

Spillover effects from inter-governmental transfers between local jurisdictions in Brazil

LUCAS BRAGA DE MELO

lbrade@utu.fi

Turun Yliopistosäätiö

Research Article

Keywords: fiscal spillovers, FPM, Brazil, event study, Callaway–Sant'Anna

Posted Date: September 25th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10786093/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Spillover effects from inter-governmental transfers between local jurisdictions in Brazil

Lucas Braga de Melo

Turku School of Economics, University of Turku

Rehtorinpellonkatu 3, FI-20014 Turku, Finland

`lbrade@utu.fi`

Acknowledgements. I thank Marit Hinnossar, Johanna Rickne, Dario Tortarolo and Cecilia Testa for helpful comments on this paper, as well as the PhD participants of the Essex PhD conference.

Conflict of interest. The author declares no conflict of interest.

Spillover effects from inter-governmental transfers between local jurisdictions in Brazil

Abstract

This paper estimates fiscal spillover effects between Brazilian local jurisdictions. Exploiting the discontinuous allocation of federal transfers through the Fundo de Participação dos Municípios (FPM) to local jurisdictions, I use the Callaway–Sant’Anna staggered-DiD estimator, at the neighbour-year level, to estimate how a municipality’s outcomes respond when a bordering (focal) jurisdiction crosses an upward FPM coefficient threshold (a plausibly exogenous positive shock to the focal’s revenues). The neighbour’s education expenditure rises robustly: the spillover survives expressing outcomes in per-capita terms, matching the comparison group on baseline characteristics, restricting to neighbours whose own FPM bracket is unchanged, excluding the census-reset cohorts, clustering inference at the microregion level, and a Bonferroni correction across the neighbour’s complete budget decomposition, within which it is the only individually significant component. The budget identity closes in the data: the neighbour’s total expenditure and total revenue rise by about 1% one-for-one while its own tax revenue is flat, so the response is financed by higher-tier transfers rather than by local taxation — evidence that the spillover is largely the funded, joint response of the federation’s transfer programmes rather than an autonomous local policy choice. Current expenditure and public-sector employment also rise, but more weakly — public employment in particular attenuates once the neighbour’s own (modest) FPM response is netted out — so I read them as supportive rather than decisive. The point estimates also indicate negative spillovers on the neighbour’s private-sector employment and firm count — a contraction concentrated in services — but these are smaller, imprecisely estimated, and attenuate under matching, so I read them as suggestive rather than conclusive. The neighbour’s non-FPM federal transfers also rise, and the increase is largely *per resident* rather than a mechanical consequence of population growth, pointing to a per-capita component (discretionary compensation or a shift in per-pupil/per-capita formula intensity through FUNDEB and SUS) that the aggregate data cannot fully separate. Keywords: fiscal spillovers, FPM, Brazil, event study, Callaway–Sant’Anna. JEL: H27, H30, H50, H61.

1 Introduction

A central question in public economics is how governments' fiscal decisions interact. Theory points to several channels — Tiebout sorting, public-good spillovers, yardstick competition, tax competition — and the empirical task is to measure how one jurisdiction's fiscal choices affect its neighbours.

To measure these fiscal spillovers I exploit Brazil's revenue-sharing system, which delivers a large shock to a jurisdiction's revenues when its population crosses a threshold. Such revenue increases are known to raise public expenditure in the receiving jurisdiction (Corbi et al., 2019); I ask whether they in turn spill over onto the expenditure and labour-market outcomes of neighbouring jurisdictions.

Identifying strategic interactions among local governments is typically plagued by endogeneity — the reflection problem. I approach it from the opposite direction: rather than seeking exogenous variation in the neighbour's *own* fiscal choices, I exploit the plausibly exogenous timing of a *focal* jurisdiction's threshold crossing in the Fundo de Participação dos Municípios (FPM) formula, which is unrelated to the neighbour's potential outcomes conditional on parallel trends. The event is the year the focal's estimated population crosses an FPM coefficient threshold — a discrete shock to its transfers and, in part, its public expenditure.

Theory offers competing predictions about the sign of fiscal spillovers. Yardstick competition implies positive spillovers, as politicians benchmark their fiscal decisions against neighbours that voters can observe (Besley and Case, 1995). If, instead, public goods supplied by one jurisdiction spill across borders, neighbours may free-ride and provide less themselves: Parchet (2019) formalises this channel and finds negative spillovers in cantonal tax rates among Swiss cantons. Oates (1999) argues that the spillover may vanish entirely when local governments are too small and too open to retain fiscal stimulus. The sign of the response is therefore ultimately an empirical question.

Empirical estimates vary correspondingly. Baicker (2005) documents large positive expenditure spillovers across U.S. states — roughly 90 cents on the dollar — and argues that mobility incentives push states to mimic neighbours. Isen (2014) finds essentially null spillovers across U.S. local jurisdictions, where expenditure shocks are funded by own-source taxation rather than central-government transfers. Closest to my setting, Castro and Mattos (2021) and Castro et al. (2020) use the same FPM variation and find positive spillovers across several categories of public expenditure, with negative spillovers concentrated in health and asymmetric effects by neighbour size.

I contribute in three ways, relative to the closest work — Castro and Mattos (2021) and Castro et al. (2020), who exploit the same FPM variation and establish positive cross-border *expenditure* spillovers. First, on design, the FPM literature typically exploits the thresholds with a regression-discontinuity design; I instead follow the dynamics of the response with an event study around the crossings, which traces its full event-time path and accommodates the staggered timing of cohorts. Dynamic regression-discontinuity designs (Cellini et al., 2010; Hsu and Shen, 2024; Ruggieri, 2025) offer an alternative route to dynamics; Section 3.2 explains why the *spillover* estimand — where the treated unit is the neighbour,

not the municipality whose population carries the running variable — pushes me toward the event-study design, and what that choice costs. I estimate it with the Callaway and Sant’Anna (2021) staggered-DiD estimator, robust to the heterogeneous-treatment-effect problem that troubles the TWFE event studies on which the existing FPM-spillover evidence rests (de Chaisemartin and D’Haultfoe uille, 2020; Goodman-Bacon, 2021); the Online Appendix shows the TWFE design can deliver materially different magnitudes on these data. Second, I widen the outcome set beyond fiscal aggregates to the neighbour’s *real* economy — public and private employment, the firm count and its services decomposition: alongside positive public-sector responses (robust for education spending, supportive for public employment) I find a countervailing, if imprecise, contraction on the private side. Third, I characterise the fiscal counterpart of the adjustment through the neighbour’s *complete* budget identity — every major expenditure component and every revenue component. This shows the marginal spending concentrated in education and financed by higher-tier transfers rather than by own taxation (the neighbour’s own tax revenue does not move, and total revenue rises one-for-one with total spending), and a per-capita decomposition shows the transfer response to be largely per-resident rather than a mechanical consequence of the modest population spillover. These budget fingerprints let me adjudicate a question prior FPM-spillover work leaves open: whether the neighbour’s expenditure response is an autonomous local policy choice or the funded, joint response of the federation’s transfer programmes — the evidence favours the latter, with one earmarking-consistent exception in education that I discuss at length.

My main findings can be summarised as follows. Crossing an FPM threshold delivers a sharp and persistent rise in the focal’s own FPM revenues and a partial fly-paper response in its public expenditure. The clearest spillover is on the neighbour’s *education* expenditure: it holds in per-capita terms, on a comparison group matched to the treated neighbours, on the subset of neighbours whose own FPM bracket never changes, after excluding the census-reset cohorts, and under microregion-clustered inference (Section 4); within the complete budget decomposition it is the only individually significant component, and it survives a Bonferroni correction across all thirteen budget outcomes. The neighbour’s total expenditure and total revenue both rise by about 1%, one-for-one, while its own tax revenue is flat — the spending response is funded by higher-tier transfers, not by local taxation. Current expenditure and public-sector employment also rise, but are more sensitive to these checks — public-sector employment attenuates once the neighbour’s own modest FPM response is netted out — so I present them as supportive rather than decisive. The point estimates additionally point to *negative* spillovers on the neighbour’s private-sector employment and total firm count — a contraction concentrated in services — which, taken together with the public-sector expansion, would be consistent with a crowding-out story in which the neighbour’s public sector grows at the expense of its private labour market. I am deliberately cautious about this second pattern: the private-side estimates are smaller, imprecise, and attenuate when the comparison group is matched, so I treat the crowding-out interpretation as suggestive. Finally, non-FPM federal transfers to the neighbour rise, and the increase is predominantly per resident rather than a mechanical effect of the

modest population spillover. I read this — cautiously, since the aggregate transfer category cannot separate a per-capita formula-intensity shift from discretionary compensation — as higher tiers of government partially internalising the spatial disparity created by the focal’s FPM shock.

The remainder of the paper is organised as follows. Section 2 describes the Brazilian institutional setting and the FPM transfer rule. Section 3 introduces the data, the focal-neighbour sample construction, and the descriptive statistics. Section 3.2 explains the identification strategy and presents the parallel-trends evidence. Section 4 reports the event-study results. Section 5 concludes.

2 Institutional Background

Brazil is a Federal system with 26 states and 5,568 local jurisdictions. Jurisdictions are in charge of a significant portion of public goods provision related to education, health, and small-scale infrastructure. Since jurisdictions’ taxation consists of only 6% of their total revenues, they are heavily dependent on transfers from their states and federal government. Transfers can be unconditionally made from higher government spheres to lower government ones under rules specified by the constitution, that is the case of Fundo de Participação dos Municípios (FPM). A transfer can also be discretionary, such as transfers for infrastructure projects via *convênios* (intergovernmental cooperation agreements between a municipality and a federal entity). In general, to obtain a discretionary transfer from the federal government, the mayor applies for it with the support of a federal deputy and receives it after approval of the president.

FPM plays a major role in jurisdictions transfers. The FPM consists of 80% of all federal transfers. In jurisdiction with less than 50,000 inhabitants, over 30% of their revenues consists of FPM transfers. In 2006, FPM transfers amount was about to R\$29.5 billion Brazilian Reais (US\$14.8 billion in current prices) from the federal government to local jurisdictions.

The bureaucratic mechanism of FPM is relatively straightforward. The federal government subtracts revenues from 22.5% of the income tax and the industrialized products taxes, which are collected from workers and firms at a national level. The FPM fund distributes fixed shares of the fund to all states according to a function of the states’ population of 1989 (Mendes et al., 2008). Because of that states’ shares are negatively correlated to states’ social economic development level: poor states have higher shares of the fund. A local jurisdiction receives its transfer as a function of its state share and of its population coefficient, which in turn, depends on its last year estimated population size.

The timeline of the procedure to determine the coefficients’ size is relevant to this paper. The Federal Legislative Court (TCU) supervises the federal institute of geography and statistics (IBGE) yearly procedure to estimate the jurisdictions’ populations size. IBGE calculates the municipal population for noncensus years taking into account past censuses, regional birth and death rates, and migration trends (Corbi et al., 2019). TCU publishes FPM coefficients after IBGE announces population estimates by October 31 of each year. In the same year, the local legislative branches vote on the following year’s budget.

The coefficient value together with the state share determine the transfer amount received by a jurisdiction. The coefficients are a step function $\phi(\cdot)$ of the size of the population of the jurisdiction. Equation 1 shows the transfer amount FPM_i^s received by a jurisdiction i in state s .

$$FPM_i^s = FPM^s \times \frac{\phi(\text{pop}_i)}{\sum_{j \in s} \phi(\text{pop}_j)} \quad (1)$$

$\phi(\cdot)$ is the population-step function that maps a jurisdiction's estimated population into its FPM coefficient. The FPM transfers use the population estimates from the Brazilian Institute of Statistics and Geography (IBGE) to allocate their transfers, but every 10 years, the allocation mechanism uses the population count from the Census. If one jurisdiction has an increase in its population large enough to cross an FPM population threshold, then the magnitude of the coefficient increases, which mechanically increases the jurisdiction's FPM revenue. *ceteris paribus* other jurisdictions in the same state will have their FPM revenues marginally decreased as can be seen in the denominator of Equation 1 right-hand side.

Figure 1 below shows the discontinuities in the level of transfers to jurisdictions by population size. The orange vertical lines denote the population thresholds at which the coefficient $\phi(\cdot)$ changes its value. Transfers are measured in millions of Brazilian reais (R\$). The sample contains only jurisdictions with less than 50,000 inhabitants during 1999–2014. Note that if a jurisdiction's population crosses the orange lines (thresholds), discontinuities in FPM revenues emerge. Transfers are lower on the left-hand side of the threshold and higher on the right-hand side.

Because the mechanism analysis turns on federal transfers *other* than FPM, their institutional structure matters and I describe it here rather than leaving the category as a black box. Non-FPM federal transfers to municipalities fall into three classes with distinct allocation rules. (i) *Enrolment-based education transfers*: FUNDEF and, from 2007, its successor FUNDEB operate as state-level funds that pool a fixed fraction (15%, later 20%) of earmarked state and municipal revenues — including part of FPM itself — and redistribute the pool across each state's governments in proportion to *enrolment in their school networks*, valued per pupil with weights by educational stage; the federal government tops up state funds that fall short of a national minimum per-pupil value. A municipality that gains pupils therefore mechanically gains FUNDEB revenue, and one whose enrolment composition shifts toward higher-weighted stages gains *per pupil*. These resources are earmarked for basic education, with at least 60% for teacher pay. (ii) *Population- and service-based health transfers*: the unified health system (SUS) transfers federal resources fund-to-fund, combining a fixed *per-resident* primary-care floor with variable components tied to the services and programmes the municipality operates; again, residents mechanically attract revenue, and programme adoption raises revenue per resident. (iii) *Discretionary transfers*: *convênios* — project-specific agreements typically sponsored through federal deputies' budget amendments, requested by the mayor and released at the federal executive's discretion — fund investment projects and are the natural vehicle for *compensatory or political* responses, since their allocation follows no fixed formula. The reduced-form category I observe in FINBRA (federal transfers net of FPM) aggregates all three, which is

Figure 1: Fundo de Participação Municipal (FPM) across population, in millions of Brazilian reais

