Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Where Chinese Finance Lands, and What It Leaves in Reported Regional Output

Quoc Lap Nguyen

lapnq344@gmail.com

University of Economics and Law

Research Article

Keywords: Chinese development finance, Regional allocation, Reported regional output, Event study

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11461681/v1

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

Additional Declarations: No competing interests reported.

Where Chinese Finance Lands, and What It Leaves in Reported Regional Output

Quoc Lap Nguyen[1,2*]

[1*]University of Economics and Law, Ho Chi Minh City, Vietnam.
[2]Vietnam National University, Ho Chi Minh City, Vietnam.

Corresponding author(s). E-mail(s): lapnq344@gmail.com, lapnqc26602-1-nc@sdh.uel.edu.vn;

Abstract

Chinese official finance is large and poorly captured by conventional statistics, yet little is known about where it lands inside recipient countries or whether it leaves a trace in the output that regional accounts report. This paper links geocoded project records to a harmonised panel of reported gross regional product covering 1,103 first-level administrative regions in 48 countries from 2000 to 2019. Within countries, first projects reach richer and, above all, more populous regions, the tilt is stronger for non-concessional than for concessional finance, and recent growth has no predictive power. Staggered difference-in-differences estimates against same-country, not-yet-treated regions show no detectable effect of the first project on reported output per capita. The baseline point estimate is slightly negative but disappears under matched controls, and confidence intervals exclude gains beyond a few percent across specifications. Project size, sector, financing terms and governance quality yield no reliable heterogeneity, and a decline among recent cohorts proves fragile. The evidence suggests that Chinese finance follows economic weight rather than need and that its local returns, if present, are small or are not recorded in regional accounts, which bears on how governance and measurement shape the evaluation of external finance.

Keywords: Chinese development finance, Regional allocation, Reported regional output, Event study

JEL Classification: F35 , O18 , R11 , H54 , O47

1 Introduction

Few external financiers have redrawn the infrastructure map of developing countries as quickly as China, and few have left a thinner statistical trail. Official statistics capture only part of Chinese lending and project finance (Strange et al. 2017), and a substantial share of China's overseas lending is absent from the debt statistics of the borrowing countries (Horn et al. 2021). Geocoded project records assembled by AidData now make it possible to observe this finance directly. In the data used here, 2,862 projects with combined commitments of about USD 463 billion (constant 2021 prices) fall inside the regions of 48 countries between 2000 and 2019 (Goodman et al. 2024). Two questions follow from this scale. The first concerns where the money physically lands inside recipient countries. The second concerns whether it leaves a measurable trace in the output that regional statistical systems record.

Existing work has largely addressed neighbouring questions. At the country level, Chinese development finance has been found to raise short-term growth (Dreher et al. 2021), while the wider aid literature remains divided on whether external finance raises growth at all (Burnside and Dollar 2000; Rajan and Subramanian 2008). National aggregates, however, conceal the spatial margin at which infrastructure and productive investment operate. Work on allocation within countries has concentrated on political explanations: leaders' home regions receive more Chinese projects in Africa (Dreher et al. 2019), which echoes the broader evidence on regional favouritism in public spending (Hodler and Raschky 2014; Franck and Rainer 2012; Burgess et al. 2015). Work on local effects relies mainly on satellite-based measures such as night lights (Dreher et al. 2021; Bluhm et al. 2025), which proxy for activity rather than measure the output that regional accounts report (Henderson et al. 2012; Gibson et al. 2021).

Reported regional output deserves separate attention. It is the quantity on which fiscal rules, intergovernmental transfers, development rankings and political evaluations are built, and it is also the quantity most exposed to misreporting when incentives are strong (Martínez 2022; Michalski and Stoltz 2013). Whether a large external project registers in these accounts therefore speaks both to real economic change and to the way local statistical systems capture it. The question connects directly to the concerns of the economics of governance, which studies how political and institutional features shape the territorial distribution of public resources and the economic outcomes that follow (Hankins et al. 2017; Melnik 2024; Cerruti 2024; Rezki 2022; Mergo et al. 2023).

This paper combines the geocoded Chinese official finance dataset (Geo-GCDF v3.0; Goodman et al. 2024) with DOSE, a harmonised database of reported subnational economic output (Wenz et al. 2023). The matched panel covers 1,103 first-level administrative (ADM1) regions in 48 countries over 2000 to 2019. The analysis has two parts. The first documents, within countries, which regions receive their first project and when, using tercile comparisons, hazard models with country-by-year fixed effects and conditional logit. The second estimates the effect of the first project on reported gross regional product (GRP) per capita with a staggered difference-in-differences design in which every treated region is compared with regions of the same country that have not yet received a project. This design avoids the biases that two-way fixed effects suffer under staggered adoption (Goodman-Bacon 2021; de Chaisemartin and

D'Haultfœuille 2020; Callaway and Sant'Anna 2021; Sun and Abraham 2021; Borusyak et al. 2024). Because treatment is spread over a limited number of countries, inference relies on wild-cluster bootstrap procedures (Cameron et al. 2008; MacKinnon and Webb 2017; Webb 2023), supplemented by permutation tests (Young 2019), equivalence bounds (Hartman and Hidalgo 2018) and sensitivity analysis for violations of parallel trends (Rambachan and Roth 2023).

Three results emerge. First, allocation follows economic weight. Regions that are richer than their national average and, above all, more populous are far more likely to receive a first project, the tilt is sharper for non-concessional than for concessional finance, and recent growth has no predictive power. Second, the first project leaves no detectable mark on reported GRP per capita over the following five years. The baseline estimate is slightly negative but statistically indistinguishable from zero, falls to zero under matched controls, and the confidence intervals exclude gains above roughly four percent in every specification. Third, none of the dimensions along which gains might be expected (project size, sector, financing terms, governance quality) produces reliable heterogeneity. The one significant pattern, a decline among regions first treated in 2014 to 2019, is fragile to the estimator and coincides with differences in pre-treatment trends across cohorts.

The paper contributes along four lines. First, it supplies multi-country evidence on the within-country geography of Chinese finance that separates economic weight, relative need and recent momentum, and distinguishes finance by its terms. To the best of the author's knowledge, allocation has not previously been studied in this way across this many countries. The result complements the political-favouritism evidence (Dreher et al. 2019, 2021) by showing how much of the geography is already accounted for by size and income before political variables enter. Second, the outcome is reported GRP rather than night lights, which links the literature on Chinese finance with research on the quality of official sub-national statistics (Martínez 2022; Wenz et al. 2023). Third, the null is treated as a result to be bounded rather than an absence of evidence: minimum detectable effects, equivalence margins and breakdown values state which effect sizes the data can and cannot exclude. Fourth, the findings bear on governance. Finance that follows scale reinforces existing concentration, and finance whose local returns cannot be seen in regional accounts leaves the fiscal and political questions raised by hidden debt (Horn et al. 2021) and local corruption (Isaksson and Kotsadam 2018) without an offsetting output benefit that could be verified.

The remainder of the paper is organised as follows. Section 2 reviews the literature and states the hypotheses. Section 3 describes the data and the construction of the sample. Section 4 sets out the empirical strategy. Section 5 reports the results on allocation, effects, robustness and heterogeneity. Section 6 discusses interpretation and limitations, and Section 7 concludes.

2 Background and hypotheses

2.1 Chinese development finance and its subnational geography

In the project data (the Global Chinese Development Finance Dataset v3.0, on which the geospatial release of Goodman et al. 2024 builds), Chinese official finance is classified into flows that resemble official development assistance (ODA-like), flows on non-concessional terms comparable to other official flows (OOF-like), and a residual category of vague official finance. The distinction matters for governance because non-concessional finance is by construction closer to commercial lending, so repayment capacity and commercial viability can be expected to weigh more heavily in the choice of location. China's state-driven loans have indeed been found to resemble commercial lending more than official lending (Horn et al. 2021). A large share of this lending escapes conventional reporting (Strange et al. 2017; Horn et al. 2021), so project-level geocoding offers the only route to a spatial analysis.

The political and institutional correlates of Chinese finance have been studied from several angles. Across countries, political considerations shape Chinese aid allocation to about the same degree as that of Western donors, while allocation appears largely independent of recipients' natural resource endowments and institutional characteristics (Dreher and Fuchs 2015), which echoes the weight of political and strategic considerations in the decisions of traditional donors (Alesina and Dollar 2000). Chinese flows have been linked to higher local corruption (Isaksson and Kotsadam 2018) and to weaker economic reform (Brazys and Vadlamannati 2021), and their consequences for state stability have been contrasted with those of World Bank projects (Gehring et al. 2022). Within countries, the birth regions of political leaders receive more Chinese projects (Dreher et al. 2019), and the local effectiveness of projects has been examined with respect to such favouritism (Dreher et al. 2021). A subnational perspective is therefore well established, but it has so far centred on Africa and on night-light outcomes.

2.2 Allocation inside countries

Four mechanisms can explain where a project lands within a country. The first is economic weight. Large and prosperous regions offer demand, port and energy hubs, absorptive capacity and bankable counterparties, so non-concessional projects in particular are drawn to them. The second is need. Finance targeted at poverty would, by construction, favour poorer regions. The third is momentum, whereby finance flows to regions that have recently performed well, in the way that donors have been argued to reward sound policy environments (Burnside and Dollar 2000). The fourth is political favouritism, documented for the allocation of public goods and transport infrastructure (Hodler and Raschky 2014; Franck and Rainer 2012; Burgess et al. 2015).

The governance literature offers a close parallel. Studies in this field have examined how party alignment and polarisation relate to federal grant spending in the United States (Hankins et al. 2017), how legislative redistricting alters the partisan distribution of transportation expenditure (Melnik 2024), whether political connections matter for the growth of Italian cities (Cerruti 2024), how political competition relates to the

economic performance of Indonesian districts (Rezki 2022), and how ethnically asymmetric political representation relates to the provision of public goods in Ethiopia (Mergo et al. 2023). These studies share a focus on how political incentives and institutional design shape the territorial distribution of public resources and their economic returns. The present data do not contain local political variables, but country-by-year fixed effects absorb national politics, so the within-country gradients in size, income and growth are cleanly identified. Because the most populous region is often also the political centre, a population gradient is compatible with both the economic-weight and the favouritism mechanisms, a limitation returned to in Section 6.

H1 (allocation). Within countries, first projects are directed to richer and more populous regions rather than to poorer ones, the tilt is stronger for OOF-like than for ODA-like finance, and recent growth does not predict receipt.

2.3 Reported regional output as the outcome

Public capital has long been argued to raise productivity (Aschauer 1989), but local evidence on infrastructure is more mixed. Highways in China reduced output growth in the peripheral counties they connected (Faber 2014), access to transport lines raised the level of per capita output only modestly and left growth unchanged (Banerjee et al. 2020), and rural roads in India produced limited gains in local economic development (Asher and Novosad 2020). For aid at the regional level, a study of World Bank projects across more than a hundred countries finds no causal effect on growth measured by night lights (Dreher and Lohmann 2015). Chinese-financed infrastructure projects have been found to diffuse economic activity across space (Bluhm et al. 2025), so that gains may be redistributed between locations and leave the totals of an administrative unit largely unchanged. Regional growth is shaped by many of the same factors as national growth, such as geography and human capital (Gennaioli et al. 2014), and regional inequality by determinants such as natural resources, labour mobility and human capital (Lessmann and Seidel 2017). Comparing regions within the same country holds national factors constant but not these regional ones.

Measurement adds a further reason for caution. Reported output is compiled by national statistical offices, and reported growth in less democratic countries exceeds what night lights imply (Martínez 2022). Governments also appear at times to misreport economic data strategically, as evidence from balance-of-payments statistics suggests (Michalski and Stoltz 2013). Night lights are a useful but imperfect alternative (Henderson et al. 2012), and their relation to economic activity differs across data sources and places (Gibson et al. 2021). A project can therefore leave real effects that the accounts do not record, or can be recorded in ways that depart from true activity. Reported GRP is nonetheless the object on which governments, lenders and voters rely.

H2 (output). The first project raises reported GRP per capita in the regional accounts within a five-year horizon. The rival expectation is of no detectable effect, because gains are small, diffuse, delayed or not recorded.

