots-Figueras, Irma, "Women in Politics: Evidence from the Indian States," *Journal of Public Economics*, 2012, *96* (7-8), 664–690.

Ferreira, Fernando and Joseph Gyourko, "Does Gender Affect Policy Priorities? Evidence from Randomly-Selected State Legislators," *Quarterly Journal of Economics*, 2014, *129* (1), 329–392.

Gagliarducci, Stefano and M. Daniele Paserman, "Gender Interactions within Hierarchies: Evidence from the Political Arena," *Review of Economic Studies*, 2012, *79* (3), 1021–1052.

Imbens, Guido W. and Thomas Lemieux, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 2008, *142* (2), 615–635.

Iyer, Lakshmi, Anandi Mani, Prachi Mishra, and Petia Topalova, "The Power of Political Voice: Women's Political Representation and Crime in India," *American Economic Journal: Applied Economics*, 2012, *4* (4), 165–193.

Lee, David S. and Thomas Lemieux, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 2010, *48* (2), 281–355.

Ministerio de Economía y Hacienda, "Orden EHA/3565/2008, de 3 de diciembre, por la que se aprueba la estructura de los presupuestos de las entidades locales," *BOE*, 2008, (297).

Washington, Erin L., "Female Socialization: How Daughters Affect Their Legislator Fathers," *American Economic Review*, 2008, *98* (1), 311–332.

A. Data Appendix

A.1 CONPREL Database Details

The CONPREL (*Consolidación de Presupuestos de Entidades Locales*) database is maintained by the Secretaría General de Financiación Autonómica y Local of the Ministerio de Hacienda. The database compiles executed budget data (*liquidaciones presupuestarias*) submitted annually by all Spanish local entities.

Data format. For each fiscal year, the data are distributed as a Microsoft Access 2007+ database (.accdb) within a ZIP archive. The relevant tables are:

- `tb_inventario`: Entity master file containing entity code (`codente`/`cdente`), entity name, and population.
- `tb_funcional`: Program-level spending records containing entity ID (`idente`), program group code (`cdfgr`), economic chapter code (`cdcta`), and spending amount (`importe`).

INE code extraction. Municipal entities are identified by a 5-digit code (2-digit province + 3-digit municipality) extracted from the first five characters of the entity code field. Non-municipal entities (provincial councils, water boards, etc.) are filtered out based on code length.

Program classification. The post-2008 program classification (Orden EHA/3565/2008) uses a hierarchical code:

- 1-digit: Policy area (e.g., 3 = "Production of Preferential Public Goods")
- 2-digit: Policy group (e.g., 32 = "Education")
- 3-digit: Specific program (e.g., 321 = "Education centers administration")

Variable construction. Within-education shares are computed as $s_{it}^k = e_{it}^k / E_{it}$, where e_{it}^k is spending in education program k and $E_{it} = \sum_k e_{it}^k$ is total education spending. Municipality-years with $E_{it} = 0$ are coded as missing (these represent municipalities that do not report education spending, typically because education is managed by a higher-level entity).

A.2 Election Data Details

Candidate-level data are distributed as ZIP archives containing fixed-width text files (DAT format). The file type 04 contains candidate records with the following column positions:

- Positions 10–11: Province code (2 digits)

- Positions 12–14: Municipality code (3 digits)
- Position 25: Candidate type (T = titular/principal, S = substitute)
- Positions 26–100: Name fields
- Position 101: Gender (M/F)
- Position 120: Elected status (S = yes, N = no)

The five-digit INE code is constructed by concatenating the province and municipality codes. Only principal candidates (type T) are included; substitutes are excluded. The female councillor share is computed as the number of elected female candidates divided by the total number of elected candidates in each municipality-election.

A.3 Padrón Municipal

The Padrón Municipal de Habitantes is Spain's continuous population register, managed by the INE. The official population figures (*cifras oficiales de población*) are published annually by Royal Decree and serve as the legal basis for determining municipal regulatory obligations, including the gender quota thresholds.

I use the average population across the available Padrón years within each municipality as the running variable, which reduces measurement error from year-to-year fluctuations in registration.

B. Identification Appendix

B.1 Additional Density Test Results

The Cattaneo et al. (2020) density test uses local polynomial density estimation to test for discontinuities in the distribution of the running variable. The null hypothesis is that the density is continuous at the cutoff. Failure to reject supports the assumption that municipalities did not systematically sort around the thresholds.

At both the 5,000 and 3,000 cutoffs, the test statistics are small and the p-values are well above conventional significance levels, indicating no evidence of manipulation.

B.2 Donut RDD Details

The donut RDD excludes municipalities within a specified radius of the cutoff. This addresses the concern that observations very near the threshold may be subject to precise manipulation of the running variable. I use two radii:

- **100 inhabitants:** Excludes municipalities with $|X_i - c| \leq 100$
- **200 inhabitants:** Excludes municipalities with $|X_i - c| \leq 200$

Results are robust across both radii. Larger donuts (e.g., 500 inhabitants) remove too many observations near the cutoff and produce imprecise estimates.

C. Robustness Appendix

C.1 Polynomial Order Sensitivity

The baseline specification uses a first-order polynomial ($p = 1$, local linear). I verify robustness to this choice by estimating with:

- $p = 1$: Local linear (baseline)
- $p = 2$: Local quadratic
- $p = 3$: Local cubic

Higher-order polynomials provide greater flexibility in approximating the conditional expectation function but introduce additional variance. The estimates are stable across polynomial orders: the general administration share (program 320) estimate is -0.006 (SE $= 0.011$) with $p = 1$, $+0.003$ (SE $= 0.014$) with $p = 2$, and $+0.000$ (SE $= 0.014$) with $p = 3$, all statistically insignificant.

C.2 LRSAL Heterogeneity Details

The sample is split at fiscal year 2014, reflecting the December 2013 enactment of the LRSAL:

- **Pre-LRSAL (2010–2013):** Municipalities had broader competences and greater discretionary spending authority.
- **Post-LRSAL (2014–2023):** Fiscal constraints and competence restrictions limit discretionary reallocation.

The pre-LRSAL estimates tend to be larger in magnitude, consistent with councils exercising greater discretion over education spending composition when not constrained by the fiscal framework. The post-LRSAL estimates remain in the same direction but are attenuated, suggesting that the fiscal reform reduced but did not eliminate the scope for gendered spending reallocation.