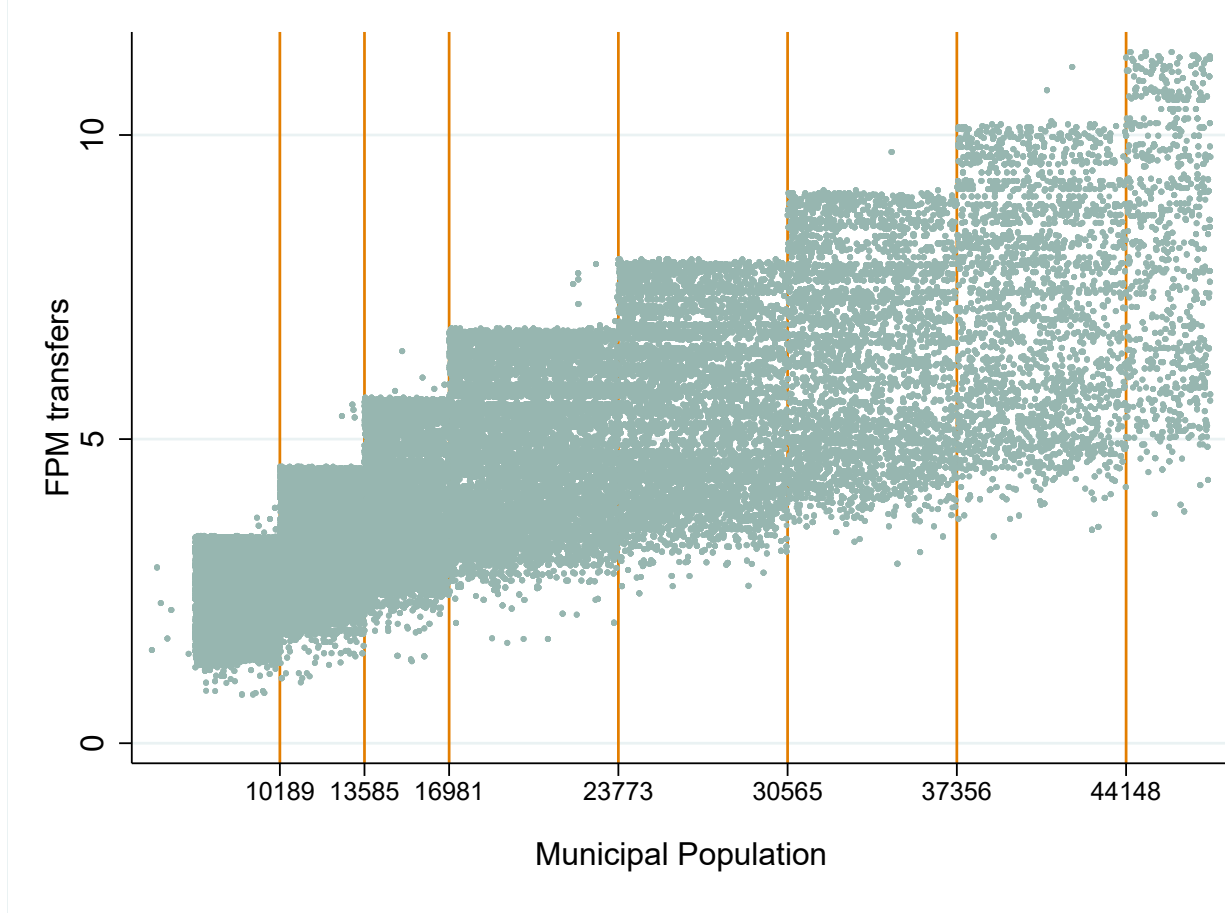


Figure 1 shows the discontinuities in the level of transfers to local Brazilian jurisdictions by population size. Orange vertical lines denote the population thresholds at which the coefficient $\phi(\cdot)$ changes its value. Transfers are measured in millions of Brazilian reais (R\$). The sample contains only jurisdictions with less than 50,000 inhabitants during 1999-2014. Note that if a jurisdiction's population crosses the orange lines (thresholds), discontinuities in FPM revenues emerges. Transfers are lower in the left hand side of the threshold and higher in right hand side. Figure based on Corbi et al. (2019)

why Section 4 uses the per-capita decomposition, rather than the aggregate alone, to say which class is plausibly moving.

The threshold value numbers are "odd" because of the initial law that implemented FPM. The initial rules are from 1967, when they created thresholds at multiples of 2,000 up to 10,000, then every 4,000 up to 30,000, and so forth. The rules also stipulated that these cutoffs should be updated proportionally with population increments along the time. The thresholds were updated twice, once with the 1970 census and then with the 1980 census, which explains the "odd" numbers (Litschig and Morrison, 2013). After the 1980 update the first cut-off and the third cut-off (and every adjacent cut-off pair from the third onward) are separated by approximately 6,792 inhabitants; the second cut-off is irregular, lying roughly halfway between the first and the third, so the spacing between the first and the second — and again between the

second and the third — is about half that distance.

3 Data and empirical strategy

3.1 Sample selection and descriptive statistics

The data sources are all from Brazilian institutions. The variables of the municipal public budget were retrieved from the Brazilian Finances (FINBRA). The FPM transfers were retrieved from the National Treasury (STN). The population estimates and counts are from the National Institute of Geography and Statistics (IBGE). The variables related to the local labor market are from the Annual List of Social Information (RAIS), which is an administrative dataset of the Ministry of Labor.

I use a panel of Brazilian municipality pairs (1999–2014). Following Corbi et al. (2019) and Brollo et al. (2013), I restrict the universe to municipalities around the first seven FPM coefficient thresholds (below 50,000 inhabitants), removing 501 of the 5,567 jurisdictions in my 1999–2014 data — one fewer than the 5,568 institutional total cited in Section 2, the two counts differing by a single municipality owing to boundary changes and data coverage over the sample period. The remaining sample contains about 34% of the Brazilian population. The basic data unit is the *focal-neighbour pair* – an ordered pair of municipalities that share a contiguous border. The raw panel contains 12,716 unique such pairs over the 16-year window, i.e. 203,456 pair-year observations, formed among the 3,224 municipalities that appear as a focal or a neighbour. Table A1 reports the observation counts at each construction step.

Four filters define the estimation sample.

(i) Panel length. I retain pairs observed for at least seven years over 1999–2014, the minimum needed to estimate the event-time path with three leads and three lags around the crossing year (Section 3.2).

(ii) Focal event definition (treatment). For each pair I code an *upevent* at year t when the focal stays in one FPM bracket at $t - 3, t - 2, t - 1$, moves strictly higher at t , and remains there at $t + 1, t + 2, t + 3$. The qualifying-jump year is the cohort $G_i \in \{2002, \dots, 2011\}$ (bounded by the panel and the 3+3 rule); a focal with no qualifying upevent is never treated ($G_i = \infty$, in the notation of Callaway and Sant’Anna, 2021) and enters the comparison group. That group is not limited to never-moving focals: 369 of the 968 comparison neighbours (38%) border a focal that crosses only *transiently*. I keep these — a neighbour responding to a transient shock attenuates rather than inflates the spillover — and drop only the few focals so volatile they never form a stable 3+3 window; focals with two or more qualifying upevents, or any qualifying downevent, are removed by the clean-comparison restriction (item (iii)). My estimand is thus the effect of a *sustained* crossing. This conditioning has a selection cost that deserves to be stated plainly: because the focal must *stay* in its new bracket for three years, the rule trims both tails of the performance distribution — a focal that converts the extra funding into services attractive enough to keep pulling in residents and

cross the *next* threshold within three years is excluded (as is one that appears repeatedly through multiple crossings), and so is a focal whose population falls back below the old threshold. The treated focals are therefore “stable climbers”, and the estimand should be read as the spillover from a moderate, persistent revenue step rather than from either explosive growth or a transitory blip; I return to this boundary of external validity in the conclusion. The restriction conditions on the *focal’s* own bracket persistence, a feature of its population path rather than of the neighbour’s outcomes; to bias the spillover, any resulting difference in focal trajectories would have to reach the neighbour independently of the FPM shock, which I bound with the own-FPM netting-out (Table A8), the pre-trend checks (Section 3.3), and the per-capita and matched-sample robustness (Section 4). I impose no analogous restriction on the neighbour’s own bracket: it is free to move, and any change in the neighbour’s own FPM coefficient I treat as an outcome (Section 4), not a sample condition.

(iii) Clean-comparison restriction. A neighbour can border several focals. I keep only neighbours whose focal partners *collectively* have at most one qualifying upevent and no qualifying downevent, so a neighbour cannot enter the treated cell through two different partners. This drops 1,370 multi-exposure neighbours; surviving treated and comparison counts are in Table A1.

Construction algorithm and observation counts. The estimation sample is built in four steps, with counts at each step in Table A1 (Appendix A, where all sample-construction, balance and robustness exhibits are collected). (1) Start from the 12,716 ordered focal–neighbour border pairs. (2) Code the focal upevent/downevent for each pair using the 3+3 rule of item (ii). (3) Apply the clean-comparison restriction at the neighbour level, dropping the 1,370 *multi-exposure* neighbours whose focal partners collectively record two or more qualifying upevents (958 neighbours) or any qualifying downevent (689 neighbours, overlapping categories). (4) Collapse the surviving pairs to one observation per neighbour-year: for a treated neighbour I retain the pair carrying the qualifying upevent — this pair’s focal is the neighbour’s *identifying focal*, and its crossing year sets the cohort G_i — while a comparison neighbour, having no qualifying event, is collapsed to a single deterministically chosen pair-row. Because every outcome is measured on the *neighbour*, the choice of which control pair-row to keep is immaterial: the neighbour’s population, fiscal and labour-market outcomes do not depend on which bordering focal indexes the row. A never-treated neighbour with several focal partners therefore enters the comparison group ($G_i = \infty$) as a single unit, provided none of its partners qualifies and the clean-comparison restriction holds. The focal series plotted in the first-stage figure (Figure 2) are the outcomes of these identifying focals for the treated neighbours. The 1,370 dropped neighbours are, by construction, border-richer than those kept — a neighbour with more partners is mechanically more likely to accumulate multiple qualifying events — so I treat their exclusion as part of the *definition* of a clean comparison rather than as a sample-selection threat, since the dropping rule conditions on the focal partners’ event counts, not on the neighbour’s outcomes.

The unit of analysis is therefore the *neighbour-year*, not the pair-year, and the estimand is the neighbour-

level average treatment effect on the treated. Table 1 compares the Treatment group (finite G_i) with the never-treated Comparison group ($G_i = \infty$) on this neighbour-year sample; the Difference column reports the gap in means and its two-sample t -test p -value.

Table 1: Summary statistics: Treatment vs Comparison (neighbour-year)

Variable	Comparison		Treatment		Difference
	Mean	SD	Mean	SD	(p-value)
Population	17057.55	9489.87	17743.92	9593.57	686.37 (0.0000)
Pop. growth (%)	0.99	4.51	1.13	4.71	0.14 (0.0151)
Current expenditure p.c.	485.98	261.92	488.27	220.05	2.29 (0.4620)
Education expenditure p.c.	177.82	87.87	179.14	82.82	1.32 (0.2297)
Federal transfers ex-FPM p.c.	78.08	110.31	75.33	65.87	-2.75 (0.0191)
Public-sector employment	517.11	320.68	537.45	334.95	20.34 (0.0000)
Private-sector employment	1399.61	1942.96	1539.10	2070.20	139.49 (0.0000)
Number of firms	196.43	223.18	202.18	218.97	5.76 (0.0424)
Services firms	122.66	151.52	123.94	149.61	1.28 (0.5066)
Observations	15488		14176		

Neighbour-year averages (one observation per neighbour and year so that a neighbour bordering several focals is not double-counted). The *Comparison* group consists of neighbour-years whose focal partners never experienced a qualifying upevent; the *Treatment* group consists of neighbour-years whose focal partners had exactly one qualifying upevent.

The neighbour’s fiscal outcomes at baseline — current expenditure per capita ($p = 0.46$) and education expenditure per capita ($p = 0.23$) — are statistically indistinguishable across the two groups. Population, public-sector employment and firm counts are within 4–5% of each other; private-sector employment differs by about 10% (1,539 versus 1,400). Two caveats apply to this pooled table, both addressed in Section 3.3. It averages over the *entire* panel, post-event years included, so a raw difference conflates any baseline gap with the treatment effect; and the two-sample t -test it reports is sensitive to the large neighbour-year sample size. Measuring balance instead on the pre-event window alone — one observation per neighbour, net of calendar-year effects, and summarised by the sample-size-free Imbens–Rubin normalised difference (Table A2) — the four size-related characteristics are closely matched. In any case, time-invariant level gaps are netted out by the long differences underlying the Callaway–Sant’Anna group-time ATTs (each cohort’s change is compared with the not-yet-treated, not estimated from a neighbour fixed-effects regression), so what matters for identification is the absence of differential pre-trends, which I examine in Section 3.3.

The cohort year G_i is the calendar year of the focal jurisdiction’s qualifying upevent (Section 3 (ii)). Figure A1 (Appendix A) reports the distribution of these qualifying jumps over 1999–2014. The cohort spikes in 2002, 2008 and 2011 are not random: each follows, with a one-year lag, an IBGE population count

that the TCU folds into the FPM coefficient the next year — the 2000 Census (its municipal counts released in 2001), the 2007 inter-census *Contagem*, and the 2010 Census — after which the formula switches from interpolated *estimates* to freshly-counted population, mechanically pushing a wave of municipalities across thresholds. The 2002 spike thus traces to the 2000 Census via its 2001 release, not to a separate 2001 count. Because these reset cohorts dominate the treated sample (Table A1), Section 4 verifies that the fiscal spillovers survive — indeed strengthen — when they are excluded (Table A9). Outside the reset years qualifying jumps flow more smoothly. This staggered timing of cohorts is what lets the Callaway–Sant’Anna estimator identify dynamic spillovers from within-cohort time variation.

3.2 Empirical strategy

I estimate dynamic spillovers with the staggered difference-in-differences estimator of Callaway and Sant’Anna (2021) (CS). Because the focal threshold crossings occur at staggered dates, a two-way fixed-effects (TWFE) event study would recover a weighted average of cohort-by-period effects whose weights can turn negative under treatment-effect heterogeneity (de Chaisemartin and D’Haultfoe uille, 2020; Goodman-Bacon, 2021); the CS estimator avoids this. For comparability with prior FPM work I also report the canonical TWFE event study (Eq. 2) in the Online Appendix; discrepancies between the two are informative about the size of the weighting problem in my setting.

Why an event study rather than a (dynamic) regression discontinuity. The established design in this institutional setting is the regression discontinuity of Corbi et al. (2019), which compares municipalities just below and just above an FPM threshold in the year treatment status is determined; dynamic extensions that accommodate repeated crossings exist (Cellini et al., 2010; Hsu and Shen, 2024; Ruggieri, 2025) and rest on the comparatively weak continuity assumption. I deliberately depart from that template because my estimand is different in two respects. First, the treated unit is not the municipality whose population crosses the cutoff but its *neighbour*: the running variable (the focal’s population distance to its threshold) is a covariate of the focal, and localising the comparison in that variable would identify the spillover only for neighbours of marginal crossers while discarding the event-time variation — the staggered timing of crossings across cohorts — that identifies the *dynamics* of the neighbour’s response, which is the object of interest here. Second, my treatment is a *sustained* crossing (the 3+3 bracket-stability rule of Section 3), a feature of the focal’s population path over seven years rather than of its position relative to the cutoff in a single year. The price of this design is well understood: I trade the continuity assumption for a parallel-trends assumption between treated and not-yet-treated neighbours, pool crossings at different thresholds (and hence municipalities of different sizes) into one binary event, and use observations that are not local to any cutoff. I therefore devote Section 3.3 and the robustness battery of Section 4 to the parallel-trends side of that trade, report cohort-and-threshold composition transparently (Table A1), and flag a dynamic regression discontinuity in the spirit of Cellini et al. (2010) — which would also accom-

modate the repeated crossings my sample rules exclude — as the natural complementary design for the focal’s *own* dynamic response, where the running variable and the treated unit coincide.