H3 (heterogeneity). Effects differ with project size, sector orientation, financing terms and governance quality, and are larger for bigger projects, for productive rather than social projects, and in better-governed countries.

3 Data and sample

3.1 Projects, output and governance

Project data come from AidData's Global Chinese Development Finance Dataset, version 3.0, in its geospatial release (Geo-GCDF v3.0; Goodman et al. 2024), which records commitments of Chinese official sector institutions together with the footprints of project sites. Only projects with status Completion or Implementation that are recommended for aggregate analysis are retained. A project is attached to a region when its footprint overlaps the region by at least 1 percent of the project area, and its adjusted commitment (constant 2021 USD) is apportioned across regions in proportion to area overlap. The treatment year is the commitment year.

Table 4 summarises the 2,862 retained projects. Commitments are highly skewed, with a median of USD 31.1 million and a mean of USD 234.8 million, and the median project touches a single region. OOF-like projects number 1,263 but account for 90.0 percent of allocated value, whereas 1,470 ODA-like projects account for 6.8 percent. Industry, mining and construction (43.0 percent), energy (29.5 percent) and transport (18.0 percent) absorb over nine-tenths of value, while social sectors are the most numerous category (1,486 projects) but represent only 1.7 percent of value. Value is also concentrated across countries: Russia alone accounts for 24.5 percent of allocated value, followed by Kazakhstan (8.0 percent) and Pakistan (7.7 percent; Table A1 in the Appendix). This concentration motivates country-level clustering and leave-one-country-out checks.

Output data come from DOSE V2.14, a harmonised database of reported subnational economic output compiled from statistical agencies and other official sources (Wenz et al. 2023). The main outcome is GRP per capita in constant 2015 USD, and agricultural, manufacturing and services GRP per capita are used for sectoral analysis. DOSE regions are matched to ADM1 boundaries from GADM version 3.6 for the spatial overlay (these boundaries are not redistributed with this paper). The panel spans 2000 to 2019 and contains 1,103 regions in 48 countries, all with a valid GRP per capita series. Table 3 shows that mean GRP per capita in the estimation sample is USD 4,860 against a median of USD 3,449, with an interquartile range of USD 1,844 to 6,065, so the sample covers regions at very different income levels.

Governance is measured with the Worldwide Governance Indicators (Kaufmann et al. 2011). A country-level index is constructed as the mean of the standardised indicators averaged over 2006 to 2008, a window that precedes more than half of all first projects, and countries are split at the sample median into better and weaker governance. All variables are defined in Table A3 of the Appendix.

3.2 Treatment definitions and estimation samples

Treatment is the first year in which a region receives a qualifying project (T1, any project). Regions whose first project falls in 2002 or earlier are classed as left-censored and excluded, since treatment may pre-date the panel (Table 1). This removes 88 regions in 29 countries and leaves 573 eligible treated regions with a first project in 2003 to 2019. The never-treated and not-yet-treated pool contains 442 regions. Alternative definitions restrict treatment to first projects that are productive or infrastructure-related (T2), social or humanitarian (T3), OOF-like (T5) or ODA-like (T6), to any project with an allocated value of at least USD 50 million (T4, labelled large), and to projects whose allocated value reaches stated shares of the region's median real GRP (Table 2). For restricted definitions the first project of any type must be of the stated type, and controls are regions with no project of any type yet.

The baseline estimate keeps treated regions with an observed pre-period at the year before treatment and with at least one not-yet-treated control in the same country. This leaves 435 regions in 38 countries for the any-project definition. Figure 1 shows the spatial pattern of treatment status in the study panel.

Table 1 Sample construction

Sample step	Regions	Countries
DOSE ADM1 regions in the panel	1,103	48
with a valid GRP per capita series	1,103	48
with at least one GCDF project (2000 to 2019)	686	48
left-censored (first project in 2002 or earlier), excluded	88	29
never treated by 2019 (regions with no project)	417	35

Notes: Regions whose first GCDF project (commitment year) falls in 2002 or earlier are excluded because treatment may pre-date the panel, which starts in 2000. Projects are assigned to ADM1 regions by area-weighted spatial overlay of Geo-GCDF project polygons with DOSE region boundaries.

4 Empirical strategy

4.1 Allocation models

The descriptive analysis compares regions across within-country terciles of initial income and initial population. Initial relative income is the mean log GRP per capita over the first three observed years minus the country mean.

The determinants of receiving the first project are then modelled with a hazard linear probability model. For region i in country c and year t, at risk from 2003 until the year of its first project,

$$F_{ict} = \alpha_{ct} + \beta_1 \,\mathrm{RelInc}_{i,t-1} + \beta_2 \,\mathrm{LogPop}_{i,t-1} + \beta_3 \,\mathrm{Growth}^3_{i,t-1} + \varepsilon_{ict}, \qquad (1)$$

where $F_{ict} = 100 \times \mathbb{1}\{t = g_i\}$ and g_i is the year of the first project. RelInc is log GRP per capita relative to the country-year mean, LogPop is log population, and Growth^3 is the average annual growth rate of GRP per capita over the preceding three years.

Table 2 Treatment definitions and estimation samples

Treatment definition	Eligible treated	Control pool	Baseline regions	Baseline countries
T1: any project	573	442	435	38
T2: first project productive/infrastructure	294	442	228	36
T3: first project social/humanitarian	244	442	183	32
T4: any project $\geq$ USD 50m	264	751	218	37
T5: first project OOF-like	261	442	229	34
T6: first project ODA-like	287	442	179	32
Any project $\geq$ 1% of regional GRP	269	746	213	37

Notes: Eligible treated regions have a first qualifying project in 2003 to 2019. Control pool is the number of regions that are never or not yet treated under the definition. For the restricted definitions (T2, T3, T5, T6) the first project of any type must be of the stated type, and controls are regions with no project of any type yet. For the size-based definitions (T4 and the relative-size row) controls are regions that have not yet received a project meeting the threshold. Baseline regions and countries refer to the estimation sample, which keeps treated regions with an observed pre-period at $g-1$ and at least one same-country, not-yet-treated control.

Table 3 Descriptive statistics of the region-year panel

Variable	Obs.	Mean	SD	P25	Median	P75
GRP per capita (2015 USD)	16,494	4,860	5,265	1,844	3,449	6,065
Agricultural GRP per capita (2015 USD)	13,441	581	499	278	450	709
Manufacturing GRP per capita (2015 USD)	13,452	2,036	3,921	486	1,120	2,179
Services GRP per capita (2015 USD)	13,452	2,650	2,623	1,153	1,971	3,293
Population	16,102	3,143,787	11,361,082	407,250	955,218	1,992,156

Notes: Region-years of the DOSE V2.14 panel for 2000 to 2019 after dropping left-censored regions (1,015 regions in 47 countries). Monetary values are in constant 2015 US dollars.

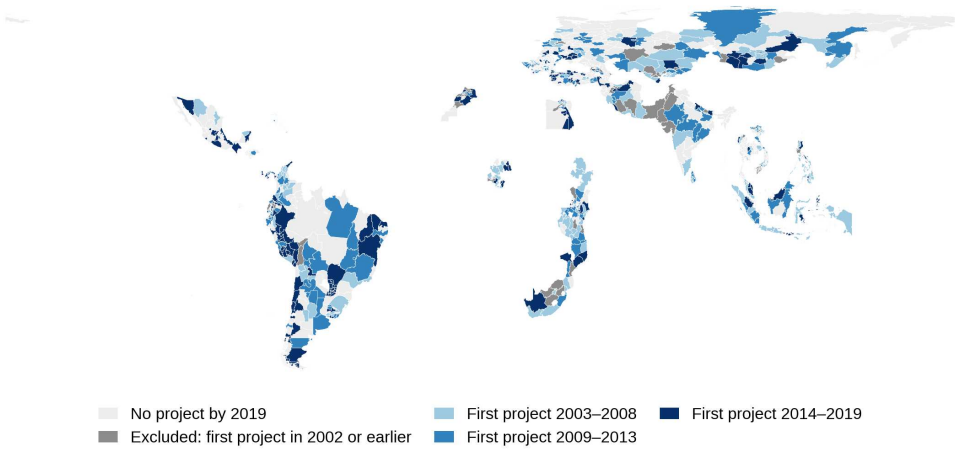


Fig. 1 Treatment status of regions in the study panel. The map distinguishes regions without a project, regions excluded as left-censored (first project in 2002 or earlier), and regions by cohort of first project (2003 to 2008, 2009 to 2013, 2014 to 2019). Source: author's calculation from Geo-GCDF v3.0 and DOSE V2.14.

Table 4 Chinese development finance projects used in the analysis

Panel A. Project-level statistics		
Projects (unique)		2,862
Total commitments (USD billion, 2021)		463.2
Median project commitment (USD million)		31.1
Mean project commitment (USD million)		234.8
Project-region pairs		3,777
Median regions per project		1
Panel B. Flow class		
Flow class	*Projects*	*Share of value (%)*
OOF-like	1,263	90.0
ODA-like	1,470	6.8
Vague (Official Finance)	129	3.2
Panel C. Sector group		
Sector group	*Projects*	*Share of value (%)*
Industry, mining, construction	294	43.0
Energy	401	29.5
Transport and storage	326	18.0
Other	200	4.0
Communications	74	3.0
Social sectors	1,486	1.7
Agriculture, forestry, fishing	81	0.9

Notes: Projects with status Completion or Implementation and recommended for aggregate analysis in GCDF v3.0 that overlap at least one DOSE region by 1% of project area or more. Amounts are AidData adjusted amounts in constant 2021 US dollars. Shares use project value allocated to regions by area weights, whereas the number of projects counts each project once. Productive and infrastructure sectors are transport, energy, communications, industry-mining-construction and agriculture; the social sectors are defined in Table A3.

All regressors are lagged one year and standardised, so that coefficients are percentage-point changes in the annual probability of receipt per standard deviation. Country-by-year fixed effects α_{ct} absorb every national shock and national political condition, so identification comes from comparisons between regions of the same country in the same year. The linear probability form keeps the high-dimensional fixed effects tractable. As complements, a conditional logit with country-by-year strata is estimated for the first project, and a cross-sectional linear model with country fixed effects is estimated for the probability of ever receiving a project. Standard errors are clustered by country, the level at which sampling and national policy vary (Abadie et al. 2023). Heteroskedasticity across countries and serial dependence within them are further reasons for clustering at this level (Bertrand et al. 2004).

4.2 Staggered difference-in-differences

Two-way fixed effects regressions with staggered adoption average heterogeneous effects with weights that can be negative, and early-treated units serve as controls for late-treated ones (Goodman-Bacon 2021; de Chaisemartin and D'Haultfœuille 2020; Baker et al. 2022). The effect of the first project is therefore estimated with a group-time design in the spirit of Callaway and Sant'Anna (2021), simplified to a single comparison per region and event year. For treated region i first treated in year g_i and

event year e,

$$\hat{\tau}_{i,e} = \left(y_{i,g_i+e} - y_{i,g_i-1}\right) - \frac{1}{|C_{i,e}|} \sum_{j \in C_{i,e}} \left(y_{j,g_i+e} - y_{j,g_i-1}\right), \tag{2}$$

where y is the log of GRP per capita, the base period is the year before treatment ($e = -1$), and $C_{i,e}$ is the set of regions in the same country that have not yet received a project in calendar year $g_i + e$. Comparisons within the same country remove national shocks, exchange-rate movements and differences in national statistical practice, which matters for reported output. The post-treatment effect for a region is $\hat{\tau}_i = \frac{1}{5} \sum_{e=0}^{4} \hat{\tau}_{i,e}$, the headline estimate is the unweighted mean of $\hat{\tau}_i$ across treated regions, and it is reported in percent as $100 \times (\exp(\hat{\tau}) - 1)$. The pre-trend measure is the mean of $\hat{\tau}_{i,e}$ over $e \in \{-5, \ldots, -2\}$, and event-study coefficients are the means of $\hat{\tau}_{i,e}$ at each e from -5 to 5. The procedure uses never-treated and not-yet-treated regions only, which is the control-group logic recommended in the recent literature (Sun and Abraham 2021; Roth et al. 2023). A second estimator imputes untreated outcomes from region and country-by-year fixed effects fitted on untreated observations (Borusyak et al. 2024).