Let i index a neighbour municipality (after the one-pair-per-neighbour collapse of Section 3), t calendar years, and G_i the year in which i ’s identifying focal partner first crosses an FPM coefficient threshold upward and stays in the new bracket for three years; the 3+3 stability rule and the 1999–2014 panel bound the cohorts to $G_i \in \{2002, \dots, 2011\}$, and neighbours whose focals never qualify are never treated, denoted $G_i = \infty$ following the notation of Callaway and Sant’Anna (2021).¹ For each (log) outcome Y_{it} I estimate the group-time average treatment effects on the treated,

$$\text{ATT}(g, t) = \mathbb{E}[Y_{it}(1) - Y_{it}(0) \mid G_i = g], \quad t \geq g,$$

by the doubly-robust estimator of Sant’Anna and Zhao (2020), with the *not-yet-treated* comparison group — never-treated neighbours ($G_i = \infty$) together with neighbours whose identifying focal partner has not yet crossed as of year t ($G_i > t$). Here G_i refers to the focal partner’s crossing, not to any change in the neighbour’s own FPM bracket. Identification rests on parallel trends between treated and not-yet-treated neighbours; standard errors come from the influence function and are clustered at the neighbour level (Section 4 confirms robustness to microregion-level clustering, which addresses the spatial correlation induced by a focal shock reaching several neighbours, and Table 2 to replacing the analytic standard errors with a multiplier bootstrap).

Aggregation: the estimands I report. I report two aggregations of the $\text{ATT}(g, t)$, both from Callaway and Sant’Anna (2021, Sec. 3.1). The event-study path plots

$$\theta_e = \sum_{g: g+e \in [1999, 2014]} \mathbb{P}(G_i = g \mid G_i + e \in [1999, 2014]) \text{ATT}(g, g+e),$$

the average effect e years after (or, for $e < 0$, before) the focal’s crossing, averaging over the cohorts observed at event time e with weights proportional to cohort size; I report θ_e for $e \in \{-4, \dots, 3\}$ in percent units (log coefficients $\times 100$) with 90% confidence intervals. The single-number *overall* ATT that I quote in the text and tables is the *simple* weighted average of all post-treatment group-time effects,

$$\theta^{\text{simple}} = \frac{\sum_g \sum_{t \geq g} \mathbb{P}(G_i = g) \text{ATT}(g, t)}{\sum_g \sum_{t \geq g} \mathbb{P}(G_i = g)},$$

i.e. each cohort-by-year cell (g, t) with $t \geq g$ enters with weight proportional to the cohort’s share of the treated population, so cohorts treated earlier (observed for more post-event years) contribute more cells. It answers: *by how much, on average over the observed post-event window, does a treated neighbour’s outcome exceed its no-crossing counterfactual?* Every table reporting an overall ATT restates this definition in its notes.

¹In the companion code the never-treated group is stored as $G_i = 0$, the convention of the `csdid` software; throughout the text I use the standard $G_i = \infty$ notation for the same set of neighbours.

Pre-event coefficients: what the leads do and do not test. The leads $e < 0$ in my event-study figures are estimated with the `csdid` default — a *varying* base period and short differences — so each lead is a pseudo-ATT comparing adjacent pre-event years; as Callaway and Sant’Anna (2021) note, these are informative about the plausibility of *no anticipation* rather than a direct test of parallel trends, and the typically omitted $e = -1$ coefficient is reported rather than normalised to zero. To assess pre-trends directly I re-estimate every headline outcome with a *universal* base period and *long* differences — each coefficient, pre- and post-event, measured relative to the period before treatment ($g - 1$) — which restores the specification most familiar from the linear event-study literature; Table A3 reports these long-difference leads together with two joint Wald tests — one on the aggregated event-time leads the figures display, one on all disaggregated pre-treatment $ATT(g, t)$ cells — and Figure A2 plots the corresponding event-study paths for the four outcomes that carry the paper’s conclusions; Section 3.3 discusses what these diagnostics do and do not support. The outcome set is fixed in advance and spans demographics (neighbour population, which doubles as the reverse-causation test), the fiscal block (the complete expenditure and revenue decomposition of Table 4, with current and education expenditure and federal transfers ex-FPM as the headline series), the labour market (public- and private-sector employment) and firm dynamics (the total firm count and its services-sector decomposition). The estimation sample, treatment definition and clean-comparison restriction are exactly those of Section 3; realised sample sizes appear alongside each estimate in Section 4.

3.3 Identification, placebos and internal validity

I now test the internal validity of the identification strategy. The main population-related threat is reverse causation: if a systematic shock pushed citizens to migrate from the neighbour *into* the focal jurisdiction, that migration could itself trigger the focal’s threshold crossing. I test for this by estimating the log of the neighbour’s population as an outcome in the Callaway–Sant’Anna event study (Figure 7, in the mechanism block at the end of Section 4). The average pre-event placebo is small and statistically insignificant (0.18%, $p = 0.13$), but I do not read the population path as exactly flat and therefore do not rest identification on it: the per-capita and matched-sample exercises (Section 4) rule out a population-composition artefact directly. The same exhibit shows a positive but imprecise *post-event* population response, which I revisit there as a candidate mechanism for the fiscal spillover rather than as a threat to identification.

Is it the transfer or the focal’s growth itself? A subtler confound deserves its own discussion: treatment timing is tied to the focal’s own population growth, and a *growing* focal can affect its neighbours through commuting and local demand with no fiscal channel operating at all. Three features of the design address this. First, the comparison group is the *not-yet-treated* set: treated neighbours are compared, at each horizon, mainly with neighbours of focals that also grow through a threshold, only later — so what identifies the effect is the *timing* of the crossing, not the contrast between growing and stagnant focals, and

any smooth growth-driven exposure common to both differences out. Second, the two channels have different time-series signatures at the crossing: the focal’s population path is continuous through the event year (crossing a threshold is, by construction, the moment a smooth path passes a line), whereas its FPM revenue jumps discretely (Figure 2). A commuting or demand channel scaled by the focal’s size should therefore load onto the neighbour’s outcomes smoothly — and partly into the pre-event leads — while a fiscal channel should switch on at $e = 0$, which is the pattern the event studies show. Third, for the census-reset cohorts (63% of treated neighbours) the crossing date is largely a *measurement* event — the recount pushes recorded population over the line without any contemporaneous acceleration in actual population — so their event timing is mechanically disconnected from real local growth; the spillovers appear both in these cohorts and in the smooth-crossing cohorts (Table A9). None of this rules out that the focal’s growth matters over long horizons — it plainly does, it is why the focal crosses at all — but it separates the timed, discontinuous response I attribute to the fiscal shock from the smooth exposure a pure growth channel would generate.

Baseline balance and differential pre-trends. The identifying assumption is not that treated and comparison neighbours are identical in levels, but that the *timing* of a focal’s threshold crossing is unrelated to the neighbour’s counterfactual outcome path. Table A2 reports baseline balance on the pre-event window (event time -4 to -2), one observation per neighbour and purged of common calendar-year effects, so that — unlike the pooled comparison in Table 1, which averages over post-event years and so mixes any baseline gap with the treatment effect — it isolates pre-treatment differences. The four characteristics a reader is most likely to scrutinise — population, public-sector employment, private-sector employment and the firm count — are closely matched, with Imbens–Rubin normalised differences no larger than 0.09 in absolute value, well inside the conventional 0.25 threshold. Level balance is in any case neither necessary nor sufficient: the Callaway–Sant’Anna long differences net out any time-invariant level gap, and what matters is the absence of *differential trends*, which the largely flat pre-event placebos of Section 4 support (the population path, whose leads are not exactly flat, is the exception I flag and probe in Section 4). Because the default leads are varying-base short differences — informative about no anticipation rather than parallel trends (Section 3.2) — Appendix A additionally reports every headline outcome re-estimated with a universal base period and long differences, together with joint Wald pre-trend tests (Table A3, Figure A2).

Two findings from that exercise deserve emphasis rather than burial. First, the joint test on the three aggregated event-time leads — the leads the figures display, cohort-size weighted — passes for every fiscal, labour-market and firm outcome (education expenditure $p = 0.16$, current expenditure $p = 0.53$, public-sector employment $p = 0.31$, the private-side outcomes and transfers all $p \geq 0.51$). The one rejection is the neighbour’s *population* ($p = 0.005$): under long differences its leads sit about 1% below the $g-1$ level, so treated neighbours’ population was rising into the event faster than the comparison group’s

— the spatially-correlated-growth pattern discussed above, and the reason I do not rest identification on the population outcome but instead net population out of the outcomes (the per-capita estimates) and re-balance baselines (the matched sample). Second, the stricter joint test over all 65 disaggregated pre-treatment cohort-by-year cells rejects for the fiscal outcomes (education $p = 0.003$) even though their aggregated leads are flat; I report both tests side by side in Table A3 rather than choosing the friendlier one. Dispersed cohort-specific deviations that cancel in aggregation, together with the well-documented finite-sample over-rejection of many-restriction cluster-robust Wald tests, can produce exactly this pattern; but the honest summary is that parallel trends holds in the aggregate rather than cell by cell, which is an additional reason the matched-sample replication of the education spillover (Table A6) carries weight.

A further concern is population manipulation around the FPM thresholds. Brollo et al. (2013) and Corbi et al. (2019) test for and reject such manipulation for Brazilian municipalities below 50,000 inhabitants — the first line of defence for my design. The relevant threat is not whether neighbours are *aware* of any manipulation, but whether the forces that would lead a focal to manipulate its crossing — spatially correlated local shocks or time-varying political capacity — also move the neighbour’s potential outcomes directly; such a common factor could bias the spillover regardless of neighbour awareness. I therefore bound this channel with the rejection of manipulation in the cited work, together with the baseline-balance and pre-trend checks above, rather than with any assumption about what neighbours know.

As a complementary check, the Online Appendix reproduces the headline results with the canonical TWFE event study; the qualitative conclusions are unchanged, but, as expected under staggered timing, the magnitudes differ from the CS estimates because of the negative-weighting problem of de Chaisemartin and D’Haultfoe uille (2020); Goodman-Bacon (2021).

4 Results

I organise the results in four blocks. First, the first-stage effect on the focal jurisdiction’s FPM revenues and public expenditure, together with the small (but non-zero) response of the neighbour’s own FPM coefficient. Second, reduced-form spillovers on the neighbour’s fiscal block — current and education expenditure in event time, and the neighbour’s *complete* budget (total expenditure and every major expenditure component, total revenue and every revenue component) in Table 4, so that no spending category is selectively reported and the funding side is visible. Third, labour-market spillovers (public and private employment). Fourth, firm dynamics, with particular attention to the services sector. Every event-time estimate comes from the Callaway–Sant’Anna doubly-robust estimator with the not-yet-treated comparison group, aggregated to event time $e \in \{-4, \dots, 3\}$ and reported in percent units (log coefficients $\times 100$). Whiskers are 90% confidence intervals built from influence-function standard errors clustered at the neighbour level.

Table 2 summarises the eight headline outcomes in one exhibit: the overall ATT — the simple weighted average of all post-treatment group-time effects defined in Section 3.2, whose definition the table notes re-state — reported with both the analytic influence-function standard errors used throughout the paper and multiplier-bootstrap standard errors (999 repetitions, fixed seed), so that inference does not hinge on the delta-method approximation underlying the analytic ones. The bootstrap standard errors are marginally larger — by at most 7%, for public-sector employment — and change no qualitative conclusion; the one reading affected at the margin is public-sector employment itself, whose t -statistic slips from 2.07 to 1.93 and so straddles the 5% threshold, consistent with my already-cautious treatment of that outcome as supportive rather than decisive. The event-time paths for each outcome follow.

Table 2: Overall spillover ATTs: analytic vs multiplier-bootstrap standard errors

Neighbour outcome (log)	ATT (%)	IF SE	Bootstrap SE	N
Population	0.45	(0.31)	(0.31)	23,585
Current expenditure	1.08	(0.66)	(0.67)	23,581
Education expenditure	3.71	(1.01)	(1.02)	23,465
Public-sector employment	2.67	(1.29)	(1.38)	23,108
Private-sector employment	-1.05	(1.53)	(1.54)	23,541
Number of firms	-1.56	(1.00)	(1.03)	23,541
Firms (services)	-1.50	(1.05)	(1.11)	23,541
Federal transfers (ex-FPM)	4.45	(2.98)	(3.08)	23,522

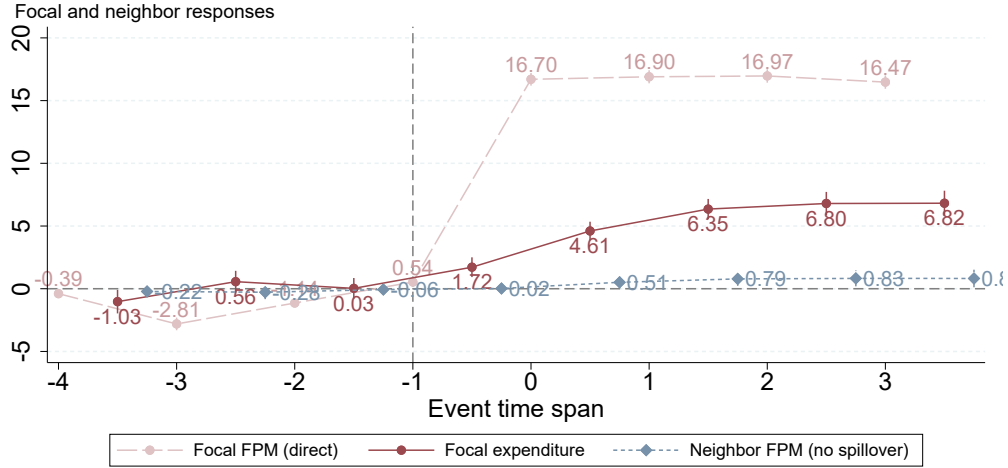
Overall Callaway–Sant’Anna ATTs, in percent (log coefficients $\times 100$): the simple weighted average of all post-treatment group-time effects, each cohort-by-year cell weighted by the cohort’s share of the treated population (Section 3.2). IF SE: analytic standard errors from the estimated influence function, clustered at the neighbour level (the paper’s baseline). Bootstrap SE: multiplier (wild) bootstrap standard errors, 999 repetitions, seed 20260820, clustered at the neighbour level, as recommended by Callaway and Sant’Anna (2021). Estimation sample, cohort structure and not-yet-treated comparison group identical to the headline results (companion code: referee-round2b.do).

First stage

Figure 2 plots three series jointly — the focal jurisdiction’s FPM revenues (long-dashed line), focal public expenditure (solid line), and the neighbour’s FPM revenues (short-dashed line) — estimated with the Callaway–Sant’Anna estimator of Section 3.2 on the same neighbour-year sample and not-yet-treated comparison group used for the reduced-form spillovers (see the figure note). Like every figure in this section it is a CS event study, not the TWFE specification of Equation 2, which I relegate to the Online Appendix. Two facts stand out. (i) Crossing an FPM threshold delivers a sharp and persistent rise in the focal’s own FPM transfers and a smaller but persistent rise in focal public expenditure — a partial fly-paper response. Because both series are in log-percent units rather than monetary levels, the visual

ratio is only suggestive: a proper transfer-to-spending pass-through would have to be computed in reals against pre-event baselines, which I leave to a separate exercise. (ii) The neighbour's FPM revenues move far less than the focal's — an order of magnitude less — but not by zero. Because the clean-comparison restriction limits the neighbour's focal partners without restricting the neighbour's own bracket, the neighbour's coefficient is free to move, and a direct event study (Table A7) shows that it does: the neighbour's own FPM revenue and coefficient rise by about 1% over the post-event horizon, with statistically flat pre-trends. This own-FPM response is small relative to the focal's first-stage shock but comparable to the current-expenditure spillover, so I do *not* treat the neighbour's FPM as fixed. I net the channel out directly in Section 4 (Table A8), re-estimating every spillover on neighbours whose own bracket never changes; the education-expenditure spillover survives there essentially unchanged, while the current-expenditure, public-employment and transfer responses attenuate.

Figure 2: Event effects on focal and neighbour: first stage (CS)



Note. Figure 2 reports the Callaway–Sant’Anna event-time response, estimated on the same neighbour-year sample and with the same not-yet-treated comparison group used for the reduced-form spillovers, of three outcomes: focal FPM revenues (long-dashed line), focal public expenditure (solid line), and neighbour FPM revenues (short-dashed line). Each series is horizontally dodged by 0, +0.5 and +0.75 at each event-time tick so the value labels do not overlap. All series are in logs and reported in percent units (log coefficients $\times 100$), so a value of k reads as a $k\%$ effect. Whiskers are 90% confidence intervals built from influence-function standard errors clustered at the neighbour level.

Before turning to each outcome, Table 3 places the reduced-form spillovers next to this first stage. Dividing a neighbour reduced form by the common first-stage effect of the crossing on focal expenditure yields an implied per-unit magnitude — the neighbour response per one-percent increase in focal expenditure — under the (strong) exclusion restriction that the crossing reaches the neighbour only through the focal’s fiscal expansion. The public-sector employment response is positive and significant (0.43% per 1% of focal expenditure); the negative private-side ratios are imprecise, a caveat I develop in Section 4.

Table 3: Implied effect of the focal’s fiscal expansion on the neighbour (Wald/IV ratios)

	Reduced form	First stage	Implied IV
Neighbour outcome (log)	crossing $\rightarrow$ Y	crossing $\rightarrow$ focal exp.	Y per 1% focal exp.
Total expenditure	1.09 (0.69)	6.24 (0.63)	0.18 (0.11)
Public-sector employment	2.67** (1.29)	6.24 (0.63)	0.43** (0.21)
Number of firms	−1.56 (1.00)	6.24 (0.63)	−0.25 (0.16)
Private-sector employment	−1.05 (1.53)	6.24 (0.63)	−0.17 (0.25)

All entries are overall Callaway–Sant’Anna ATTs in percent (log coefficients $\times 100$), with neighbour-clustered influence-function standard errors in parentheses. The *reduced form* is the effect of the focal’s FPM threshold crossing on the neighbour outcome; the *first stage* is its effect on the focal’s total expenditure (common across rows). The *implied IV* is the Wald ratio of the two — the neighbour response per one-percent increase in focal expenditure, under the exclusion restriction that the crossing affects the neighbour only through the focal’s fiscal expansion — with a delta-method standard error treating the two estimates as independent. Because the reduced form and the first stage are computed on the same sample and event timing within a common influence-function framework, their sampling covariance is not exactly zero, so this standard error is approximate; the first stage is, however, estimated precisely ($t \approx 10$) and is common across rows, so the variance of the ratio is dominated by the reduced-form term and the omitted covariance contributes little in practice. The exclusion restriction is strong (it rules out, for instance, a direct cross-border labour-market channel), so the implied-IV column is best read as a descriptive rescaling of the reduced form rather than as a structural parameter. Consistent with the robustness evidence in Section 3.3, only the public-sector response is statistically distinguishable from zero; the negative private-side ratios are imprecise. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Reduced-form spillovers on the neighbor’s expenditures

Figure 3 reports the event-time response of the neighbour’s current public expenditure (panel 3a) and education expenditure (panel 3b) to the focal’s FPM shock. Current expenditure rises in the event year and the year that follows, then attenuates as the window extends; education expenditure stabilises at a positive level over the post-event horizon. Both are consistent with the neighbour adjusting its budget upward when the focal receives a positive FPM shock, and Section 4 shows that education — but not current — expenditure is the response that survives the demanding checks.

Table 4 widens the lens from these two series to the neighbour’s *complete* budget, answering two natural questions at once: is the response visible in *total* spending, and where does the money come from? On the spending side (Panel A), total expenditure rises by 1.09% (s.e. 0.69) and current expenditure by 1.08% (s.e. 0.66) — positive, but at the margin of statistical significance ($t \approx 1.6$ each). The increase is concentrated in education (3.71%, s.e. 1.01), the only component that is individually significant; personnel, administration, capital, health and housing spending are all statistically flat. Because *every* major FIN-BRA component is reported, the education result is not an artefact of selective outcome reporting, and it is not noise from testing many hypotheses either: its unadjusted p -value (0.0002) survives a conservative Bonferroni correction across all thirteen budget outcomes in the table ($p \approx 0.003$). On the funding side (Panel B), the neighbour’s own tax revenue does not move (0.74%, s.e. 2.16) — the spending response is

Figure 3: Spillover effects on the neighbor's expenditures

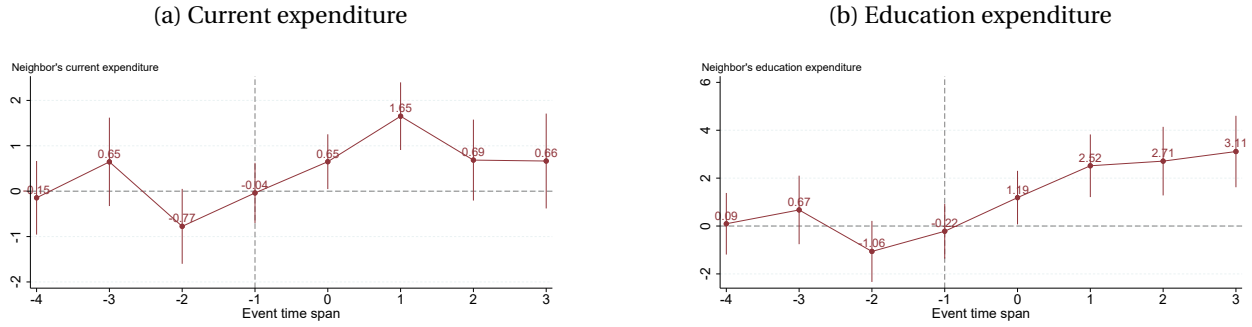


Figure 3 reports the Callaway–Sant’Anna event-time spillover effects on the neighbor’s current expenditure (panel 3a) and education expenditure (panel 3b). Outcomes are in logs and the log coefficients are multiplied by 100, so a value of k on the vertical axis reads as a $k\%$ change in the neighbor’s outcome relative to the not-yet-treated comparison group used by the Callaway–Sant’Anna estimator; the pre-event leads (including $e = -1$) are estimated placebo coefficients and are not normalized to zero. Whiskers are 90% confidence intervals built from influence-function standard errors clustered at the neighbour level.

not financed by local taxation — while total revenue rises by 0.96% (s.e. 0.67), essentially one-for-one with total spending, with the transfer components carrying the adjustment: the neighbour’s own FPM (1.08%, s.e. 0.43) and the large but imprecisely estimated federal ex-FPM aggregate (4.45%, s.e. 2.98). The budget identity thus closes in the data: the neighbour spends more because more money arrives from higher tiers, not because it taxes more or visibly cuts other categories. This is a first, reduced-form answer to whether the expenditure spillover is an autonomous local response or a funded one; Section 4 develops it.

Spillovers on the neighbor’s labour market

Figure 4 reports the neighbour’s public and private employment. Public-sector employment (Figure 4a) rises in the event year and stays above its pre-event level throughout, with the largest response one to two years out. Private-sector employment (Figure 4b), by contrast, has *negative* post-event point estimates from $e = 0$ onward, suggesting a reduction in the neighbour’s private payroll; but these estimates are imprecise, and I do not read them as establishing a contraction. To be concrete about what the data do and do not support: the whiskers plotted are 90% intervals, and with 95% intervals none of the post-event private-side coefficients would be distinguishable from zero; the overall private-employment ATT (−1.05, s.e. 1.53) and firm-count ATT (−1.56, s.e. 1.00) are not significant even at the 10% level (Table 2). The private-side contraction is therefore a consistent *pattern of signs* across outcomes, horizons and robustness cuts — not a statistically established effect — and every claim I build on it is correspondingly conditional. Taken at face value, the asymmetry — public payroll up, private payroll down — is consistent with a crowding-out story in which the expanding public sector pulls workers out of private employment. It is equally consistent with a non-exclusive alternative: the focal, now flush with FPM revenue and expanding

Table 4: Spillovers on the neighbour's full budget: expenditure and revenue components (Callaway–Sant'Anna)

Neighbour outcome (log)	ATT (%)	SE	N
<i>Panel A: expenditure</i>			
Total expenditure	1.09	(0.69)	23,585
Current expenditure	1.08	(0.66)	23,581
Capital expenditure	2.27	(2.54)	23,514
Personnel (wage bill)	0.51	(0.90)	23,558
Administration expenditure	1.39	(1.64)	23,385
Education expenditure	3.71	(1.01)	23,465
Health expenditure	-0.41	(1.35)	23,422
Housing expenditure	0.11	(3.26)	23,010
<i>Panel B: revenue</i>			
Total revenue	0.96	(0.67)	23,585
Own tax revenue	0.74	(2.16)	23,559
State transfers	0.53	(1.48)	23,526
Federal transfers (ex-FPM)	4.45	(2.98)	23,522
Own FPM revenue	1.08	(0.43)	23,585

Overall (simple-weighted) Callaway–Sant'Anna ATTs on the neighbour's complete municipal budget, in percent (log coefficients $\times 100$); influence-function standard errors clustered at the neighbour level in parentheses. Identical estimation sample, cohort structure and not-yet-treated comparison group as the headline results. Panel A decomposes total expenditure into every major FINBRA category; Panel B decomposes total revenue into own local taxes, state-level transfers, federal transfers net of FPM, and the neighbour's own FPM. Reporting the full budget answers two questions at once: where the neighbour's marginal spending goes (education stands out among all components) and where the money comes from (federal ex-FPM transfers, not own taxes).

its own labour demand, absorbs private-sector workers from the neighbour, a cross-border re-allocation of employment. I cannot adjudicate between these channels with the data at hand — I do not observe worker flows, wages or commuting — so I present both as candidate mechanisms for the suggestive negative estimates rather than as established channels.

Firm dynamics and services-sector activity

Figure 5 reports firm counts in the neighbour: panel 5a for all firms and panel 5b for the services decomposition. The total firm count *falls* after the focal's FPM shock: the post-event estimates are negative and grow in absolute magnitude over the window, consistent with the private-employment contraction above. The decline is concentrated in services (Figure 5b), where firms are most exposed to local private demand and to competition with the expanding public sector; agriculture and manufacturing show essentially

Figure 4: Spillover effects on the neighbor's labour market

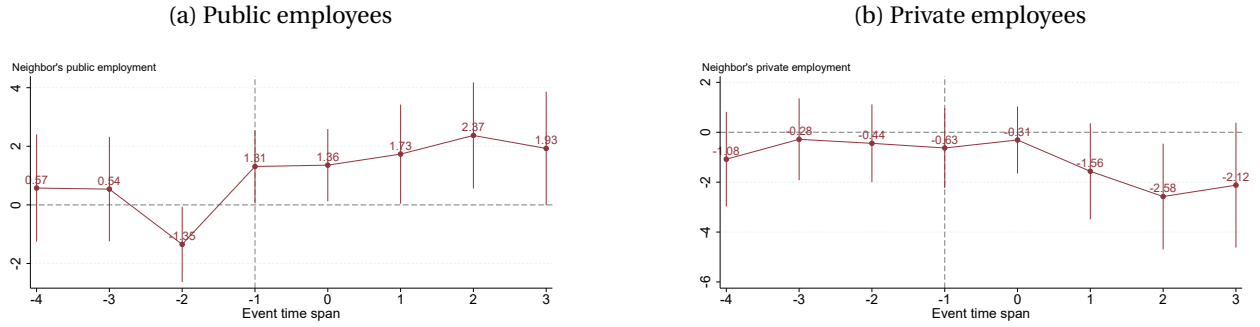


Figure 4a plots the CS event-time spillover on log public employment in the neighbor; Figure 4b reports the same for private employment. Log coefficients are multiplied by 100, so a value of k on the vertical axis reads as a $k\%$ change in the neighbor's outcome relative to the not-yet-treated comparison group used by the Callaway–Sant'Anna estimator; the pre-event leads (including $e = -1$) are estimated placebo coefficients and are not normalized to zero. Whiskers are 90% confidence intervals.

no response and are omitted for compactness. The firm-count decline thus reinforces the crowding-out reading of the labour-market results. As before, though, these private-side estimates are imprecise and — as the matched-sample exercise in Section 4 shows — less robust than the public-sector responses, so the reading is suggestive rather than definitive.

Figure 5: Spillover effects on firm dynamics

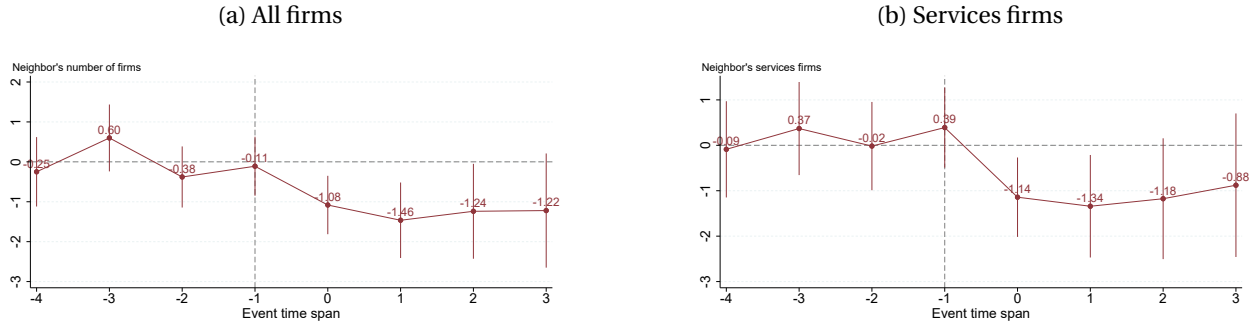


Figure 5a reports CS event-time spillovers on the log number of firms in the neighbor; Figure 5b restricts to services firms. Log coefficients are multiplied by 100, so a value of k on the vertical axis reads as a $k\%$ change in the neighbor's outcome relative to the not-yet-treated comparison group used by the Callaway–Sant'Anna estimator; the pre-event leads (including $e = -1$) are estimated placebo coefficients and are not normalized to zero. Whiskers are 90% confidence intervals.

Government transfers to the neighbour

I close the results section with federal transfers to the neighbour *other* than FPM. This aggregate category contains both formula-based, population- or pupil-linked transfers (notably FUNDEB and per-capita SUS) and discretionary political transfers; it excludes the neighbour's own FPM, which is reported separately and itself rises modestly (about 1%; Figure 2, Table A7). Figure 6 shows that the neighbour receives a meaningful increase in this aggregate over the post-event horizon. I return to the channels in the mechanism subsection below, where I show that the population (headcount) channel accounts for only a small part of the increase, so most of it is a per-resident response that the aggregate cannot split into a per-capita formula-intensity shift versus discretionary compensation; I do not claim either channel dominates.