Treatment is the first project, so the estimand is the effect of becoming exposed to Chinese finance, including any later projects that follow in the same region. That feature makes the estimate policy relevant, because it measures what happens to regions after Chinese finance arrives, but it does not isolate the effect of one project.

4.3 Inference, power and sensitivity

Treatment occurs in 38 countries at most, with very unequal numbers of regions per country. Cluster-robust inference is unreliable when only a few groups are treated (Conley and Taber 2011) and when clusters differ greatly in size (MacKinnon and Webb 2017). Confidence intervals and p-values therefore come from wild-cluster bootstrap-t procedures with Webb weights and clusters defined by country (Cameron et al. 2008; MacKinnon and Webb 2017; Webb 2023), using 9,999 draws for the main estimates and 4,999 for event studies. P-values impose the null, and intervals use the unrestricted bootstrap-t distribution. A country-level pairs bootstrap and a permutation test that reassigns first-project timing across regions within countries (500 draws) provide checks that do not rely on the same asymptotics (Young 2019). Sectoral results form a family of twelve tests (three treatment definitions and four outcomes), and Holm's step-down adjustment is applied across that family to control the family-wise error rate (Holm 1979; Anderson 2008).

Pre-trends are summarised by the mean of the leads and by a clustered Wald test that all leads are jointly zero. Such tests are an incomplete safeguard, because pre-testing can distort subsequent inference (Roth 2022) and because pre-event trends can reflect confounding that persists after treatment (Freyaldenhoven et al. 2019). Three further devices therefore accompany the estimates. The minimum detectable effect (MDE) at 80 percent power and 5 percent size is 2.8× the standard error. Equivalence bounds follow the logic of equivalence testing (Hartman and Hidalgo 2018): the bound δ is the smallest symmetric margin that contains the 90 percent bootstrap-t interval, so effects larger than δ in absolute value are rejected at the 5 percent level. Finally,

sensitivity intervals widen the baseline interval by $3M\Delta$, where Δ is the largest change between consecutive pre-period event-study coefficients and M scales how far post-treatment departures from parallel trends may exceed that change. If the violation grows by at most $M\Delta$ per year, its average over event years 0 to 4 does not exceed $3M\Delta$. The breakdown value M^* is the value of M at which the upper bound reaches +2 percent. This is a conservative device in the spirit of relative-magnitude restrictions (Rambachan and Roth 2023) and is not the fixed-length confidence interval of that paper.

4.4 Robustness

The baseline is compared with the imputation estimator, with matched controls that satisfy a caliper on initial population and initial relative income, in the spirit of covariate matching (Abadie and Imbens 2006) (wide: within 1.0 log points of population and 0.5 of relative income; narrow: 0.5 and 0.25), with controls that are not adjacent to any region with a project through four years after treatment, with controls that exclude direct neighbours of treated regions, with population weights, and with the pre-period linear trend removed. Sample restrictions drop the most populous region in each country, shorten the sample period, exclude region-years flagged for structural change, require at least ten years of GRP data or at least five regions per country, and combine the three data-quality restrictions. Leave-one-country-out estimates assess dependence on single countries.

5 Results

5.1 Where the first project lands

Table 5 and Fig. 2 give the descriptive picture within countries. In the richest third of regions, 72.0 percent receive at least one project, against 53.3 percent in the poorest third, and the richest third absorbs 51.1 percent of allocated value while the poorest third absorbs 16.3 percent. The gradient by population is as steep: 75.5 percent of the most populous third of regions host a project and take 54.2 percent of value, against 53.1 percent of the least populous third. The most populous region of each country stands apart. In all 45 such regions, 95.6 percent host a project, the mean number of projects is 20.4, and they take 23.9 percent of value, although they represent only 4 percent of the regions in the table. The pattern is therefore one of concentration in large and prosperous places.

The hazard models in Table 6 and Fig. 3 separate size and income from recent performance within the same country and year. A one standard deviation (0.50 log points) higher relative income raises the annual probability of receiving the first project by 1.01 percentage points, and a one standard deviation (1.30 log points) higher population raises it by 3.24 percentage points, against a mean event rate of 4.52 points per year. The income effect is therefore equivalent to about a fifth of the average hazard and the population effect to about seven-tenths of it. Past growth has no predictive power (0.05 points, standard error 0.26). The same ordering holds for productive projects (0.64 and 1.42 points), social projects (0.42 and 1.68), and large projects (0.32 and

Table 5 Where projects land: allocation within countries

Group	Regions	With ≥ 1 project (%)	Mean projects	Mean value (USD m)	Share of value (%)
Panel A. Within-country tercile of initial income					
Poorest third	366	53.3	1.98	201.9	16.3
Middle third	355	60.6	2.47	417.1	32.6
Richest third	378	72.0	5.49	612.9	51.1
Panel B. Within-country tercile of initial population					
Least populous third	356	53.1	1.92	338.6	28.5
Middle third	345	59.7	2.21	211.1	17.2
Most populous third	368	75.5	5.96	622.5	54.2
Panel C. Most populous region of each country					
Most populous region	45	95.6	20.38	2,412.8	23.9
All other regions	1,054	60.6	2.62	327.4	76.1

Notes: Terciles are computed within each country (countries with at least three regions with data). Initial income is the mean log GRP per capita over the first three observed years relative to the country mean. Allocated value is project value apportioned to regions by area overlap (USD 2021). Sample: 46 countries.

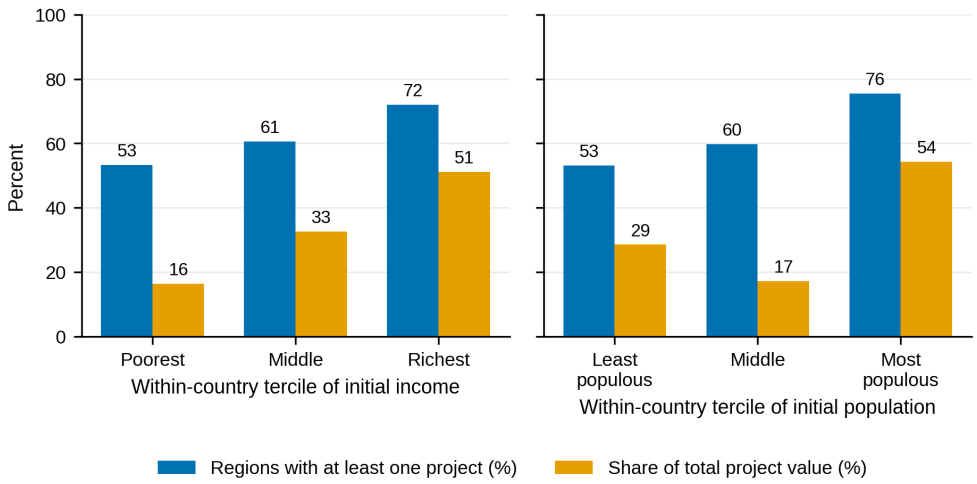


Fig. 2 Allocation of Chinese projects across within-country terciles of initial income (left) and initial population (right). Blue bars show the share of regions in each tercile with at least one project; orange bars show the tercile's share of total allocated project value. Terciles are formed within each country. Source: author's calculation from Geo-GCDF v3.0 and DOSE V2.14.

0.83). Terms of finance matter: for OOF-like first projects the income and population gradients are 0.87 and 2.20 points, whereas for ODA-like first projects the income gradient is absent (−0.01) and the population gradient is smaller and only marginally significant (0.95). The conditional logit in Table 7 gives the same ranking, with odds ratios of 1.27 for relative income and 2.58 for population per standard deviation and an odds ratio of 1.02 (95% interval 0.89 to 1.16) for past growth.

The gradients are stable over time and across specifications. By sub-period, the income coefficient is 1.19 for 2003 to 2008, 1.05 for 2009 to 2013 and 0.60 for 2014 to 2019, while the population coefficient stays between 2.7 and 4.1 points, and none of the

Table 6 Determinants of receiving the first project: hazard linear probability models

First project of type	Relative income	Log population	Past growth	Region-years	Regions	Countries	Mean rate (pp/yr)
Any project	1.01^{***} (0.21)	3.24^{***} (0.46)	0.05 (0.26)	8,406	832	38	4.52
Productive/infrastructure (first project)	0.64^{***} (0.16)	1.42^{***} (0.22)	0.01 (0.26)	8,406	832	38	2.36
Social/humanitarian (first project)	0.42^{***} (0.12)	1.68^{***} (0.33)	−0.04 (0.13)	8,406	832	38	1.90
Large project $\geq$ USD 50m	0.32^{**} (0.15)	0.83^{***} (0.16)	−0.04 (0.16)	8,406	832	38	1.31
First project OOF-like	0.87^{***} (0.16)	2.20^{***} (0.30)	0.09 (0.29)	8,406	832	38	2.43
First project ODA-like	−0.01 (0.14)	0.95^{*} (0.47)	−0.08 (0.13)	8,406	832	38	1.76

Notes: The dependent variable is $100 \times \mathbf{1}\{\text{year of the region's first project of the given type}\}$; coefficients are percentage-point changes in the annual probability per one standard deviation of the regressor (SD of relative income = 0.504 log points, of log population = 1.296, of past growth = 5.331 pp per year). Regressors are lagged one year. All models include country-by-year fixed effects; at-risk region-years run from 2003 until the first project. Standard errors clustered by country are in parentheses. Stars denote cluster-robust p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

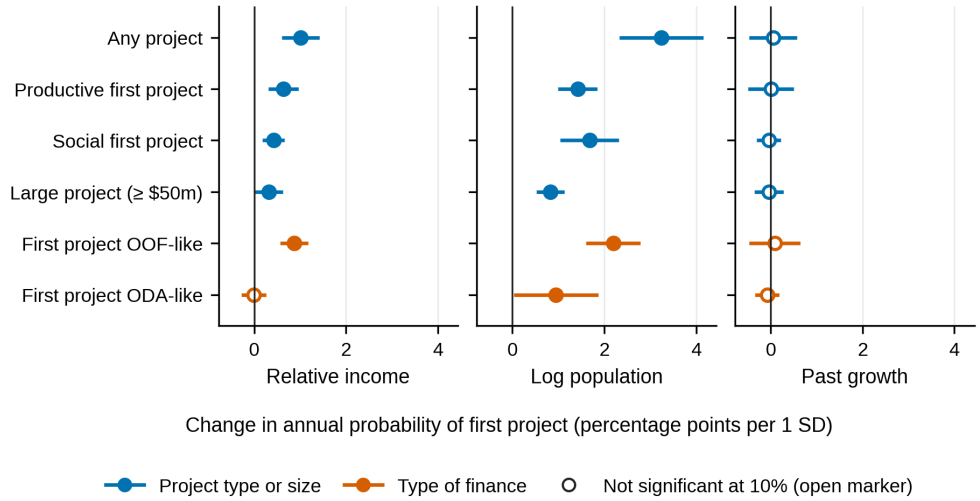


Fig. 3 Determinants of receiving the first project. Each panel reports the change in the annual probability of the first project (percentage points per one standard deviation of the regressor) from the hazard linear probability models with country-by-year fixed effects, together with 95 percent confidence intervals based on country-clustered standard errors. Blue markers refer to project type or size, vermilion markers to the type of finance, and open markers denote coefficients not significant at the 10 percent level. Source: author's calculation.

Table 7 Conditional logit for receiving the first project

Regressor (per SD)	Odds ratio	95% CI	p-value
Relative income	1.27	[1.09, 1.49]	0.003
Log population	2.58	[2.12, 3.13]	<0.001
Past growth (3-year)	1.02	[0.89, 1.16]	0.810

Notes: Conditional logit with country-by-year strata (378 first-project events, 181 informative strata, 38 countries). Confidence intervals and p-values come from a country-level pairs bootstrap with 199 replications.

interactions with period is individually distinguishable from the base period except a marginal rise in the population gradient in 2009 to 2013 (Table A4). The cross-sectional model with country fixed effects points to the same conclusion for ever receiving a project: 7.13 percentage points per standard deviation of initial relative income and 18.24 per standard deviation of initial population, with no role for population density (Table A5). Total project value per head rises with income and population as well.