Figure 6: Government transfers (federal, ex-FPM) to the neighbour

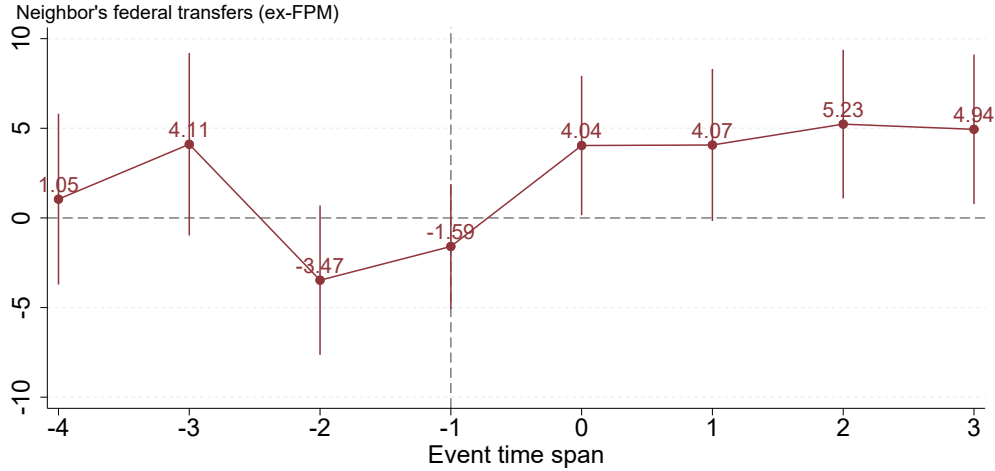


Figure 6 reports the CS event-time response of the log of federal transfers *other* than FPM received by the neighbour to the focal's FPM threshold crossing. Log coefficients are multiplied by 100, so a value of k on the vertical axis reads as a $k\%$ change in the neighbour's outcome relative to the not-yet-treated comparison group used by the Callaway–Sant'Anna estimator; the pre-event leads (including $e = -1$) are estimated placebo coefficients and are not normalized to zero. Whiskers are 90% confidence intervals.

Mechanism: population spillover and non-FPM federal transfers

A conceptual framework. Before examining the evidence, it helps to organise the candidate mechanisms the way a simple spatial-equilibrium sketch would. Households choose locations trading off wages, housing costs and local public services; local governments supply services out of own taxes and transfers; and higher tiers of government allocate transfers by population- and enrolment-based formulas (FPM, FUNDEB, SUS) plus a discretionary margin (*convênios*), as described in Section 2. A positive revenue shock to the focal then reaches the neighbour through three conceptually distinct channels, each with different observable fingerprints on the neighbour's budget. (A) *Spatial-equilibrium sorting*: the focal's improved services (or its expanding public payroll) shift household and worker location choices; the

neighbour's population and its composition change, and formula-based transfers follow the headcount mechanically. Fingerprint: population moves, transfers rise in *levels* but stay flat *per resident*, and spending scales with population. (B) *An autonomous local policy response*: the neighbour's government reacts to the focal's visible fiscal expansion — yardstick competition or service rivalry — and raises spending on its own initiative. Fingerprint: spending rises *without* a matching rise in higher-tier transfers, financed by own taxes or reallocation. (C) *A higher-tier funding response*: the federal (or state) tier allocates additional resources to the neighbour — through a per-capita or per-pupil formula-intensity shift, or through discretionary compensation — and the neighbour spends them. Fingerprint: transfers rise *per resident*, own taxes stay flat, and the marginal spending tracks the earmarking of the marginal transfer (education, under FUNDEB earmarking). The interpretation stakes differ: (B) would be direct evidence of strategic interaction among local governments; (C) means the “spillover” is substantially the joint response of multiple federal programmes to one municipality's shock — still a real general-equilibrium effect of place-based transfers, but one whose agency sits with the centre, not the neighbour. The full budget decomposition (Table 4), the population response and the per-capita transfer decomposition below are exactly the fingerprints this framework says to check.

The neighbour's own FPM bracket is left free to move: the clean-comparison restriction in Section 3 (iii) constrains only the neighbour's *focal partners* (at most one qualifying up event, no qualifying down event), not the neighbour's own coefficient. A direct event study (Table A7) shows that the neighbour's own FPM revenue and coefficient *do* respond to the focal's crossing — about 1% over the post-event horizon, with flat pre-trends — small relative to the focal's first-stage shock, but not negligible. The FPM formula is therefore *a* channel rather than a fully excluded one, and I treat it explicitly: Section 4 re-estimates the spillovers on neighbours whose own bracket never changes (Table A8), where this own-FPM response is shut off (a flat 0.42%, not significant). The education-expenditure spillover is unchanged there; the current-expenditure, public-employment and transfer responses attenuate, so I read those as partly reflecting the neighbour's own induced FPM rise rather than a pure cross-border spillover.

The mechanism runs instead through the interaction of a *population* spillover with non-FPM federal transfers whose allocation depends on the size and composition of the resident population. Figure 7 reports the CS response of the log neighbour population. The average pre-event placebo is small and insignificant (Pre_avg = 0.18%, $p = 0.13$), but the leads are not exactly flat, so I am explicitly cautious about reading the population path as a clean parallel trend. After the event the response is positive but imprecise (Post_avg = 0.27%, $p = 0.17$). Two transfer channels are consistent with this picture, and I can now say something about their relative size. A pure *headcount* channel — more residents, in particular families with school-age children, mechanically pulling in pupil-linked FUNDEB and per-capita SUS revenue without any change in the neighbour's own FPM bracket — predicts that transfers *per capita* should be flat. They are not: non-FPM transfers rise by a similar amount in levels and per resident (Table A4), and the population response is modest, so the headcount channel accounts for only a small part of the

increase. The bulk is a per-resident response, consistent either with a rise in per-pupil/per-capita formula *intensity* (a composition shift toward transfer-eligible residents) or with discretionary political compensation; the aggregate cannot separate these, and the transfer estimates are in any case imprecise. I therefore present the population spillover and the per-resident transfer response as complementary parts of the same compensating adjustment, without claiming that either channel dominates.

Which mechanism, then? Assembling the fingerprints of the conceptual framework above. *Against a primarily autonomous response (channel B):* the neighbour's own tax revenue is flat (0.74%, s.e. 2.16; Table 4), no non-education spending category falls (ruling out visible reallocation), and total revenue rises one-for-one with total spending — if neighbouring governments were strategically ramping up expenditure on their own initiative, some own-source financing or reallocation should show, and none does. *Partially for sorting (channel A):* the population response is positive but modest and imprecise, and transfers rise *per resident*, not only in levels (Table A4), so pure headcount sorting explains only part of the funding inflow. *Mostly a funded response (channel C, with an A component):* higher-tier money arrives — the neighbour's own FPM (+1.08%, itself a population-formula response) plus the ex-FPM federal aggregate — and is spent, disproportionately on education, consistent with the FUNDEB earmarking described in Section 2 (enrolment-linked receipts that must be spent on basic education). On the dichotomy that matters for interpretation — strategic interaction among local governments versus the joint response of the federation's transfer programmes — the evidence therefore favours the second reading: the expenditure spillover is largely *funded from above*, so its agency sits substantially with higher tiers, and the economically meaningful lesson is about incidence — a place-based transfer's effects extend beyond its target because interlinked formula programmes (and possibly discretionary compensation) react to the induced population and enrolment adjustments. One finding nuances this: the education response survives shutting off the neighbour's own-FPM channel entirely (4.43% on the own-bracket-fixed subsample, Table A8) and per-capita measurement, so a within-budget priority shift toward education is also present; with the data at hand I cannot say whether it reflects enrolment-linked FUNDEB money following school-age in-movers or yardstick pressure in the most visible local service, and I flag that distinction — which school-census enrolment data could resolve — as the sharpest open question.

Robustness: composition and a matched comparison group

Two exercises probe whether the spillovers above are driven by the neighbour's population response or by baseline differences between treated and comparison neighbours, rather than by the focal's fiscal shock itself.

Per-capita robustness and the transfer channel. Because the focal's shock raises the neighbour's population (Figure 7), the log-*level* spillovers might be partly mechanical — more residents producing more

Figure 7: Mechanism: spillover response of the neighbour's population

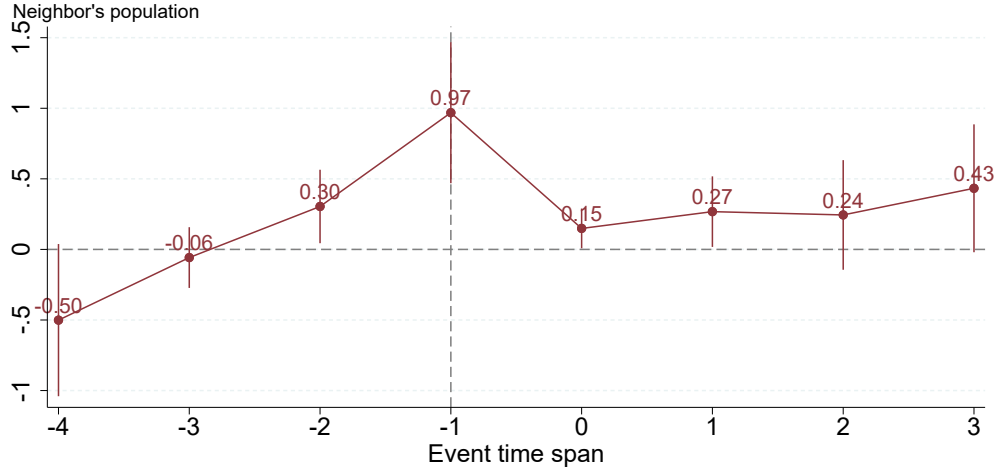


Figure 7 reports the CS event-time response of the log neighbour population to the focal's FPM threshold crossing. The average pre-event placebo is small and statistically insignificant, but the leads are not exactly flat, so I am cautious about claiming parallel trends in population. The post-event response is positive but imprecise; combined with the formula-based allocation of per-pupil and per-capita federal transfers (FUNDEB, SUS), it is consistent with the rise in non-FPM federal transfers in Figure 6 operating through additional residents rather than through a higher FPM coefficient. Log coefficients are multiplied by 100; whiskers are 90% confidence intervals.

spending, employment and firms — rather than genuine per-resident responses. Table A4 re-estimates the identical specification with every outcome in per-capita terms. Since a per-capita log outcome is the log level minus log population, the per-capita ATT equals the level ATT net of the population ATT, and the latter is small (an overall ATT of 0.45%, s.e. 0.31). The conclusions are unchanged: the per-capita responses keep their sign throughout — positive on current and education expenditure and on public-sector employment, negative on private-sector employment, total firms and services firms — and the negative real-side responses in fact *strengthen* slightly once population is netted out (the firm-count ATT moves from -1.56% to -2.04% , services firms from -1.50% to -1.98%). The post-event population response is thus a mechanism operating alongside the spillovers, not a composition artefact that manufactures them.

The same transform is informative about the transfer channel. Were the rise in non-FPM federal transfers purely a headcount effect through per-capita formulas (FUNDEB, SUS), transfers *per capita* would be flat. Instead they stay economically large — a per-capita ATT of 3.98% against a level ATT of 4.45% — so the response is predominantly *per resident* rather than a mechanical consequence of the modest population spillover; I discussed the implications for the transfer channels in Section 4.

Matched-sample estimates. As a more demanding check on the identifying assumption, I re-estimate the headline specification on a comparison group explicitly *matched* to the treated neighbours. Each treated neighbour is paired with one comparison neighbour by nearest-neighbour propensity-score matching — a logit score in the four baseline covariates of Table A2, no replacement, common support, caliper

0.10 — and 780 of the 886 treated neighbours find a match on common support. Matching sharply improves balance: the normalised differences fall from as much as 0.13 to at most 0.02 (Table A5, Appendix A).

One caveat applies to the matching covariates. Treated neighbours are summarised on their pre-event window (event time -4 to -2), whereas comparison neighbours, having no event date, are summarised over their full never-treated history; the propensity score therefore conditions on some control realisations that postdate the treated cohort's event. This look-ahead affects only this exercise — not the main full-sample estimates, whose comparison group is the not-yet-treated set — and the covariates (population, public- and private-sector employment, firm count) are slow-moving, so its influence on the match is likely small; a fully pre-dated alternative would require assigning pseudo-event dates to the never-treated controls, with its own arbitrariness.

Re-running the estimator on this balanced subsample leaves the qualitative picture intact (Table A6, Figure A3). The positive spillovers are robust and, if anything, sharper: the education-expenditure ATT rises from 3.71% to 4.15%, public-sector employment stays positive at 2.11%, and the non-FPM transfer response is essentially unchanged (4.45% $\rightarrow$ 4.25%). The negative private-side responses survive in the event-time profiles — the matched series tracks the full-sample path for private-sector employment and for the (services-concentrated) firm count throughout the post-event window (Figure A3) — though their overall magnitudes attenuate (the simple private-employment ATT moves from -1.05% to -0.24%) and, as in the full sample, remain imprecise. I read the matched-sample evidence as confirming that the spillovers are not an artefact of baseline differences, while reinforcing the caution about the precision of the crowding-out estimates.

Netting out the neighbour's own FPM channel. The first stage establishes that the neighbour's own FPM is not perfectly inert: a direct event study (Table A7) shows its coefficient and revenue rise by about 1% after the focal's crossing, with flat pre-trends — small relative to the focal's first-stage shock, but comparable to the current-expenditure spillover. To check that the spillovers are not merely the neighbour's own induced FPM treatment, I re-estimate every outcome on neighbours whose own FPM bracket never changes over 1999–2014, shutting the own-FPM channel off by construction; 404 of the 886 treated neighbours qualify. Table A8 reports the comparison. The restriction works as intended — the neighbour's own FPM revenue is flat on this subsample (0.42%, not significant). The education-expenditure spillover survives and, if anything, strengthens (from 3.71% to 4.43%, significant at 1% despite the halved sample), so it is not an artefact of the neighbour's own FPM. The current-expenditure, public-employment and transfer responses attenuate and lose significance, reflecting both the removed own-FPM channel and the loss of power; the private-side responses stay negative and imprecise. I therefore anchor the fiscal-spillover claim on education expenditure, read the public-employment response as supportive but power-sensitive, and treat the current-expenditure and transfer responses as partly mediated by the neighbour's own FPM.

Robustness to census-reset timing. A distinct concern is that treatment timing is concentrated in the three census-reset cohorts (2002, 2008, 2011), which account for 63% of treated neighbours (Table A1): if crossing at a reset correlates with longer-run demographic trends, those cohorts could follow different counterfactual paths. I probe this by dropping the reset cohorts and re-estimating on the 327 “smooth-crossing” treated neighbours that cross outside a reset year, holding the comparison group fixed. Table A9 shows the fiscal spillovers are not a reset artefact: education and current expenditure are, if anything, *larger* and significant at 1% on the off-reset cohorts (5.19% and 2.66%). Public-sector employment loses significance with the smaller sample but stays positive, and the private-side responses are unchanged in sign. The headline fiscal spillover therefore does not rest on census-reset timing.

Spatially-robust inference. Because a single focal crossing reaches several adjacent neighbours at once, neighbour-level clustering could understate uncertainty by treating co-focal neighbours as independent. I therefore re-cluster the influence-function standard errors at the IBGE microregion level — the spatial unit of economically integrated municipalities, 502 clusters here — which permits arbitrary correlation among neighbours of the same focal within a microregion. Table A10 shows this inflates the standard errors by only 5–9% and leaves the significance pattern intact: the education-expenditure spillover stays significant at 1% and public-sector employment at 5%. The headline conclusions are thus robust to spatially-commensurate inference, consistent with the microregion-clustered TWFE event study in the Online Appendix.