These results support H1. They are consistent with the cross-country finding that Chinese aid allocation is largely independent of recipients' institutional characteristics and resource endowments (Dreher and Fuchs 2015), and they show that within countries projects, above all non-concessional ones, go where economic mass is. They also differ from what a need-based or growth-chasing allocation would produce, since poorer regions are less likely to be selected and recent growth carries no information about selection. Because the most populous region is often the political centre, the

population gradient may in part reflect the favouritism documented for leaders' home regions (Dreher et al. 2019; Hodler and Raschky 2014); the data used here cannot separate the two channels.

5.2 Effect of the first project on reported output

Table 8 reports the baseline estimates. For any first project, the average effect over event years 0 to 4 on GRP per capita is −1.01 percent (standard error 0.78, 95% interval −2.52 to 0.51, wild-bootstrap $p = 0.210$) in 435 regions and 38 countries. The point estimates are −1.38 percent for productive first projects, −0.54 percent for large projects, −1.92 percent for social projects ($p = 0.059$, the only coefficient that reaches the 10 percent level) and +0.35 percent for projects of at least 1 percent of regional GRP. No definition yields a significant positive effect, so H2 receives no support in its positive form.

The event study in Table 9 and Fig. 4 shows why the baseline sign should not be read as a causal reduction. For the any-project definition, the lead coefficients are close to zero at event years −5 and −4 (−0.12 and −0.06) and then turn negative at −3 (−0.97, $p < 0.10$) and −2 (−0.73, $p < 0.05$). The post-treatment coefficients, −0.68, −0.58, −0.45, −0.88 and −1.38 for event years 0 to 4, are of the same order and none is distinguishable from zero. The joint test of the leads has $p = 0.088$ for the any-project definition and larger p-values for productive and large projects. The post-treatment path continues the pre-treatment drift rather than departing from it: treated regions were already falling slightly behind their same-country comparators before the first project arrived.

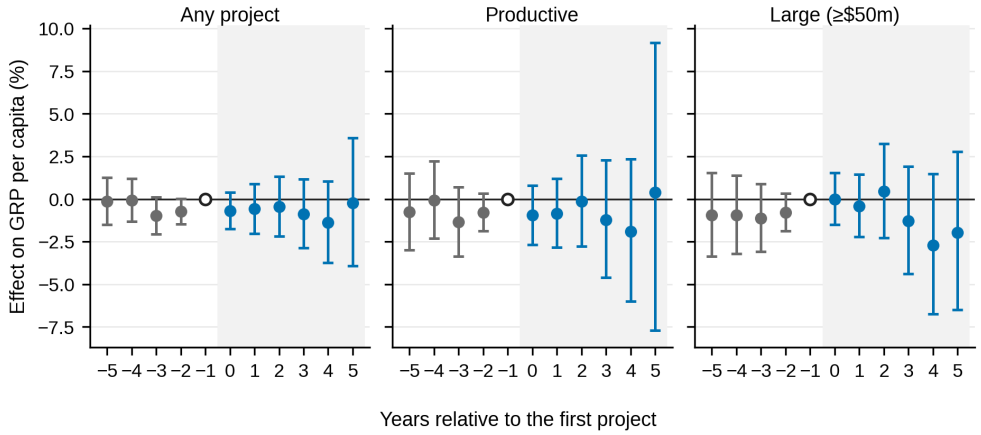


Fig. 4 Event-study estimates of the effect of the first project on GRP per capita (percent), relative to same-country not-yet-treated regions. Panels refer to any project, productive or infrastructure first projects, and large projects (at least USD 50 million). Grey markers are pre-treatment coefficients, blue markers are post-treatment coefficients, the hollow marker is the reference year ($e = -1$), and the shaded area marks the post-treatment period. Bars are 95 percent wild-cluster bootstrap-t intervals. Source: author's calculation.

Table 8 Effect of the first Chinese development project on reported GRP per capita

Treatment definition	Effect (%)	SE	95% CI	p (wild)	Regions	Countries	Pre-trend (%)	p (pre)	p (leads)	MDE (%)
Any project	−1.01	0.78	[−2.52, 0.51]	0.210	435	38	−0.59	0.248	0.088	2.22
Productive/infrastructure project (first project)	−1.38	1.26	[−3.94, 1.26]	0.323	228	36	−0.91	0.340	0.325	3.60
Social/humanitarian project (first project)	-1.92^{*}	0.96	[−3.83, 0.03]	0.059	183	32	−0.05	0.930	0.509	2.72
Large project (≥ USD 50m)	−0.54	0.93	[−2.38, 1.34]	0.577	218	37	−0.86	0.303	0.726	2.65
Project ≥ 1% of regional GRP	0.35	0.99	[−1.65, 2.39]	0.726	213	37	−0.63	0.463	0.147	2.81

Notes: Effects are the average over event years 0 to 4 of region-level differences-in-differences in the log of the outcome relative to same-country controls, converted to percent as $100 \times (\exp(b) - 1)$. Confidence intervals are bootstrap-t intervals from 9,999 Webb-weight wild-cluster draws (clusters are countries); standard errors are cluster-robust (CR1) in log points times 100. Pre-trend is the average over event years −5 to −2; p (leads) tests that these leads are jointly zero (clustered Wald). MDE is the minimum detectable effect at 80% power and 5% size (2.8×SE). Stars denote wild-cluster-bootstrap p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 9 Event-study estimates for log GRP per capita

	Any project			Productive			Large (≥USD 50m)		
Event year	Effect (%)	95% CI	Regions	Effect (%)	95% CI	Regions	Effect (%)	95% CI	Regions
−5	−0.12	[−1.50, 1.27]	350	−0.77	[−2.98, 1.49]	197	−0.94	[−3.37, 1.55]	191
−4	−0.06	[−1.31, 1.20]	395	−0.07	[−2.32, 2.24]	211	−0.93	[−3.21, 1.40]	198
−3	−0.97*	[−2.06, 0.13]	427	−1.34	[−3.35, 0.71]	227	−1.12	[−3.08, 0.89]	209
−2	−0.73**	[−1.48, 0.01]	435	−0.77	[−1.88, 0.34]	227	−0.79	[−1.89, 0.33]	211
−1	0 (ref.)		447	0 (ref.)		236	0 (ref.)		223
0	−0.68	[−1.76, 0.40]	428	−0.95	[−2.69, 0.81]	222	−0.00	[−1.51, 1.52]	215
1	−0.58	[−2.04, 0.90]	399	−0.84	[−2.82, 1.19]	202	−0.41	[−2.23, 1.44]	191
2	−0.45	[−2.19, 1.32]	383	−0.15	[−2.78, 2.56]	192	0.45	[−2.27, 3.24]	181
3	−0.88	[−2.87, 1.15]	356	−1.22	[−4.60, 2.27]	172	−1.27	[−4.38, 1.93]	164
4	−1.38	[−3.74, 1.04]	325	−1.92	[−6.01, 2.35]	152	−2.72	[−6.75, 1.49]	138
5	−0.24	[−3.91, 3.58]	285	0.38	[−7.70, 9.17]	136	−1.97	[−6.49, 2.77]	126
Joint test of leads (p)		0.088			0.325			0.726	

Notes: Each coefficient is the mean over treated regions of the region's difference-in-differences relative to same-country not-yet-treated controls, with base period $g-1$ (event year -1), in percent. Confidence intervals are bootstrap-t intervals from 4,999 wild-cluster draws. Stars denote wild-cluster-bootstrap p-values: $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

5.3 Robustness

The robustness exercises in Table 10 and Fig. 5 reinforce a selection reading of the baseline. When controls are matched on initial population and relative income, the any-project estimate becomes +0.18 percent with the wide caliper (95% interval −1.10 to 1.48) and 0.00 percent with the narrow caliper (−1.48 to 1.51). The imputation estimator gives −0.84 percent (−2.90 to 1.27), removing the linear pre-trend gives −1.35 percent (−3.62 to 0.55), excluding the most populous region of each country gives −1.15 percent, and a population-weighted estimate is +0.47 percent (−2.80 to 3.85). Leave-one-country-out point estimates lie between −1.23 and −0.52 percent, so no single country drives the result. The only specification with a statistically significant coefficient for the any-project definition is the sample restricted to 2005 to 2019 (−1.93 percent, interval −3.50 to −0.33), whereas the sample restricted to 2000 to 2015 yields −0.33 percent. The picture is similar for productive and large projects. Across all specifications and treatment definitions in the table, no confidence interval has an upper bound above 3.85 percent.

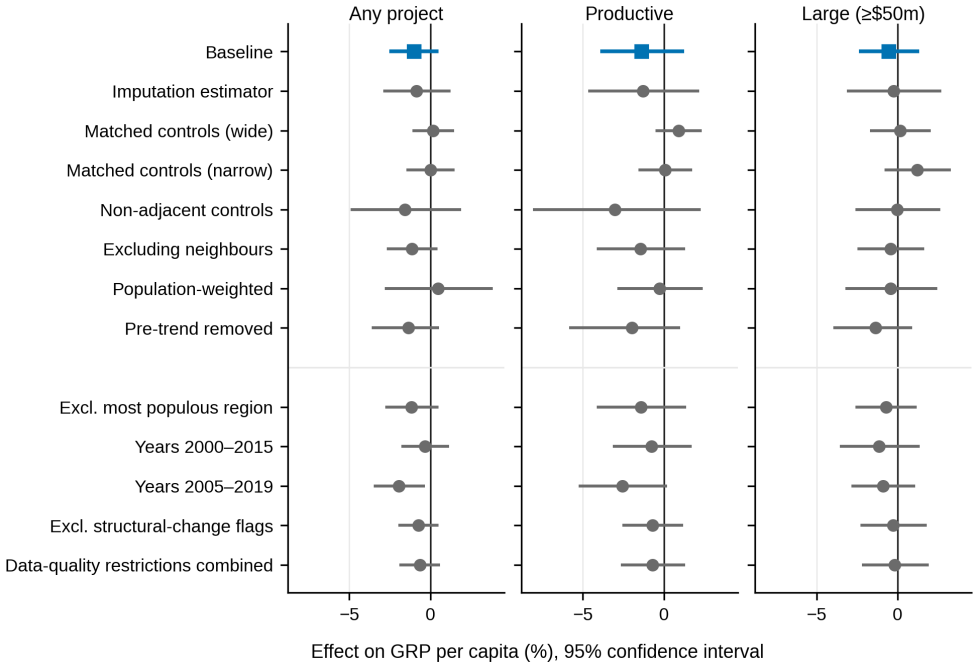


Fig. 5 Robustness of the effect on GRP per capita (percent) across estimators, control groups and samples. Panels refer to any project, productive or infrastructure first projects, and large projects. The blue square is the baseline specification, grey circles are alternative specifications, and horizontal bars are 95 percent confidence intervals. Source: author's calculation.

Table 10 Robustness of the effect on log GRP per capita to estimator, control group and sample

	Any project			Productive			Large (≥USD 50m)		
Specification	Effect (%)	95% CI	Regions	Effect (%)	95% CI	Regions	Effect (%)	95% CI	Regions
Panel A. Estimator and control group									
Baseline (Callaway and Sant'Anna type, same-country controls)	−1.01	[−2.52, 0.51]	435	−1.38	[−3.94, 1.26]	228	−0.54	[−2.38, 1.34]	218
Imputation estimator (Borusyak, Jaravel and Spiess type)	−0.84	[−2.90, 1.27]	458	−1.30	[−4.66, 2.17]	242	−0.24	[−3.11, 2.71]	219
Matched controls, wide caliper	0.18	[−1.10, 1.48]	366	0.91	[−0.50, 2.35]	195	0.15	[−1.71, 2.05]	191
Matched controls, narrow caliper	0.00	[−1.48, 1.51]	285	0.08	[−1.56, 1.75]	148	1.23	[−0.80, 3.30]	154
Controls not adjacent to any region with a project	−1.57	[−4.92, 1.91]	311	−3.04	[−8.08, 2.27]	164	−0.01	[−2.60, 2.66]	161
Excluding direct neighbours of treated regions from controls	−1.13	[−2.69, 0.46]	427	−1.46	[−4.15, 1.30]	224	−0.43	[−2.47, 1.66]	213
Population-weighted	0.47	[−2.80, 3.85]	435	−0.26	[−2.86, 2.41]	228	−0.42	[−3.22, 2.46]	218
Linear pre-trend removed	−1.35	[−3.62, 0.55]	435	−1.98	[−5.86, 1.00]	228	−1.37	[−3.97, 0.90]	218
Panel B. Sample restrictions									
Excluding the most populous region in each country	−1.15	[−2.79, 0.51]	410	−1.42	[−4.14, 1.38]	219	−0.72	[−2.59, 1.20]	202
Sample period 2000 to 2015	−0.33	[−1.79, 1.17]	350	−0.76	[−3.16, 1.71]	167	−1.12	[−3.55, 1.36]	153
Sample period 2005 to 2019	-1.93^{**}	[−3.50, −0.33]	361	-2.56^{**}	[−5.25, 0.21]	204	−0.90	[−2.85, 1.09]	208
Excluding region-years with structural-change flag	−0.74	[−1.98, 0.50]	407	−0.70	[−2.56, 1.19]	212	−0.27	[−2.28, 1.79]	191
Regions with at least 10 years of GRP	−0.92	[−2.49, 0.68]	411	−1.36	[−4.01, 1.36]	218	−0.42	[−2.39, 1.59]	202
Countries with at least 5 regions	−1.01	[−2.51, 0.51]	435	−1.38	[−3.93, 1.25]	228	−0.55	[−2.40, 1.33]	216
All three data-quality restrictions combined	−0.65	[−1.90, 0.62]	387	−0.70	[−2.65, 1.30]	205	−0.16	[−2.21, 1.92]	182
Leave-one-country-out (range of point estimates)		[−1.23, −0.52]			[−1.90, −0.43]			[−0.85, −0.15]	

Notes: Effects are the average over event years 0 to 4 of region-level differences-in-differences in the log of the outcome relative to same-country controls, converted to percent as $100 \times (\exp(b) - 1)$. Confidence intervals are bootstrap-t intervals from 9,999 Webb-weight wild-cluster draws (clusters are countries); standard errors are cluster-robust (CR1) in log points times 100. Wide caliper: controls within 1.0 log points of initial population and 0.5 of relative income; narrow caliper: 0.5 and 0.25. Clean controls are not adjacent to any region with a project through $g + 4$. The imputation estimator fits region and country-by-year fixed effects on untreated observations and averages imputed effects over event years 0 to 4. The linear-pre-trend row removes the pre-period slope (bootstrap over countries, 4,999 draws; no wild-bootstrap p-value, hence no stars). The last row reports the range of point estimates when each country is dropped in turn. Stars denote wild-cluster-bootstrap p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

5.4 Timing and relative size

Two features of the treatment definition could mask an effect: the timing of the project and its size relative to the regional economy. Dating treatment by the implementation start rather than the commitment year gives −1.35 percent (interval −3.23 to 0.57), close to the baseline (Table 11). Restricting treatment to first projects whose value reaches a given share of the region's median real GRP produces estimates of +0.05, +0.22, +0.35, +0.94 and +0.59 percent at thresholds of 0.25, 0.5, 1, 2 and 5 percent, none of which is significant (the 2 percent estimate has $p = 0.422$). The point estimates rise slightly with the threshold, but the intervals are wide and the rise does not form a dose-response pattern: the 5 percent threshold, the most demanding, yields an estimate smaller than the 2 percent threshold. Larger projects relative to the regional economy do not leave a clearer trace in reported output.

5.5 Sectors and population

A positive effect might be concentrated in one sector, so Table 12 reports sectoral outcomes. Agricultural GRP per capita falls by 3.23 percent after any first project (interval −5.91 to −0.47, wild $p = 0.024$) and by 4.76 percent after a productive first project. Neither survives adjustment for the twelve tests of the family: the Holm-adjusted p-values are 0.284 and 0.348. Manufacturing and services show no effect (effects of −0.20 and +0.60 percent for the any-project definition), although the manufacturing estimate carries a significant pre-trend ($p = 0.023$). A decline in agricultural output per head would be compatible with labour moving out of agriculture, a margin documented for rural roads in India (Asher and Novosad 2020), but the estimate does not pass the multiplicity correction and is treated as suggestive only.

Table 13 turns to population and total GRP. Population rises by 0.69 percent after any first project ($p = 0.030$), but the pre-trend is also significant ($p = 0.019$), so the coefficient cannot be read as a migration response. Total GRP does not change significantly (−0.44 percent). Population and total output therefore do not offer evidence of a hidden increase that per capita measures obscure.

5.6 Heterogeneity and spillovers

Table 14 and Fig. 6 explore the dimensions along which H3 expects differences. By cohort of first project, the any-project estimate is +1.33 percent for 2003 to 2008, −1.50 percent for 2009 to 2013 and −2.86 percent for 2014 to 2019 ($p = 0.009$), and the difference between the last and the first cohort is −4.22 percentage points ($p = 0.007$). By governance, the any-project estimate is −0.75 percent in better-governed countries and −1.40 percent in weaker-governed countries, with +0.77 and −1.49 for large projects and −0.16 and −2.52 for productive projects, and none of the six is significant. Point estimates are more negative under weaker governance, but the intervals overlap widely, so H3 receives no clear support. By financing terms, OOF-like first projects yield −0.79 percent and ODA-like first projects −1.70 percent, both insignificant. Regions without projects that are adjacent to treated regions show no change in GRP per capita (−0.84 percent, $p = 0.326$) or population (−0.03 percent),

Table 11 Treatment timing and relative project size

Definition of treatment	Eligible treated	Effect (%)	95% CI	p (wild)	Regions	Countries	p (pre)
Any project, implementation-start year	585	−1.35	[−3.23, 0.57]	0.205	442	39	0.192
Any project, commitment year (baseline)	573	−1.01	[−2.52, 0.51]	0.210	435	38	0.248
Any project with value $\geq$ 0.25% of regional GRP	358	0.05	[−1.56, 1.68]	0.949	281	39	0.587
Any project with value $\geq$ 0.50% of regional GRP	314	0.22	[−1.53, 2.01]	0.799	246	38	0.410
Any project with value $\geq$ 1.00% of regional GRP	269	0.35	[−1.65, 2.39]	0.726	213	37	0.463
Any project with value $\geq$ 2.00% of regional GRP	206	0.94	[−1.35, 3.28]	0.422	162	33	0.423
Any project with value $\geq$ 5.00% of regional GRP	146	0.59	[−1.34, 2.57]	0.532	115	27	0.420

Notes: Effects are the average over event years 0 to 4 of region-level differences-in-differences in the log of the outcome relative to same-country controls, converted to percent as $100 \times (\exp(b) - 1)$. Confidence intervals are bootstrap-t intervals from 9,999 Webb-weight wild-cluster draws (clusters are countries); standard errors are cluster-robust (CR1) in log points times 100. Relative size is the project value allocated to the region divided by the region's median real GRP over the panel; regions are treated when their first project meets the threshold, and controls are regions without such a project yet. Stars denote wild-cluster-bootstrap p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 12 Sectoral outcomes (family of twelve tests)

Outcome	Effect (%)	95% CI	p (wild)	p (Holm)	Regions	Countries	p (pre)
Treatment: any project							
Log GRP per capita	−1.01	[−2.52, 0.51]	0.210	1.000	435	38	0.248
Log agricultural GRP per capita	−3.23**	[−5.91, −0.47]	0.024	0.284	343	31	0.483
Log manufacturing GRP per capita	−0.20	[−2.81, 2.48]	0.896	1.000	345	31	0.023
Log services GRP per capita	0.60	[−0.75, 1.97]	0.382	1.000	345	31	0.949
Treatment: productive first project							
Log GRP per capita	−1.38	[−3.94, 1.26]	0.323	1.000	228	36	0.340
Log agricultural GRP per capita	−4.76**	[−8.66, −0.70]	0.032	0.348	188	29	0.205
Log manufacturing GRP per capita	−1.92	[−6.59, 2.98]	0.450	1.000	189	29	0.051
Log services GRP per capita	0.68	[−1.00, 2.39]	0.531	1.000	189	29	0.462
Treatment: large project (USD 50m or more)							
Log GRP per capita	−0.54	[−2.38, 1.34]	0.577	1.000	218	37	0.303
Log agricultural GRP per capita	3.09	[−2.53, 9.04]	0.385	1.000	178	31	0.062
Log manufacturing GRP per capita	2.31	[−1.97, 6.77]	0.285	1.000	181	31	0.062
Log services GRP per capita	−0.30	[−1.50, 0.92]	0.612	1.000	181	31	0.849

Notes: Family of 12 tests (3 treatments × 4 outcomes); the Holm adjustment is applied to the wild-bootstrap p-values across the whole family. Effects are the average over event years 0 to 4 of region-level differences-in-differences in the log of the outcome relative to same-country controls, converted to percent as $100 \times (\exp(b) - 1)$. Confidence intervals are bootstrap-t intervals from 9,999 Webb-weight wild-cluster draws (clusters are countries); standard errors are cluster-robust (CR1) in log points times 100. Stars denote wild-cluster-bootstrap p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 13 Secondary outcomes: population and total GRP

Outcome	Effect (%)	95% CI	p (wild)	Regions	Countries	Pre-trend (%)	p (pre)
Treatment: any project							
Log population	0.69^{**}	[0.09, 1.30]	0.030	430	38	−0.50	0.019
Log total GRP	−0.44	[−2.04, 1.20]	0.603	430	38	−1.08	0.053
Treatment: productive first project							
Log population	0.63	[−0.22, 1.49]	0.161	223	36	−0.57	0.049
Log total GRP	−0.98	[−3.69, 1.81]	0.473	223	36	−1.45	0.107
Treatment: large project (USD 50m or more)							
Log population	0.61	[−0.28, 1.50]	0.200	213	37	−0.26	0.321
Log total GRP	0.06	[−2.14, 2.31]	0.957	213	37	−1.31	0.172

Notes: Population effects should not be read as migration responses where pre-trends are significant (see the p-value of the pre-trend). Effects are the average over event years 0 to 4 of region-level differences-in-differences in the log of the outcome relative to same-country controls, converted to percent as $100 \times (\exp(b) - 1)$. Confidence intervals are bootstrap-t intervals from 9,999 Webb-weight wild-cluster draws (clusters are countries); standard errors are cluster-robust (CR1) in log points times 100. Stars denote wild-cluster-bootstrap p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

so there is no evidence of spillovers or diversion at the ADM1 scale, which would otherwise be a channel for the diffusion described by Bluhm et al. (2025).

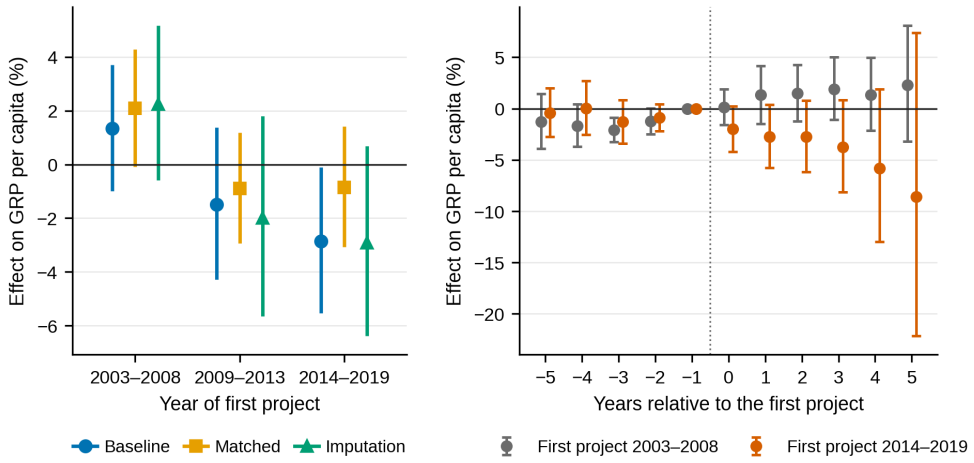


Fig. 6 Cohort heterogeneity of the effect on GRP per capita (percent). The left panel shows estimates by cohort of first project for the baseline estimator (blue circles), the matched-control estimator (orange squares) and the imputation estimator (green triangles) with 95 percent intervals. The right panel shows event-study estimates for regions first treated in 2003 to 2008 and in 2014 to 2019. Source: author's calculation.