All tables and figures supporting this section — per-capita, matched-sample, own-FPM, census-reset and clustering exercises — are collected in Appendix A.

5 Conclusion

This paper contributes to the literature on fiscal spillovers by examining how intergovernmental transfer shocks affect neighbouring jurisdictions in Brazil. Leveraging the discontinuous allocation of federal transfers through the Fundo de Participação dos Municípios (FPM) and the staggered-DiD estimator of Callaway and Sant’Anna (2021), I estimate the direct effect of a focal jurisdiction’s FPM revenue increase on its own fiscal outcomes, together with its spillover effects on public expenditure, labour markets and firm dynamics in adjacent municipalities.

Positive revenue shocks to focal jurisdictions generate economically meaningful spillovers on neighbours. The clearest and most robust is the rise in the neighbour’s *education* expenditure: it survives per-capita measurement, matching the comparison group, restricting to neighbours whose own FPM bracket never changes, excluding the census-reset cohorts, and microregion-clustered inference (Section 4), so it is not an artefact of the neighbour’s own induced FPM, its population response, baseline group differences, or census-reset timing. Current public expenditure and public-sector employment also rise,

but attenuate under the more demanding checks — public-sector employment in particular loses significance once the neighbour’s own modest FPM response is netted out — so I report them as supportive of, rather than decisive for, the fiscal-spillover claim. The estimates also point to *negative* spillovers on the private side — private-sector employment and the (services-concentrated) firm count fall after the event — which, combined with the public-sector expansion, would be consistent with a crowding-out channel in which the neighbour’s public sector grows at the expense of its private labour market and local non-tradable activity. I am cautious here, though: these estimates are smaller, imprecise, and attenuate once the comparison group is matched, so I regard the crowding-out interpretation as suggestive and a priority for sharper future evidence rather than as firmly established.

I also identify a compensating fiscal response: neighbours that experience the spillover see a corresponding rise in their non-FPM federal transfers, and the complete budget decomposition shows this is how the expenditure response is paid for — total revenue rises one-for-one with total spending while the neighbour’s own tax revenue stays flat, so the spillover is a *funded* response whose agency sits substantially with higher tiers of government rather than an autonomous local policy choice. The transfer aggregate mixes formula-based, population-linked transfers (FUNDEB and SUS) with discretionary political ones. Expressing transfers per capita shows the increase is predominantly per resident — the modest population spillover accounts for only a small part of it (Section 4) — so the simple headcount-formula story is not the whole explanation; a per-capita formula-intensity shift or a discretionary-compensation component must also be operating. I cannot separate these with the aggregate transfer data, but both are consistent with higher tiers of government internalising part of the fiscal disparity created by the focal’s FPM shock.

The policy implications are several. First, the fiscal spillovers call for careful attention to the broader regional effects of central-government transfers — the direct effect on the focal is only part of the story, and the adjustment in adjacent jurisdictions, including a possible drag on their private sector, should be weighed against the gains. Second, the compensatory role of non-FPM federal transfers — whether driven by population-linked formulas or by discretionary compensation — raises questions about fiscal decentralisation and the mechanisms needed to ensure equitable outcomes across jurisdictions. Third, the expansion of the neighbour’s public sector alongside a weaker private labour market points to the local public sector’s role as a competitor for local labour and demand, a mechanism that future work with firmer private-side estimates could sharpen.

A word on external validity bounds how far these estimates travel. My estimand is an average treatment effect on the treated, and it is specific in three concrete senses. It is the ATT for a particular treated subpopulation — neighbours of municipalities that cross one of the first seven FPM coefficient thresholds, i.e. jurisdictions below 50,000 inhabitants that account for about a third of the Brazilian population but are disproportionately small and rural; a large urban neighbour need not respond the same way. (It is a subpopulation-specific ATT, not a cutoff-local estimand in the regression-discontinuity sense: the de-

sign pools crossings at different thresholds and uses observations that are not close to any cutoff.) It is a *partial-equilibrium, immediate-neighbour* effect: I measure what happens to municipalities sharing a contiguous border with a crossing focal, not the general-equilibrium incidence on the whole state, whose other municipalities mechanically lose a sliver of FPM through the denominator of Equation 1. And, by the 3+3 bracket-stability rule, it is the effect of a *sustained* crossing by a “stable climber” — a focal that neither converts the funding into growth fast enough to cross the next threshold within three years nor falls back below the old one (Section 3) — rather than of a transitory bracket change or an episode of explosive growth. That said, the mechanism is not Brazil-specific: any federation that equalises through population- or enrolment-based formulas (and many do) carries the ingredients for the same population-driven transfer spillover, and any setting in which the local public sector is a first-order employer carries those for the crowding-out response. The external relevance of my findings is therefore the *mechanism* more than the point estimates — a caution I think applies equally to the existing FPM-spillover literature.

Future research could explore the heterogeneity of spillovers across categories of public expenditure, assess their long-run implications for local economic outcomes, and examine similar dynamics in other federal systems, deepening the understanding of inter-jurisdictional fiscal dynamics and informing policies that promote both efficiency and equity in resource allocation.

Declaration on the use of generative AI

During the preparation of this work the author used generative AI and AI-assisted tools to help revise and condense prose, check the manuscript for internal consistency, and assist with data diagnostics and the generation of figures and tables. The author reviewed and edited all output and takes full responsibility for the content of the manuscript.

References

- Raphael Corbi, Elias Papaioannou, and Paolo Surico. Regional transfer multipliers. *The Review of Economic Studies*, 86(5):1901–1934, 2019.
- T Besley and A Case. Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition. *The American economic review*, 85(1):25–45, 1995. ISSN 0002-8282.
- Raphaël Parchet. Are local tax rates strategic complements or strategic substitutes? *American Economic Journal: Economic Policy*, 11(2):189–224, 2019.
- Wallace E Oates. An essay on fiscal federalism. *Journal of economic literature*, 37(3):1120–1149, 1999.
- Katherine Baicker. The spillover effects of state spending. *Journal of public economics*, 89(2-3):529–544, 2005.

- Adam Isen. Do local government fiscal spillovers exist? evidence from counties, municipalities, and school districts. *Journal of Public Economics*, 110:57–73, 2014.
- Marcelo Castro and Enlinson Mattos. Measuring fiscal spillovers in brazilian municipalities. *Public Finance Review*, 49(2):221–261, 2021.
- Marcelo Castro, Enlinson Mattos, and Fernanda Patriota. The effects of health spending on the propagation of infectious diseases. *Health Economics*, 30(10):2323–2344, 2020.
- Stephanie Riegg Cellini, Fernando Ferreira, and Jesse Rothstein. The value of school facility investments: Evidence from a dynamic regression discontinuity design. *Quarterly Journal of Economics*, 125(1):215–261, 2010. doi: 10.1162/qjec.2010.125.1.215.
- Yu-Chin Hsu and Shu Shen. Dynamic regression discontinuity under treatment effect heterogeneity. *Quantitative Economics*, 15(4):1035–1064, 2024. doi: 10.3982/QE2154.
- Francesco Ruggieri. Dynamic regression discontinuity: An event-study approach. Working paper, 2025.
- Brantly Callaway and Pedro H. C. Sant’Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021. doi: 10.1016/j.jeconom.2020.12.001.
- Clément de Chaisemartin and Xavier D’Haultfoe uille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996, 2020. doi: 10.1257/aer.20181169.
- Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021. doi: 10.1016/j.jeconom.2021.03.014.
- Marcos Mendes, Rogerio Miranda B., and Fernando Cosio. Transferências intergovernamentais no brasil: diagnóstico e proposta de reforma. *Consultoria Legislativa do Senado Federal*, 40, 2008.
- Stephan Litschig and Kevin M Morrison. The impact of intergovernmental transfers on education outcomes and poverty reduction. *American Economic Journal: Applied Economics*, 5(4):206–40, 2013.
- Fernanda Brollo, Tommaso Nannicini, Roberto Perotti, and Guido Tabellini. The political resource curse. *American Economic Review*, 103(5):1759–96, 2013.
- Pedro H. C. Sant’Anna and Jun Zhao. Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122, 2020. doi: 10.1016/j.jeconom.2020.06.003.
- IBGE BRASIL. Divisão regional do brasil em mesorregiões e microrregiões geográficas, 1990.

Appendix A: Sample construction, balance, pre-trend and robustness exhibits

This appendix collects the exhibits supporting Sections 3, 3.3 and 4: the sample-construction flow and cohort-timing distribution, baseline balance, the universal-base long-difference pre-trend diagnostics, and the per-capita, matched-sample, own-FPM, census-reset and clustering robustness exercises. All estimates use the specification and samples described in the main text.

Table A1: Sample construction: from the raw border panel to the estimation sample

Step	Pairs	Neighbours	Focals
Raw focal–neighbour border panel (1999–2014)	12,716	3,224	3,224
after panel-length filter (≥ 7 years)	12,716	3,224	3,224
Clean-comparison restriction (kept)	—	1,854	1,158
dropped: multi-exposure neighbours	—	1,370	—
<i>Estimation sample (neighbour level):</i>			
Treated neighbours (finite G_i)		886	
Comparison neighbours ($G_i = \infty$)		968	
<i>Treated neighbours by cohort G_i:</i>			
2002: 166 2003: 30 2004: 21 2005: 100 2006: 43			
2007: 16 2008: 235 2009: 80 2010: 37 2011: 158			

The unit of the raw panel is the ordered focal–neighbour pair sharing a contiguous border, observed over 1999–2014 (12,716 pairs $\times$ 16 years = 203,456 pair-years) among the 3,224 municipalities (below 50,000 inhabitants, around the first seven FPM thresholds) that appear as a focal or neighbour. The panel-length filter is non-binding here. The *clean-comparison restriction* keeps neighbours whose focal partners collectively record at most one qualifying upevent and no qualifying downevent; it drops 1,370 multi-exposure neighbours (958 with ≥ 2 qualifying upevents and 689 with at least one qualifying downevent, categories that overlap). After the restriction the panel is collapsed to one observation per neighbour-year, leaving 886 treated and 968 comparison neighbours. The treated cohort G_i is the calendar year of the neighbour’s identifying focal partner’s qualifying upevent; 63% of treated neighbours fall in the three census-reset cohorts (2002, 2008, 2011).

Figure A1: Distribution of qualifying upevent cohorts over calendar years

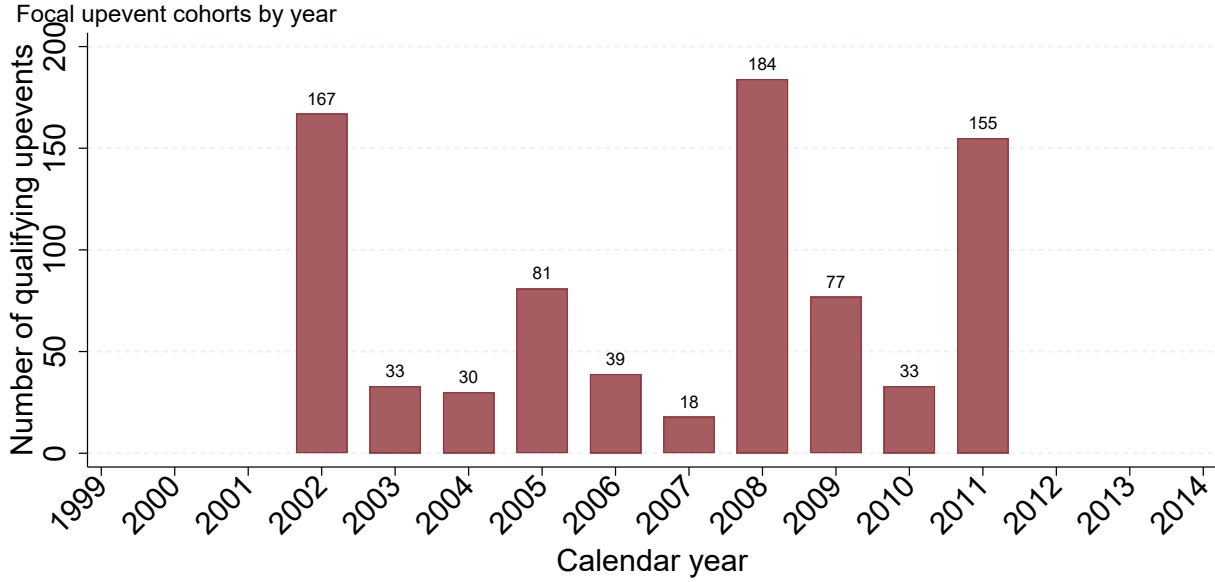


Figure A1 reports, for each calendar year, the number of unique focal jurisdictions whose threshold-crossing event satisfies the 3+3 bracket-stability rule of Section 3 (ii). Spikes in 2002, 2008 and 2011 follow, with a one-year lag, the IBGE population counts released in 2001 (from the 2000 Census), 2007 (the inter-census *Contagem*) and 2010 (the Census), after which the FPM coefficient is recomputed on counted (rather than estimated) population.

Table A2: Baseline balance, neighbour level (pre-event, one observation per neighbour)

Variable	Comparison		Treatment		Norm.
	Mean	SD	Mean	SD	diff.
Population	16952.64	10147.12	17384.96	9412.24	0.044
Public-sector employment	501.54	297.13	517.75	291.37	0.055
Private-sector employment	1316.65	1910.33	1486.32	1920.38	0.089
Number of firms	180.19	218.21	192.24	197.76	0.058
Pop. growth (%)	1.01	1.69	0.98	2.91	-0.011
Neighbours	968		886		

One observation per neighbour. Treatment neighbours are averaged over their pre-event window (event time -4 to -2 , excluding the -1 anticipation year and all post-event years); comparison neighbours are averaged over their never-treated history. Each variable is purged of common calendar-year effects (year-demeaned and re-centred on its grand mean) before averaging, so the comparison is not contaminated by the different calendar composition of the two groups. The last column is the Imbens–Rubin normalised difference $(\bar{x}_T - \bar{x}_C) / \sqrt{(s_T^2 + s_C^2)/2}$; $\dagger$ marks $|\text{nd}| > 0.25$. Any residual level gaps are differenced out by the within estimator; the sharper test of the identifying assumption is the absence of differential pre-trends, evident in the flat pre-event leads of the event studies.