The decline among 2014 to 2019 cohorts is the only pattern that stands out, and its fragility is considerable (Tables A6 and A7). With matched controls the estimate shrinks to −0.86 percent ($p = 0.384$), and with the imputation estimator it is −2.91 percent but loses significance ($p = 0.114$). Splitting the post window leaves the estimate unchanged (−2.87 percent for cohorts with a full window and −2.85 percent for cohorts with a truncated window), and leave-one-country-out estimates stay between −3.45 and −1.86 percent. The event studies by cohort (Table A8) clarify the contrast. The 2003 to 2008 cohort has significant negative leads at $e = -3$ (−2.07) and $e = -2$ (−1.25) and positive post-treatment coefficients thereafter, which resembles a recovery from a trough. Because that cohort's post-treatment coefficients are measured against a base year that follows the decline, the contrast between early and late cohorts partly reflects differences in pre-treatment trends rather than differences in treatment effects. Cohort heterogeneity was explored after the pooled analysis and is therefore treated as exploratory.

5.7 What the null can and cannot exclude

Table 15 states the information content of the null. For the any-project definition, the 90 percent interval is −2.29 to 0.28, so effects larger than 2.29 percent in absolute value are rejected at the 5 percent level in the equivalence sense; the corresponding bounds are 3.54 percent for productive projects and 2.07 percent for large projects. The MDE is 2.22, 3.60 and 2.65 percent. Permutation p-values (0.220, 0.267 and

Table 14 Heterogeneity and spillovers

Treatment	Group	Effect (%)	95% CI	p (wild)	Regions	Countries
Panel A. First-project cohort						
Any project	2003 to 2008	1.33	[−1.00, 3.72]	0.278	152	27
Any project	2009 to 2013	−1.50	[−4.24, 1.33]	0.289	127	31
Any project	2014 to 2019	-2.86^{***}	[−5.51, −0.13]	0.009	156	33
Any project	Difference 2014 to 2019 minus 2003 to 2008 (pp)	−4.22	[−8.40, −0.91]	0.007	308	
Any project	Difference 2009 to 2013 minus 2003 to 2008 (pp)	−2.83	[−7.11, 0.67]	0.121	279	
Any project	Difference 2014 to 2019 minus 2009 to 2013 (pp)	−1.39	[−4.87, 1.90]	0.416	283	
Productive	2003 to 2008	2.42	[−1.02, 5.98]	0.217	58	17
Productive	2009 to 2013	−2.36	[−6.15, 1.58]	0.223	76	27
Productive	2014 to 2019	−2.86	[−7.52, 2.03]	0.199	94	28
Productive	Difference 2014 to 2019 minus 2003 to 2008 (pp)	−5.29	[−12.40, 0.25]	0.062	152	
Productive	Difference 2009 to 2013 minus 2003 to 2008 (pp)	−4.78	[−12.12, 0.14]	0.062	134	
Productive	Difference 2014 to 2019 minus 2009 to 2013 (pp)	−0.51	[−5.15, 4.37]	0.861	170	
Panel B. Initial governance (WGI 2006 to 2008, country median)						
Any project	Better governance	−0.75	[−2.29, 0.80]	0.369	199	19
Any project	Weaker governance	−1.40	[−4.07, 1.34]	0.337	229	18
Productive	Better governance	−0.16	[−1.53, 1.22]	0.825	108	18
Productive	Weaker governance	−2.52	[−8.11, 3.41]	0.342	119	17
Large (≥USD 50m)	Better governance	0.77	[−1.00, 2.58]	0.382	95	18
Large (≥USD 50m)	Weaker governance	−1.49	[−4.79, 1.93]	0.361	120	18
Panel C. Type of first project						
OOF-like	All	−0.79	[−2.74, 1.20]	0.564	229	34
ODA-like	All	−1.70	[−4.03, 0.68]	0.151	179	32
Panel D. Spillovers to neighbouring regions without projects						
Neighbours of treated regions	Log GRP per capita	−0.84	[−2.59, 0.94]	0.326	223	21
Neighbours of treated regions	Log population	−0.03	[−0.78, 0.73]	0.939	222	21

Notes: Effects are the average over event years 0 to 4 of region-level differences-in-differences in the log of the outcome relative to same-country controls, converted to percent as $100 \times (\exp(b) - 1)$. Confidence intervals are bootstrap-t intervals from 9,999 Webb-weight wild-cluster draws (clusters are countries); standard errors are cluster-robust (CR1) in log points times 100. Cohort differences are differences of group means (in percentage points) with a country-level pairs bootstrap (4,999 draws; percentile interval). Panel D uses regions that never receive a project and are adjacent (within 2 km) to a region with one; controls are never-project regions with no adjacent project region. Stars denote wild-cluster-bootstrap p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

0.615) and country-pairs bootstrap intervals agree with the wild-bootstrap inference. Under the sensitivity analysis, the any-project interval is −3.84 to 1.90 percent when post-treatment departures from parallel trends may be half as large as the largest pre-treatment change ($M = 0.5$) and −5.16 to 3.31 percent when they may be as large ($M = 1$). The breakdown value at which the upper bound reaches +2 percent is 0.54 for any project, 0.19 for productive projects and 0.28 for large projects. The data therefore rule out sizeable positive effects on reported output per capita for a typical first project, but they cannot rule out effects of one or two percent, nor larger effects for productive projects in the presence of modest violations of parallel trends.

6 Discussion

6.1 Finance follows economic weight

The allocation results sketch a simple rule. Within a country, the first Chinese project goes to the region that is larger and richer than its peers, and it makes little difference whether that region has recently grown. The rule is sharper for non-concessional finance, for which both gradients are large, than for concessional finance, for which only population matters. This ordering is what would be expected if non-concessional finance is selected more on commercial viability, and it fits the cross-country evidence that aid decisions are steered by considerations other than recipient need (Dreher and Fuchs 2015; Alesina and Dollar 2000). A need-based allocation would instead predict a pro-poor tilt.

For governance, two implications follow. First, finance directed to the largest region in each country concentrates fiscal and political attention where political power already sits, which may reinforce regional favouritism rather than alleviate it (Hodler and Raschky 2014; Dreher et al. 2019). The same logic appears in research on how partisan alignment and electoral boundaries shape the territorial distribution of public funds (Hankins et al. 2017; Melnik 2024). Second, if finance flows to places that already have mass and income, it is unlikely to narrow regional gaps within countries, which converge only slowly, at roughly two percent per year (Gennaioli et al. 2014). Neither implication depends on the causal effect of the projects, and both hold even if returns are positive.

6.2 Interpreting the absence of a detectable effect

Four readings of the null merit consideration. The first is that the first project has no effect on output per capita within the horizon studied. Evidence on the local effects of transport infrastructure points the same way: output levels respond modestly and growth not at all (Banerjee et al. 2020), peripheral locations can lose rather than gain (Faber 2014), and rural roads in India mainly moved workers out of agriculture without clear gains in income or assets (Asher and Novosad 2020). Evidence on aid at the regional level is also weak (Dreher and Lohmann 2015). A small average effect is therefore not out of line with the wider literature, even though the aggregate association between Chinese finance and country growth is positive (Dreher et al. 2021).

Table 15 Inference, equivalence bounds and sensitivity to violations of parallel trends

Treatment	Effect (%)	CI (wild-t)	CI (pairs)	p (wild)	p (perm.)	90% CI (equiv.)	δ (%)	MDE (%)	Sens. CI, $M = 0.5$	Sens. CI, $M = 1$	M^*
Any project	−1.01	[−2.52, 0.51]	[−2.67, 0.41]	0.210	0.220	[−2.29, 0.28]	2.29	2.22	[−3.84, 1.90]	[−5.16, 3.31]	0.54
Productive	−1.38	[−3.94, 1.26]	[−4.12, 0.82]	0.323	0.267	[−3.54, 0.84]	3.54	3.60	[−5.77, 3.22]	[−7.56, 5.23]	0.19
Large (≥$50m)	−0.54	[−2.38, 1.34]	[−2.35, 1.33]	0.577	0.615	[−2.07, 1.03]	2.07	2.65	[−3.53, 2.55]	[−4.67, 3.77]	0.28

Notes: Permutation p-values reassign first-project timing across regions within countries (500 draws). The equivalence bound δ is the smallest symmetric margin containing the 90% bootstrap-t interval. Sensitivity intervals widen the baseline interval by $3 \cdot M \cdot \Delta$, where Δ is the largest change between consecutive pre-period event-study coefficients; this is a conservative bound in the spirit of relative-magnitude restrictions and not the Rambachan and Roth fixed-length confidence interval. M^* is the value of M at which the upper bound reaches +2% (a negative value would mean that the upper bound already exceeds +2% at $M = 0$). Stars denote wild-cluster-bootstrap p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The second reading is spatial. Chinese-financed infrastructure redistributes activity across locations (Bluhm et al. 2025), so a project can raise activity in one district while lowering it elsewhere in the same region, leaving the ADM1 total close to unchanged. The test on neighbouring regions without projects finds no detectable spillover, but it addresses diffusion across regions and not within them. The third reading concerns the treatment itself. The first project marks the start of exposure, which can involve further projects of different types, so the estimand mixes projects of different size, sector and timing. The relative-size and sectoral analyses do not isolate a subset with detectable gains.

The fourth reading concerns measurement. Reported regional output depends on national statistical practice, and there is evidence that official growth is overstated where accountability is weak (Martínez 2022) and that governments adjust data strategically (Michalski and Stoltz 2013). If a project's effects are concentrated in activities that regional accounts record poorly, the accounts will understate them. Night lights provide an independent check but have their own limits (Henderson et al. 2012; Gibson et al. 2021). The present results do not show that Chinese finance has no economic effect. They show that a first project does not leave a detectable effect larger than a few percent on the output that regional statistical systems report, which is the quantity that fiscal and political institutions use.

6.3 Governance implications

Together, the two sets of results bear on how external finance is governed. Finance that follows economic weight yet leaves no visible trace in regional accounts is difficult to evaluate from official data alone. This matters because the borrowing side of the ledger is also opaque: a substantial share of Chinese lending is not captured in debt statistics (Horn et al. 2021), Chinese projects have been related to local corruption and to weaker reform (Isaksson and Kotsadam 2018; Brazys and Vadlamannati 2021), and their consequences for the stability of recipient states differ from those of multilateral lenders (Gehring et al. 2022). An environment in which neither the cost nor the benefit is easily observed raises the value of independent, project-level monitoring and of subnational statistics that can be audited. The economics of governance literature on territorial allocation and local performance (Cerruti 2024; Rezki 2022; Mergo et al. 2023) offers a natural framework for such monitoring, because it links the political determinants of where resources go with the local outcomes that follow.

6.4 Limitations

Several limitations qualify the conclusions. First, the treatment is the first project, which bundles later projects with the first, and treatment is dated by commitment rather than disbursement. Second, the unit of analysis is the ADM1 region, which can obscure gains that diffuse within or across region boundaries. Third, power is limited: the MDE is 2.2 percent for the any-project definition and 3.6 percent for productive projects, and the post-treatment window of five years may be too short for benefits that arrive late. Fourth, the governance split is made at the country level on a time-invariant index, which leaves 18 or 19 countries in each group, so heterogeneity

estimates by governance have little power. Fifth, local political connections, which the favouritism literature finds important, are not observed, and the population gradient cannot be separated from political centrality. Sixth, reported GRP is subject to the measurement concerns discussed above. Seventh, sensitivity bounds are a conservative device rather than the fixed-length interval of Rambachan and Roth (2023), and the cohort analysis is exploratory.

7 Conclusion

This paper has asked where Chinese development finance lands inside recipient countries and whether it leaves a trace in reported regional output. Using geocoded projects and a harmonised panel of reported GRP for 1,103 regions in 48 countries, the answer to the first question is that finance follows economic weight: first projects go to richer and more populous regions, the tilt is strongest for non-concessional finance, and recent growth plays no role. The answer to the second question is that the first project does not produce a detectable increase in reported GRP per capita. The baseline estimate is slightly negative, is consistent with a pre-treatment drift that matched controls bring to zero, and is bounded away from sizeable gains in every specification. Project size, sector, financing terms and governance quality do not produce reliable heterogeneity, and the one exception, a decline among recent cohorts, does not survive changes of estimator.

These results suggest that debates about Chinese finance should pay at least as much attention to its selection of places and to the observability of its effects as to its average growth impact. For recipient governments, the findings imply that concentrating external finance in the largest region does not by itself generate gains that regional accounts can verify. For researchers, they call for designs that combine reported output with independent proxies, disbursement-based timing and political variables at the subnational level, so that economic weight can be separated from political centrality and real effects from recorded ones. The bounds reported here provide a benchmark against which such work can be judged.

Acknowledgements. This research is funded by University of Economics and Law, Ho Chi Minh City, Vietnam, and Vietnam National University, Ho Chi Minh City, Vietnam.

Declarations

Funding: This research is funded by University of Economics and Law, Ho Chi Minh City, Vietnam, and Vietnam National University, Ho Chi Minh City, Vietnam.

Competing interests: The author declares no competing interests.

Ethics approval and consent to participate: Not applicable. The study uses only publicly available secondary data on projects and regions and involves no human participants or personal data.

Consent for publication: Not applicable.

Data availability: The project data (GCDF v3.0 and Geo-GCDF v3.0) are publicly available from AidData (https://www.aiddata.org). The sub-national output data (DOSE V2.14) are publicly available from Zenodo (Wenz et al. 2023). Administrative boundaries (GADM version 3.6) are available from https://gadm.org and are not redistributed. The Worldwide Governance Indicators are available from the World Bank (Kaufmann et al. 2011). The analysis data set constructed from these sources is available from the author on reasonable request.

Materials availability: Not applicable.

Code availability: The analysis was run in Python in Kaggle notebooks. The code is available from the author on reasonable request.

Author contribution: Quoc Lap Nguyen is the sole author and was responsible for conceptualisation, data curation, methodology, software, formal analysis, visualisation, and writing (original draft, review and editing).

Appendix A Additional tables

This appendix reports the composition of the estimation sample by country (Table A1) and by cohort (Table A2), the definitions of the variables (Table A3), additional allocation results by period and in the cross-section (Tables A4 and A5), and the robustness and event-study results for the cohort analysis (Tables A6, A7 and A8).

Table A1: Estimation sample by country

Country (ISO3)	Regions	With project	Projects	Value (USD m)	Left-censored	Treated (T1)	In baseline	Share (%)
RUS	79	29	92	113,298	0	26	26	24.5
KAZ	14	11	79	36,999	2	9	9	8.0
PAK	4	4	191	35,657	4	0	0	7.7
BRA	27	16	76	33,577	0	15	15	7.2
IDN	33	29	164	31,487	1	28	26	6.8
IRN	31	19	69	24,506	5	14	13	5.3
VNM	63	50	104	21,298	16	30	29	4.6
LKA	9	9	137	14,311	1	8	0	3.1
ARG	24	14	52	11,676	0	14	14	2.5
UZB	14	14	65	11,303	5	9	8	2.4
KEN	47	34	94	10,289	4	30	9	2.2
BLR	6	6	82	9,721	1	5	0	2.1
ECU	24	21	84	8,915	4	15	15	1.9
TUR	81	40	44	8,769	0	38	38	1.9
PER	24	20	31	8,579	0	20	20	1.9
NGA	22	18	56	8,179	2	16	4	1.8
ETH	2	2	73	7,815	0	2	0	1.7
IND	33	10	26	6,878	1	9	9	1.5
PHL	17	17	126	6,697	2	15	14	1.4
ZAF	9	9	72	6,331	4	5	4	1.4
SRB	25	16	57	5,439	0	15	4	1.2
EGY	27	19	38	4,904	1	18	12	1.1
MYS	15	14	97	4,181	0	14	7	0.9
KGZ	9	9	54	3,820	2	7	4	0.8
BHS	3	3	24	3,607	0	3	0	0.8
MOZ	11	10	69	3,147	2	8	8	0.7
MNG	22	18	96	2,937	3	15	15	0.6
UKR	27	14	37	2,468	1	13	13	0.5

Continued on next page

Table A1: Estimation sample by country (continued)

Country (ISO3)	Regions	With project	Projects	Value (USD m)	Left-censored	Treated (T1)	In baseline	Share (%)
BOL	9	9	64	2,319	2	7	6	0.5
TZA	24	21	113	2,125	3	18	16	0.5
LAO	2	2	12	1,720	0	2	0	0.4
BIH	3	3	23	1,582	0	3	0	0.3
NPL	7	6	109	1,496	1	5	0	0.3
MKD	8	6	24	1,361	1	5	5	0.3
COL	32	29	45	1,154	0	20	20	0.2
CHL	15	8	12	871	0	8	8	0.2
THA	77	20	52	761	2	15	15	0.2
MAR	25	20	20	724	11	9	5	0.2
MEX	32	14	35	458	0	14	14	0.1
HND	18	1	1	336	0	1	0	0.1
GEO	11	9	25	322	4	5	4	0.1
ROU	42	8	19	252	1	7	7	0.1
URY	19	16	30	239	0	16	7	0.1
PAN	10	10	12	218	0	10	9	0.0
AZE	10	3	17	192	1	2	2	0.0
BGR	28	7	24	164	0	7	7	0.0
PRY	17	14	3	71	0	14	5	0.0
ALB	12	5	33	18	1	4	1	0.0
Total	1,103	686	2,862	463,169	88	573	447	100.0

Notes: Value is the sum of project commitments apportioned to regions by area overlap (USD million, constant 2021 prices). Share is the country share of total allocated value. Left-censored regions have a first project in 2002 or earlier; Treated (T1) counts regions with a first any-project in 2003 to 2019; In baseline counts treated regions entering the baseline any-project estimate (the event-year −1 sample of Table 9, which is not restricted to a post-period observation).

Table A2 Number of eligible treated regions by cohort

Year of first project	Any project	Productive	Social	Large
2003	25	6	18	5
2004	43	11	22	4
2005	51	15	22	7
2006	41	20	18	10
2007	32	13	18	13
2008	32	16	13	16
2009	37	21	17	18
2010	48	21	28	20
2011	29	15	12	13
2012	28	24	4	26
2013	18	15	3	22
2014	34	18	14	20
2015	51	27	22	15
2016	26	18	7	25
2017	33	26	6	19
2018	20	13	9	14
2019	25	15	11	17
Total	573	294	244	264

Notes: Counts of regions whose first qualifying project is committed in the given year, before the baseline-sample restrictions.

Table A3 Variable definitions and sources

Variable	Definition	Source
GRP per capita	Gross regional product per capita in constant 2015 USD; logs are natural logs	DOSE V2.14 (Wenz et al. 2023)
Sectoral GRP per capita	Agricultural, manufacturing and services GRP per capita in constant 2015 USD	DOSE V2.14
Population	Regional population	DOSE V2.14
Project commitment	Adjusted amount in constant 2021 USD, dated by commitment year	AidData GCDF v3.0 (Goodman et al. 2024)
Project footprint	Project polygons used for area-weighted allocation to ADM1 regions	AidData Geo-GCDF v3.0
Region boundaries	ADM1 boundaries used for the spatial overlay (DOSE regions matched to GADM version 3.6, level 1; not redistributed)	GADM 3.6
Treatment (T1)	First year with a project of status Completion or Implementation that is recommended for aggregate analysis; regions treated in 2002 or earlier are excluded	Author's construction
Productive/ infrastructure	First project in transport and storage, energy, communications, industry, mining and construction, or agriculture, forestry and fishing	Author's construction
Social/ humanitarian	First project in education, health, emergency response, other social infrastructure, government and civil society, food aid, disaster prevention, reconstruction relief or population policies	Author's construction
Large project	Project-region allocated value of USD 50 million or more	Author's construction
Relative size	Allocated value divided by the region's median real GRP over the panel	Author's construction
Initial relative income	Mean log GRP per capita over the first three observed years minus the country mean	Author's construction
Governance index	Mean of the standardised Worldwide Governance Indicators averaged over 2006 to 2008; countries are split at the sample median	WGI (Kaufmann et al. 2011); governance and fragility panel
Effect (%)	$100 \times (\exp(b) - 1)$, where b is the average event-year 0 to 4 difference-in-differences in log points	Author's construction

Notes: Source labels refer to the datasets used. The social sector definition follows the sector classification of the project data.

Table A4 Allocation over time: hazard of the first project by period

Period	Relative income	Log population	Past growth	Region-years
2003 to 2008	1.19^{***} (0.34)	2.72^{***} (0.49)	0.33 (0.47)	3,144
2009 to 2013	1.05^{**} (0.41)	4.14^{***} (0.73)	−0.41 (0.48)	2,670
2014 to 2019	0.60 (0.57)	2.91^{***} (0.87)	0.09 (0.75)	2,592
Difference 2009 to 2013 vs 2003 to 2008	−0.14 (0.55)	1.42^{*} (0.75)	−0.74 (0.80)	8,406
Difference 2014 to 2019 vs 2003 to 2008	−0.60 (0.65)	0.18 (0.84)	−0.24 (1.00)	8,406

Notes: Same specification as Table 6 for the event "first project of any type, estimated by sub-period. The last two rows are interaction coefficients from the pooled model (country-by-year fixed effects absorb the period main effects). Stars denote cluster-robust p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A5 Cross-sectional allocation: ever receiving a project, with country fixed effects

Dependent variable	Relative income	Log population	Log density	Regions	Countries	Mean dep. var.
Any project (0/100)	7.13^{***} (1.41)	18.24^{***} (1.65)	−1.90 (1.99)	1,068	45	63.01
Productive project (0/100)	6.37^{***} (2.02)	18.97^{***} (2.23)	−1.95 (1.82)	1,068	45	45.60
Social project (0/100)	6.70^{***} (1.11)	17.56^{***} (2.70)	−1.12 (2.33)	1,068	45	40.07
asinh(value per capita), any	0.44^{***} (0.12)	0.73^{***} (0.10)	−0.12 (0.09)	1,068	45	2.51

Notes: Linear regressions across regions with country fixed effects and country-clustered standard errors in parentheses; regressors are initial values standardized to unit variance. The first three dependent variables are $100 \times \mathbf{1}\{\text{region has at least one project of the stated kind in 2000 to 2019}\}$; the last is the inverse hyperbolic sine of allocated project value per head. The sample is regions in countries with at least three regions. Stars denote cluster-robust p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A6 Cohort heterogeneity: robustness of the 2014 to 2019 estimate, any project

Cohort	Effect (%)	95% CI	p (wild)	Regions	Countries
Baseline					
2003 to 2008	1.33	[−0.99, 3.70]	0.282	152	27
2009 to 2013	−1.50	[−4.29, 1.38]	0.293	127	31
2014 to 2019	-2.86^{***}	[−5.54, −0.10]	0.007	156	33
2014 to 2015 (full post-window)	-2.87^{**}	[−5.82, 0.18]	0.038	71	23
2016 to 2019 (truncated post-window)	-2.85^{**}	[−5.91, 0.32]	0.018	85	27
Matched controls (wide)					
2003 to 2008	2.08^{*}	[−0.08, 4.29]	0.064	132	26
2009 to 2013	−0.90	[−2.94, 1.19]	0.395	104	29
2014 to 2019	−0.86	[−3.07, 1.41]	0.384	130	29
Imputation estimator					
2003 to 2008	2.25	[−0.58, 5.16]	0.140	163	28
2009 to 2013	−1.99	[−5.65, 1.81]	0.291	130	31
2014 to 2019	−2.91	[−6.39, 0.69]	0.114	165	33
Linear pre-trend removed					
2003 to 2008	−0.37	[−2.70, 2.27]		152	27
2009 to 2013	−0.38	[−3.82, 2.96]		127	32
2014 to 2019	−3.23	[−6.93, −0.47]		156	33
Baseline, leave-one-country-out (range of point estimates)					
2014 to 2019		[−3.45, −1.86]		156	33

Notes: Effects are the average over event years 0 to 4 of region-level differences-in-differences in the log of the outcome relative to same-country controls, converted to percent as $100 \times (\exp(b) - 1)$. Confidence intervals are bootstrap-t intervals from 9,999 Webb-weight wild-cluster draws (clusters are countries); standard errors are cluster-robust (CR1) in log points times 100. Cohort heterogeneity was explored after the pooled analysis and should be read as exploratory. The imputation estimator fits region and country-by-year fixed effects on untreated observations. The linear-pre-trend row reports a bootstrap interval only, so no wild-bootstrap p-value is shown. Stars denote wild-cluster-bootstrap p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