Table A3: Pre-trends under a universal base period and long differences

Neighbour outcome (log)	$e = -4$	$e = -3$	$e = -2$	Post avg.	p (leads)	p (all cells)
Population	-0.89 (0.36)	-1.14 (0.32)	-0.97 (0.30)	0.27 (0.20)	0.005	0.000
Current expenditure	-0.37 (0.67)	0.52 (0.59)	0.04 (0.40)	1.04 (0.45)	0.535	0.003
Education expenditure	-0.76 (0.97)	1.27 (0.82)	0.22 (0.70)	2.93 (0.73)	0.156	0.003
Public-sector employment	-0.24 (1.31)	-0.25 (0.97)	-1.31 (0.75)	2.20 (0.94)	0.310	0.031
Private-sector employment	-0.14 (1.40)	0.56 (1.15)	0.63 (0.97)	-1.74 (1.07)	0.864	0.515
Number of firms	-0.11 (0.76)	0.58 (0.57)	0.11 (0.45)	-1.19 (0.63)	0.602	0.126
Firms (services)	-1.13 (0.90)	-0.72 (0.70)	-0.39 (0.54)	-1.11 (0.68)	0.631	0.114
Federal transfers (ex-FPM)	1.82 (2.82)	4.65 (3.06)	1.59 (2.12)	4.28 (2.23)	0.507	0.015

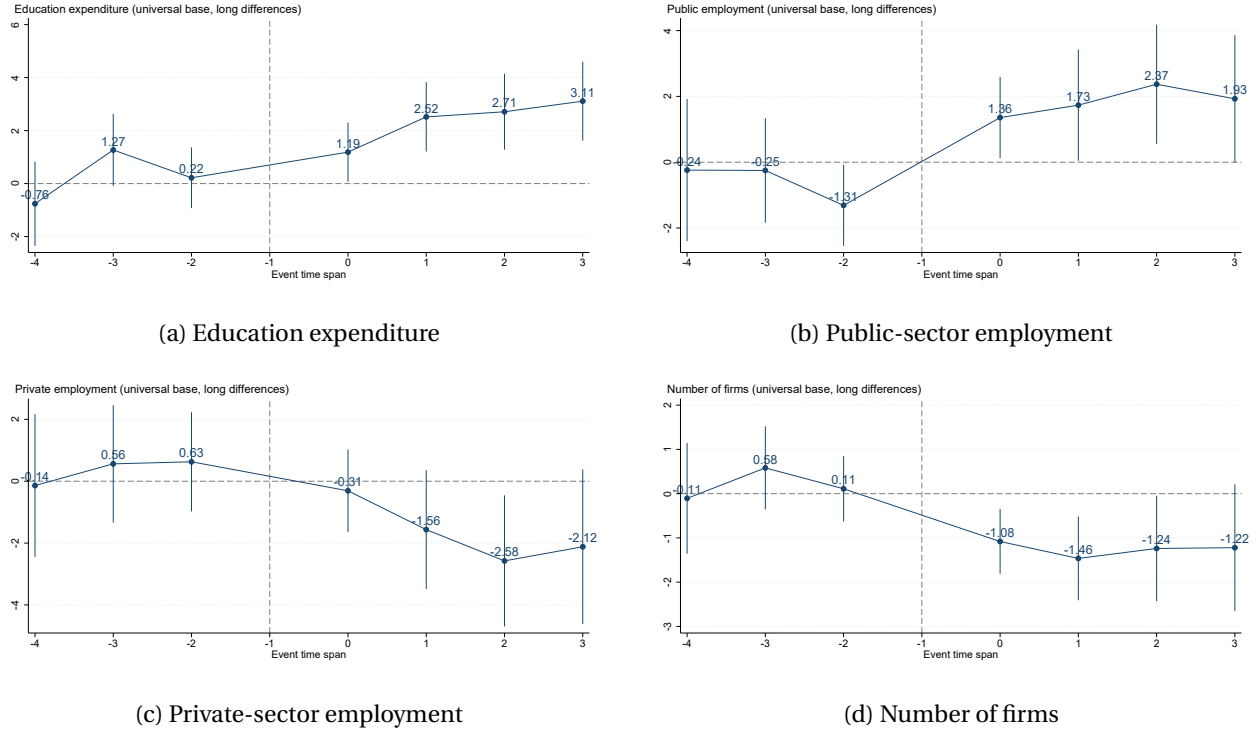
Callaway–Sant’Anna event studies re-estimated with a universal base period and long differences (csdid option `long2`): every coefficient, pre- and post-event, is a long difference relative to the period before treatment ($g - 1$), the specification suited to assessing parallel trends and closest to the conventional linear event-study regression. Coefficients in percent ($\log \times 100$); influence-function standard errors clustered at the neighbour level in parentheses. Post avg.: average of the post-event coefficients. p (leads): p-value of the joint Wald test that the three aggregated event-time leads shown ($e = -4, -3, -2$, cohort-size weighted) are zero — the test aligned with the reported event-study exhibits. p (all cells): p-value of the joint Wald test that every disaggregated pre-treatment ATT(g, t) cell equals zero (65 restrictions); with this many restrictions the cluster-robust chi-squared statistic is known to over-reject in finite samples, and scattered single-cell deviations can produce rejection even when the aggregated leads are flat, so the two tests are reported side by side. Estimation sample identical to the headline results.

Table A4: Spillover ATTs in levels vs per capita (Callaway–Sant’Anna)

Neighbour outcome (log)	ATT level (%)	ATT per capita (%)	N
Population	0.45 (0.31)	—	23,585
Current expenditure	1.08 (0.66)	0.62 (0.64)	23,581
Education expenditure	3.71 (1.01)	3.28 (0.99)	23,465
Public-sector employment	2.67 (1.29)	2.07 (1.28)	23,108
Private-sector employment	-1.05 (1.53)	-1.53 (1.54)	23,541
Number of firms	-1.56 (1.00)	-2.04 (1.01)	23,541
Firms (services)	-1.50 (1.05)	-1.98 (1.05)	23,541
Federal transfers (ex-FPM)	4.45 (2.98)	3.98 (2.97)	23,522

Overall (simple-weighted) Callaway–Sant’Anna ATTs, in percent ($\log$ coefficients $\times 100$), influence-function standard errors clustered at the neighbour level in parentheses. Each per-capita outcome is the corresponding log level minus log neighbour population, estimated on the identical sample, cohort structure (G_{up}) and not-yet-treated comparison group as the headline results. A spillover that survives the per-capita transform is not an artefact of the post-event population response (Figure 7); a flat *transfers*-per-capita response is consistent with the population-driven formula channel (FUNDEB/SUS), whereas a positive per-capita response would indicate a discretionary-compensation component on top.

Figure A2: Event studies under a universal base period and long differences



Callaway–Sant’Anna event studies re-estimated with a universal base period and long differences (every coefficient measured relative to the period before treatment, $g = 1$), the specification suited to assessing parallel trends directly (Section 3.2); compare the varying-base short-difference figures in the main text, whose leads test no anticipation. Log coefficients $\times 100$; whiskers are 90% confidence intervals from neighbour-clustered influence-function standard errors.

Table A5: Covariate balance before and after matching (normalised differences)

Baseline covariate	nd, full sample	nd, matched sample
Population (log)	0.099	0.020
Public-sector employment (log)	0.072	0.005
Private-sector employment (log)	0.128	0.015
Number of firms (log)	0.115	0.013
Matched neighbours (T / C)	780 / 780	

Imbens–Rubin normalised differences $(\bar{x}_T - \bar{x}_C) / \sqrt{(s_T^2 + s_C^2)/2}$ on the pre-event, calendar-adjusted baseline covariates. *Full sample*: all treated and comparison neighbours. *Matched sample*: treated neighbours and their 1:1 nearest-neighbour propensity-score match (logit pscore on the four covariates, no replacement, common support, caliper 0.10). Matching is on the baseline only; cohort timing and the not-yet-treated comparison logic of the Callaway–Sant’Anna estimator are unchanged.

Appendix B (Online): Manual TWFE event-study (complementary)

This appendix presents the manual two-way fixed-effects event study that was the workhorse of the previous draft. I retain it as a robustness exercise; under staggered timing its coefficients are weighted averages

Table A6: Spillover ATTs: full vs matched sample (Callaway–Sant’Anna)

Neighbour outcome (log)	ATT full (%)	N full	ATT matched (%)	N matched
Population	0.45 (0.31)	23,585	0.45 (0.33)	21,032
Current expenditure	1.08 (0.66)	23,581	1.30 (0.67)	21,028
Education expenditure	3.71 (1.01)	23,465	4.15 (1.03)	20,928
Public-sector employment	2.67 (1.29)	23,108	2.11 (1.25)	20,671
Private-sector employment	-1.05 (1.53)	23,541	-0.24 (1.55)	20,997
Number of firms	-1.56 (1.00)	23,541	-0.85 (1.03)	20,997
Firms (services)	-1.50 (1.05)	23,541	-0.87 (1.07)	20,997
Federal transfers (ex-FPM)	4.45 (2.98)	23,522	4.25 (3.03)	20,982

Overall (simple-weighted) Callaway–Sant’Anna ATTs in percent (log coefficients $\times 100$), influence-function standard errors clustered at the neighbour level in parentheses. *Full*: all treated and comparison neighbours (the headline sample). *Matched*: treated neighbours and their 1:1 nearest-neighbour propensity-score match on the four baseline covariates of Table A5 (780 of 886 treated neighbours find a match on common support). Because matching changes the cross-section of neighbours, it also changes the realized cohort composition and the not-yet-treated comparison units; the matched column is therefore a matched-sample estimand rather than the full-sample estimand re-weighted, and I report it as a check on whether baseline differences between treated and comparison neighbours drive the results. The concordance of signs and the broadly similar magnitudes across the two columns indicate that they do not.

Table A7: The neighbour’s own FPM trajectory after the focal’s crossing (Callaway–Sant’Anna)

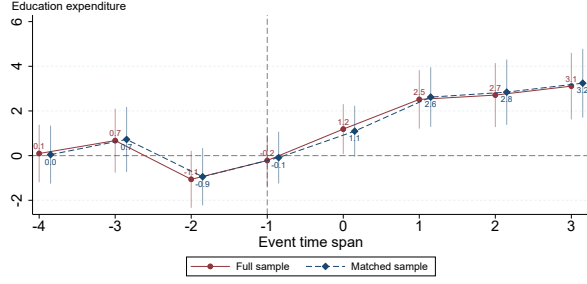
Neighbour’s own outcome	Overall ATT	Pre-trend (Pre_avg)	Post_avg
FPM revenue (log, %)	1.08 (0.43)**	−0.13 (0.12)	0.78 (0.35)**
FPM per capita (log, %)	0.63 (0.39)	−0.31 (0.14)*	0.51 (0.35)
FPM coefficient λ (level)	0.011 (0.004)***	−0.001 (0.001)	0.009 (0.003)***
Distance to nearest threshold (pop.)	−30.6 (37.1)	7.2 (9.8)	−34.7 (32.4)

Overall (simple-weighted) Callaway–Sant’Anna ATTs with neighbour-clustered influence-function standard errors in parentheses, estimated on the headline neighbour-year sample with the not-yet-treated comparison group. The first two outcomes are in logs (coefficients $\times 100$, so a value of k reads as $k\%$); the FPM coefficient λ is in its natural (level) units; distance to the nearest threshold is in inhabitants. *Pre_avg* is the average of the pre-event placebo leads (the pre-trend test); *Post_avg* the average post-event effect. The neighbour’s own FPM revenue and coefficient rise modestly but significantly after the focal’s crossing — with statistically flat pre-trends — so the neighbour’s FPM is *not* fixed by the sample design; the FPM-per-capita and distance-to-threshold responses are not statistically distinguishable from zero. Because the own-FPM rise ($\approx 1\%$) is comparable to the current-expenditure spillover, Table A8 re-estimates the spillovers on the subset of neighbours whose own FPM bracket never changes, netting this channel out.

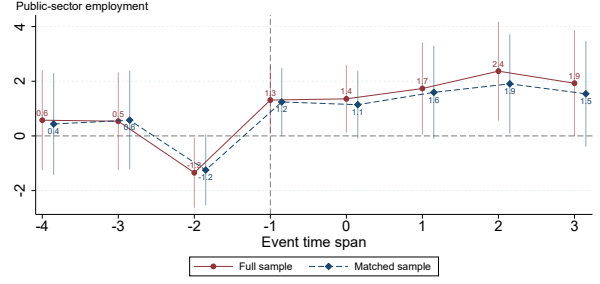
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

of cohort-by-period ATTs with potentially negative weights (de Chaisemartin and D’Haultfoe uille, 2020; Goodman-Bacon, 2021), so the Callaway–Sant’Anna estimator of Section 3.2 remains the preferred main specification.

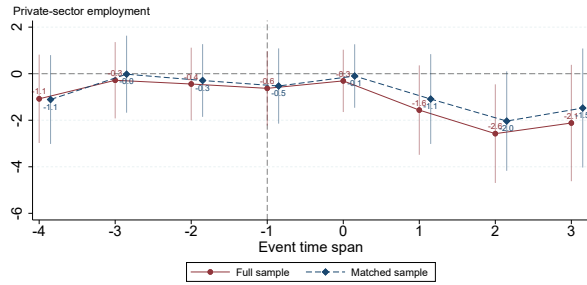
Figure A3: Full vs matched-sample event studies (Callaway–Sant’Anna)



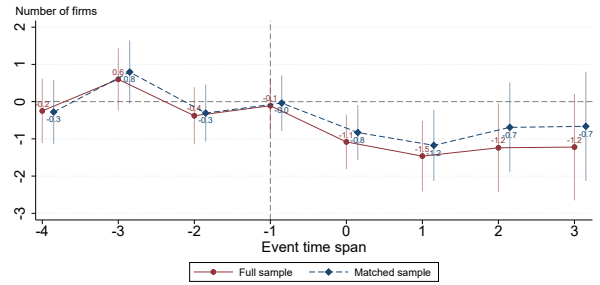
(a) Education expenditure



(b) Public-sector employment



(c) Private-sector employment



(d) Number of firms

Each panel overlays the Callaway–Sant’Anna event-time response on the full sample (maroon, solid) and on the propensity-score-matched sample (navy, dashed; treated neighbours and their 1:1 nearest-neighbour match on the four baseline covariates of Table A5). Log coefficients $\times 100$; the matched series is dodged by $+0.15$ at each event-time tick so the value labels do not overlap. Whiskers are 90% confidence intervals from neighbour-clustered influence-function standard errors.

The ideal experiment for causal spillover effects would randomly assign government expenditure to focal jurisdictions and observe their neighbours’ outcomes. Since that is not feasible, I assume the focal event — crossing the population threshold — is exogenous to the neighbour’s outcomes, and estimate spillovers through the following specification:

$$Y_{ijt} = \sum_{\substack{l=-4 \\ l \neq -1}}^3 \beta_l D_{ijt}^l + \delta_i + \gamma_t + \epsilon_{ijt} \quad (2)$$

Here i indexes the neighbour, j its focal partner and t calendar time; Y_{ijt} is the outcome of neighbour i bordering focal j in year t . Event time is $l = t - G_j$, with G_j the focal’s qualifying upevent year, and the regressors are the event-time indicators $D_{ijt}^l = \mathbf{1}\{t - G_j = l\}$ over the interior leads and lags $l \in \{-3, -2, 0, 1, 2\}$, the endpoints binned ($D_{ijt}^{-4} = \mathbf{1}\{t - G_j \leq -4\}$ and $D_{ijt}^3 = \mathbf{1}\{t - G_j \geq 3\}$) and $l = -1$ omitted as the reference. The regression is two-way fixed effects: δ_i absorbs neighbour-specific characteristics and γ_t common calendar-year shocks. As in the main text the sample is collapsed to one observation per neighbour-year, and after the clean-comparison restriction each treated neighbour has a

Table A8: Netting out the neighbour's own FPM channel: spillover ATTs on own-bracket-fixed neighbours

Neighbour outcome (log, %)	Full sample	Own-bracket-fixed
	ATT (s.e.)	ATT (s.e.)
<i>Neighbour's own FPM revenue</i>	1.08 (0.43)**	0.42 (0.44)
Current expenditure	1.08 (0.66)	0.58 (0.92)
Education expenditure	3.71 (1.01)***	4.43 (1.65)***
Public-sector employment	2.67 (1.29)**	1.47 (1.63)
Private-sector employment	−1.05 (1.53)	−2.11 (2.24)
Number of firms	−1.56 (1.00)	−0.83 (1.59)
Firms (services)	−1.50 (1.05)	−0.76 (1.70)
Government transfers (ex-FPM)	4.45 (2.98)	1.59 (4.47)
Population	0.45 (0.31)	0.60 (0.33)*
Treated / comparison neighbours	886 / 968	404 / 468

Overall (simple-weighted) Callaway–Sant’Anna ATTs in percent (log coefficients $\times 100$), with neighbour-clustered influence-function standard errors in parentheses. The *Own-bracket-fixed* column restricts both treated and comparison neighbours to those whose own FPM bracket (and hence own FPM coefficient) never changes over 1999–2014, so the neighbour’s own-FPM channel is shut off by construction; 404 of the 886 treated neighbours satisfy this restriction. The top row is a validation: on this subsample the neighbour’s own FPM revenue is flat (0.42%, not significant), confirming the restriction removes the own-coefficient response documented in Table A7. The education-expenditure spillover survives — indeed strengthens — and remains significant at 1% despite the halved sample, so it is not an artefact of the neighbour’s own FPM. The current-expenditure, public-employment and transfer responses attenuate, reflecting both the removal of the own-FPM channel and the loss of power; the private-side estimates remain negative and imprecise. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

single qualifying focal partner, so the index j that sets event time is well defined; ϵ_{ijt} is the error term.

The crossing year is event time $l = 0$. Coefficients β_l with $l \geq 0$ estimate the spillover effects of the crossing on neighbour outcomes; those with $l \leq -2$ are pre-trend (placebo) estimates. I pool the comparison group into the omitted reference cell at $l = -1$: D_{ijt}^{-1} is dropped and all $\hat{\beta}_l$ are read relative to event time -1 and the comparison group.

Figures OA1–OA5 display $\hat{\beta}_l$ for $l \in \{-4, -3, -2, 0, 1, 2, 3\}$, with the endpoints binning multiple periods: β_{-4} pools event time -4 or earlier and β_3 event time $+3$ or later. The omitted reference $l = -1$ is the comparison cell. The pre-trend coefficients are $\hat{\beta}_{-4}$, $\hat{\beta}_{-3}$ and $\hat{\beta}_{-2}$; the post-event spillover coefficients are $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$.

I estimate Equation 2 by OLS with standard errors clustered at the microregion level, following Corbi et al. (2019). IBGE defines a microregion as *groups of economically integrated municipalities sharing borders and the structure of production* (BRASIL, 1990). In the neighbour subsample there are 477 microregions, and 45% of bordering pairs lie in the same one. The rationale is that spillovers are more likely between economically and geographically integrated jurisdictions.

Table A9: Robustness to census-reset timing: spillover ATTs excluding the 2002/2008/2011 cohorts

Neighbour outcome (log, %)	Full sample	Off-reset cohorts
	ATT (s.e.)	ATT (s.e.)
Current expenditure	1.08 (0.66)	2.66 (0.90)***
Education expenditure	3.71 (1.01)***	5.19 (1.33)***
Public-sector employment	2.67 (1.29)**	1.78 (1.66)
Private-sector employment	−1.05 (1.53)	−0.91 (2.23)
Number of firms	−1.56 (1.00)	−0.82 (1.38)
Firms (services)	−1.50 (1.05)	−1.41 (1.41)
Government transfers (ex-FPM)	4.45 (2.98)	5.75 (4.77)
Population	0.45 (0.31)	0.86 (0.52)
Treated neighbours	886	327

Overall Callaway–Sant’Anna ATTs in percent (log coefficients $\times 100$), with neighbour-clustered influence-function standard errors in parentheses. The *Off-reset cohorts* column drops treated neighbours whose cohort G_i is a census-reset year (2002, 2008, 2011); 559 of the 886 treated neighbours fall in those three cohorts, leaving 327 “smooth-crossing” treated neighbours, with the comparison group unchanged. The fiscal spillovers are not driven by census-reset timing: education and current expenditure are, if anything, *larger* and significant at the 1% level on the off-reset cohorts. Public-sector employment loses significance with the smaller treated sample but retains a positive point estimate; the private-side responses remain negative and imprecise. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The TWFE coefficients of Equation 2 are reported below, one panel per outcome. Each plots $\hat{\beta}_l$ for $l \in \{-4, \dots, 3\}$ (with $l = -1$ the reference), scaled by 100 so that a value of k reads as a $k\%$ change relative to event time -1 . Whiskers are 90% confidence intervals from microregion-clustered standard errors. Figures OA1–OA5 mirror the outcome groupings of Section 4, so the TWFE and CS profiles can be compared side by side. Where signs and shapes agree, the CS conclusions are unchanged; where the magnitudes diverge, the gap is one possible diagnostic for the weighting issue of de Chaisemartin and D’Haultfoe uille (2020); Goodman-Bacon (2021), but I do not claim it isolates the contribution of negative-weighted 2×2 comparisons — aggregation, comparison-group selection and event-window support also differ across the two specifications and could account for part of the gap.

A second caveat applies to the TWFE pre-event leads. Where these are non-negligible, the TWFE specification fails its own pre-trend test for that outcome, and the corresponding CS estimate of Section 4 — whose own pre-event placebos I report — should be preferred. This bears with particular force on the mechanism panels, Figures OA4 (non-FPM federal transfers) and OA5 (neighbour population), which carry substantive interpretation in the main text, but it is not confined to them: the binned endpoint coefficients ($l = -4$) of the firm-count panels (Figure OA3) are also sizeable and warrant the same caution, since a binned endpoint pools several event times. A TWFE pre-event failure is not in itself evidence against the CS estimate, but neither is it purely a mechanical artefact, since such leads can reflect gen-

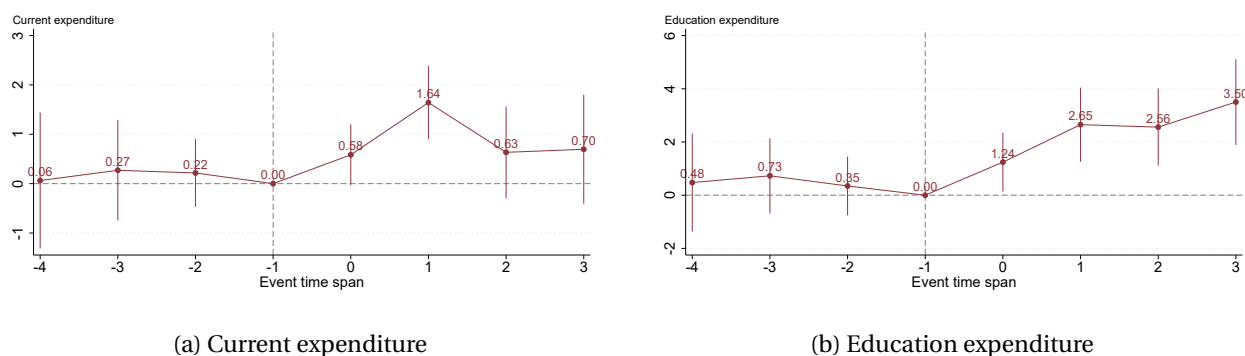
Table A10: Spatially-robust inference: neighbour- vs microregion-clustered standard errors

Neighbour outcome (log, %)	ATT	SE (neighbour)	SE (microregion)
Current expenditure	1.08	(0.66)	(0.68)
Education expenditure	3.71	(1.01)***	(1.08)***
Public-sector employment	2.67	(1.29)**	(1.35)**
Private-sector employment	−1.05	(1.53)	(1.55)
Number of firms	−1.56	(1.00)	(1.10)
Firms (services)	−1.50	(1.05)	(1.14)
Government transfers (ex-FPM)	4.45	(2.98)	(3.22)
Population	0.45	(0.31)	(0.33)
Clusters		1,854 neighbours	502 microregions

The Callaway–Sant’Anna ATTs (in percent; log coefficients $\times 100$) are identical across columns; only the clustering of the influence-function standard errors differs — at the neighbour level (the headline) and at the microregion level. Microregion clustering allows arbitrary correlation among neighbours of the same focal that lie in the same IBGE microregion (the relevant spatial unit, with 502 clusters here), addressing the concern that a single focal shock reaches several adjacent neighbours. It inflates the standard errors by only 5–9% and leaves the significance pattern unchanged: the education-expenditure spillover remains significant at the 1% level and public-sector employment at the 5% level. ** $p < 0.05$, *** $p < 0.01$.

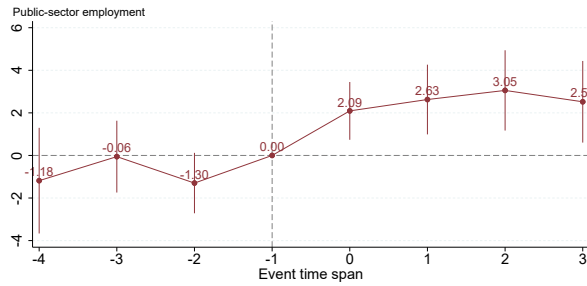
uine differential cohort trends. I therefore judge parallel trends for each outcome from the CS pre-event diagnostics directly, not from the TWFE-versus-CS contrast.

Figure OA1: TWFE spillover effects on the neighbour’s expenditures

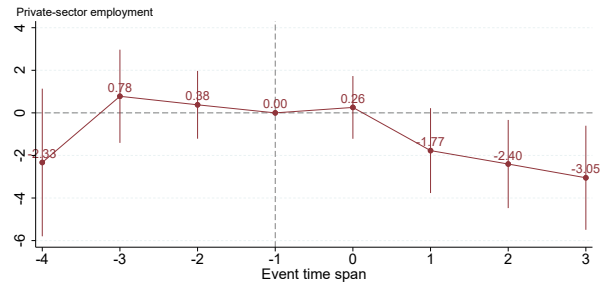


TWFE event-study counterpart of Figure 3.

Figure OA2: TWFE spillover effects on the neighbour's labour market



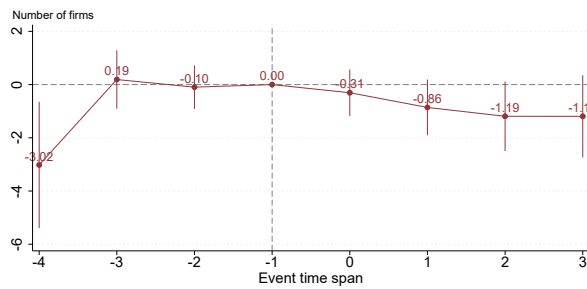
(a) Public-sector employment



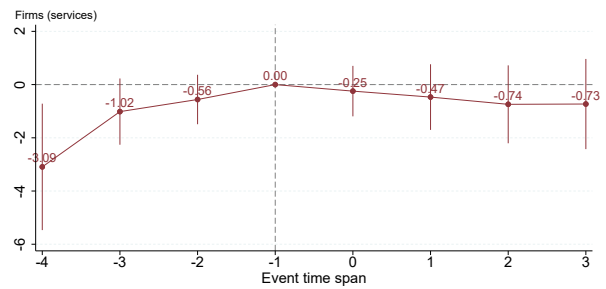
(b) Private-sector employment

TWFE event-study counterpart of Figure 4.

Figure OA3: TWFE spillover effects on firm dynamics



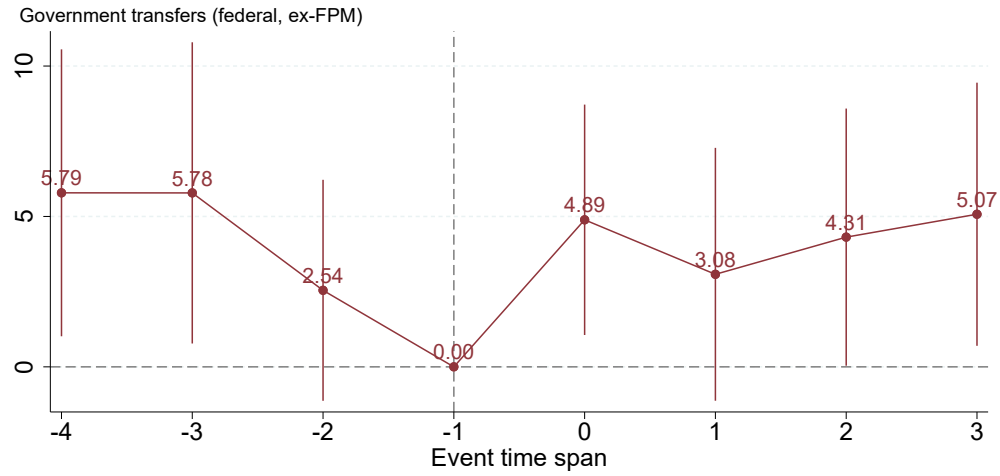
(a) Number of firms



(b) Services firms

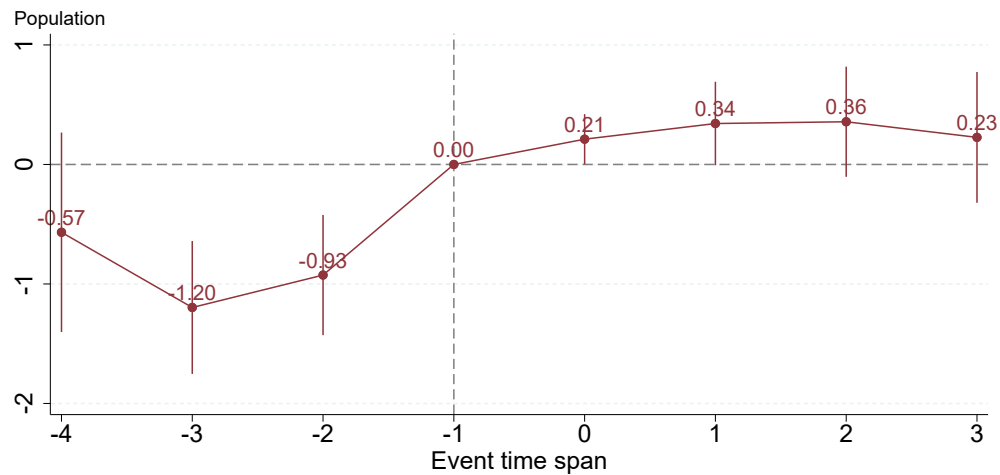
TWFE event-study counterpart of Figure 5.

Figure OA4: TWFE response of non-FPM federal transfers to the neighbour



TWFE event-study counterpart of Figure 6.

Figure OA5: TWFE response of the neighbour's population



TWFE event-study counterpart of Figure 7.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [descstatsup.tex](#)
- [refereeclusterup.tex](#)
- [csoverallup.tex](#)
- [refereelong2up.tex](#)
- [refereeownfixedup.tex](#)
- [methodologycs.tex](#)
- [refereeflowup.tex](#)
- [refereematchedup.tex](#)
- [ivtableup.tex](#)
- [sumstattable.tex](#)
- [refereebootup.tex](#)
- [refereebalanceup.tex](#)
- [refereematchbalup.tex](#)