References

Abadie, A., Athey, S., Imbens, G.W., Wooldridge, J.M.: When should you adjust standard errors for clustering? Quarterly Journal of Economics **138**(1), 1–35 (2023) https://doi.org/10.1093/qje/qjac038

Alesina, A., Dollar, D.: Who gives foreign aid to whom and why? Journal of Economic Growth **5**(1), 33–63 (2000) https://doi.org/10.1023/A:1009874203400

Abadie, A., Imbens, G.W.: Large sample properties of matching estimators for average treatment effects. Econometrica **74**(1), 235–267 (2006) https://doi.org/10.1111/j.1468-0262.2006.00655.x

Asher, S., Novosad, P.: Rural roads and local economic development. American Economic Review **110**(3), 797–823 (2020) https://doi.org/10.1257/aer.20180268

Anderson, M.L.: Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training projects. Journal of the American Statistical Association **103**(484), 1481–1495 (2008) https://doi.org/10.1198/016214508000000841

Table A7 Cohort heterogeneity: robustness of the 2014 to 2019 estimate, productive first project

Cohort	Effect (%)	95% CI	p (wild)	Regions	Countries
Baseline					
2003 to 2008	2.42	[−1.03, 5.99]	0.217	58	17
2009 to 2013	−2.36	[−6.18, 1.61]	0.221	76	27
2014 to 2019	−2.86	[−7.50, 2.02]	0.191	94	28
2014 to 2015 (full post-window)	−2.11	[−8.21, 4.40]	0.488	33	15
2016 to 2019 (truncated post-window)	−3.26*	[−7.96, 1.67]	0.055	61	23
Matched controls (wide)					
2003 to 2008	3.26*	[−0.86, 7.55]	0.084	51	15
2009 to 2013	−0.73	[−2.79, 1.38]	0.486	65	25
2014 to 2019	0.78	[−1.63, 3.24]	0.511	79	26
Imputation estimator					
2003 to 2008	3.27	[−1.10, 7.84]	0.196	62	17
2009 to 2013	−3.64	[−8.86, 1.87]	0.192	77	27
2014 to 2019	−2.23	[−7.02, 2.81]	0.398	103	28
Linear pre-trend removed					
2003 to 2008	−0.13	[−3.28, 3.44]		58	17
2009 to 2013	−1.25	[−6.44, 3.41]		76	28
2014 to 2019	−3.61	[−9.62, 0.64]		94	29

Notes: Effects are the average over event years 0 to 4 of region-level differences-in-differences in the log of the outcome relative to same-country controls, converted to percent as $100 \times (\exp(b) - 1)$. Confidence intervals are bootstrap-t intervals from 9,999 Webb-weight wild-cluster draws (clusters are countries); standard errors are cluster-robust (CR1) in log points times 100. Cohort heterogeneity was explored after the pooled analysis and should be read as exploratory. The imputation estimator fits region and country-by-year fixed effects on untreated observations. The linear-pre-trend row reports a bootstrap interval only, so no wild-bootstrap p-value is shown. Stars denote wild-cluster-bootstrap p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Aschauer, D.A.: Is public expenditure productive? Journal of Monetary Economics **23**(2), 177–200 (1989) https://doi.org/10.1016/0304-3932(89)90047-0

Burnside, C., Dollar, D.: Aid, policies, and growth. American Economic Review **90**(4), 847–868 (2000) https://doi.org/10.1257/aer.90.4.847

Bluhm, R., Dreher, A., Fuchs, A., Parks, B.C., Strange, A.M., Tierney, M.J.: Connective financing: Chinese infrastructure projects and the diffusion of economic activity in developing countries. Journal of Urban Economics **145**, 103730 (2025) https://doi.org/10.1016/j.jue.2024.103730

Bertrand, M., Duflo, E., Mullainathan, S.: How much should we trust differences-in-differences estimates? Quarterly Journal of Economics **119**(1), 249–275 (2004) https://doi.org/10.1162/003355304772839588

Banerjee, A., Duflo, E., Qian, N.: On the road: Access to transportation infrastructure and economic growth in China. Journal of Development Economics **145**, 102442 (2020) https://doi.org/10.1016/j.jdeveco.2020.102442

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., Miquel, G.: The value of democracy: Evidence from road building in Kenya. American Economic Review **105**(6), 1817–1851 (2015) https://doi.org/10.1257/aer.20131031

Table A8 Event studies by first-project cohort (any project)

	First project 2003 to 2008			First project 2014 to 2019		
Event year	Effect (%)	95% CI	Regions	Effect (%)	95% CI	Regions
−5	−1.27	[−3.89, 1.43]	95	−0.41	[−2.76, 1.99]	140
−4	−1.67	[−3.71, 0.42]	128	0.04	[−2.56, 2.70]	155
−3	−2.07***	[−3.23, −0.90]	148	−1.31	[−3.41, 0.84]	156
−2	−1.25**	[−2.50, 0.03]	150	−0.90	[−2.21, 0.43]	158
−1	0 (ref.)		152	0 (ref.)		164
0	0.14	[−1.58, 1.89]	150	−2.00*	[−4.18, 0.24]	156
1	1.32	[−1.47, 4.18]	147	−2.76*	[−5.79, 0.37]	126
2	1.48	[−1.22, 4.25]	149	−2.75	[−6.16, 0.78]	110
3	1.92	[−1.09, 5.01]	148	−3.78*	[−8.15, 0.81]	84
4	1.35	[−2.13, 4.94]	143	−5.83*	[−12.97, 1.90]	58
5	2.28	[−3.20, 8.08]	148	−8.57	[−22.16, 7.40]	20

Notes: Same construction as Table 9, restricted to the cohorts shown. Stars denote wild-cluster-bootstrap p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Borusyak, K., Jaravel, X., Spiess, J.: Revisiting event-study designs: Robust and efficient estimation. Review of Economic Studies **91**(6), 3253–3285 (2024) https://doi.org/10.1093/restud/rdae007

Baker, A.C., Larcker, D.F., Wang, C.C.Y.: How much should we trust staggered difference-in-differences estimates? Journal of Financial Economics **144**(2), 370–395 (2022) https://doi.org/10.1016/j.jfineco.2022.01.004

Brazys, S., Vadlamannati, K.C.: Aid curse with Chinese characteristics? Chinese development flows and economic reforms. Public Choice **188**(3), 407–430 (2021) https://doi.org/10.1007/s11127-020-00836-z

Cerruti, G.: Politics and city growth in Italy: Do connections matter? Economics of Governance **25**(4), 415–485 (2024) https://doi.org/10.1007/s10101-024-00319-7

Cameron, A.C., Gelbach, J.B., Miller, D.L.: Bootstrap-based improvements for inference with clustered errors. Review of Economics and Statistics **90**(3), 414–427 (2008) https://doi.org/10.1162/rest.90.3.414

Callaway, B., Sant'Anna, P.H.C.: Difference-in-differences with multiple time periods. Journal of Econometrics **225**(2), 200–230 (2021) https://doi.org/10.1016/j.jeconom.2020.12.001

Conley, T.G., Taber, C.R.: Inference with "difference in differences" with a small number of policy changes. Review of Economics and Statistics **93**(1), 113–125 (2011) https://doi.org/10.1162/REST_a_00049

Chaisemartin, C., D'Haultfœuille, X.: Two-way fixed effects estimators with heterogeneous treatment effects. American Economic Review **110**(9), 2964–2996 (2020) https://doi.org/10.1257/aer.20181169

Dreher, A., Fuchs, A.: Rogue aid? An empirical analysis of China's aid allocation. Canadian Journal of Economics **48**(3), 988–1023 (2015) https://doi.org/10.1111/caje.12166

Dreher, A., Fuchs, A., Hodler, R., Parks, B.C., Raschky, P.A., Tierney, M.J.: African leaders and the geography of China's foreign assistance. Journal of Development Economics **140**, 44–71 (2019) https://doi.org/10.1016/j.jdeveco.2019.04.003

Dreher, A., Fuchs, A., Hodler, R., Parks, B.C., Raschky, P.A., Tierney, M.J.: Is favoritism a threat to Chinese aid effectiveness? A subnational analysis of Chinese development projects. World Development **139**, 105291 (2021) https://doi.org/10.1016/j.worlddev.2020.105291

Dreher, A., Fuchs, A., Parks, B., Strange, A., Tierney, M.J.: Aid, China, and growth: Evidence from a new global development finance dataset. American Economic Journal: Economic Policy **13**(2), 135–174 (2021) https://doi.org/10.1257/

pol.20180631

Dreher, A., Lohmann, S.: Aid and growth at the regional level. Oxford Review of Economic Policy **31**(3–4), 420–446 (2015) https://doi.org/10.1093/oxrep/grv026

Faber, B.: Trade integration, market size, and industrialization: Evidence from China's National Trunk Highway System. Review of Economic Studies **81**(3), 1046–1070 (2014) https://doi.org/10.1093/restud/rdu010

Freyaldenhoven, S., Hansen, C., Shapiro, J.M.: Pre-event trends in the panel event-study design. American Economic Review **109**(9), 3307–3338 (2019) https://doi.org/10.1257/aer.20180609

Franck, R., Rainer, I.: Does the leader's ethnicity matter? Ethnic favoritism, education, and health in sub-Saharan Africa. American Political Science Review **106**(2), 294–325 (2012) https://doi.org/10.1017/S0003055412000172

Goodman-Bacon, A.: Difference-in-differences with variation in treatment timing. Journal of Econometrics **225**(2), 254–277 (2021) https://doi.org/10.1016/j.jeconom.2021.03.014

Gehring, K., Kaplan, L.C., Wong, M.H.L.: China and the World Bank: How contrasting development approaches affect the stability of African states. Journal of Development Economics **158**, 102902 (2022) https://doi.org/10.1016/j.jdeveco.2022.102902

Gennaioli, N., La Porta, R., Silanes, F., Shleifer, A.: Growth in regions. Journal of Economic Growth **19**(3), 259–309 (2014) https://doi.org/10.1007/s10887-014-9105-9

Gibson, J., Olivia, S., Boe-Gibson, G., Li, C.: Which night lights data should we use in economics, and where? Journal of Development Economics **149**, 102602 (2021) https://doi.org/10.1016/j.jdeveco.2020.102602

Goodman, S., Zhang, S., Malik, A.A., Parks, B.C., Hall, J.: AidData's geospatial global Chinese development finance dataset. Scientific Data **11**, 529 (2024) https://doi.org/10.1038/s41597-024-03341-w

Hartman, E., Hidalgo, F.D.: An equivalence approach to balance and placebo tests. American Journal of Political Science **62**(4), 1000–1013 (2018) https://doi.org/10.1111/ajps.12387

Hankins, W., Hoover, G., Pecorino, P.: Party polarization, political alignment, and federal grant spending at the state level. Economics of Governance **18**(4), 351–389 (2017) https://doi.org/10.1007/s10101-017-0196-6

Holm, S.: A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics **6**(2), 65–70 (1979)

Hodler, R., Raschky, P.A.: Regional favoritism. Quarterly Journal of Economics **129**(2), 995–1033 (2014) https://doi.org/10.1093/qje/qju004

Horn, S., Reinhart, C.M., Trebesch, C.: China's overseas lending. Journal of International Economics **133**, 103539 (2021) https://doi.org/10.1016/j.jinteco.2021.103539

Henderson, J.V., Storeygard, A., Weil, D.N.: Measuring economic growth from outer space. American Economic Review **102**(2), 994–1028 (2012) https://doi.org/10.1257/aer.102.2.994

Isaksson, A.-S., Kotsadam, A.: Chinese aid and local corruption. Journal of Public Economics **159**, 146–159 (2018) https://doi.org/10.1016/j.jpubeco.2018.01.002

Kaufmann, D., Kraay, A., Mastruzzi, M.: The Worldwide Governance Indicators: Methodology and analytical issues. Hague Journal on the Rule of Law **3**(2), 220–246 (2011) https://doi.org/10.1017/S1876404511200046

Lessmann, C., Seidel, A.: Regional inequality, convergence, and its determinants – A view from outer space. European Economic Review **92**, 110–132 (2017) https://doi.org/10.1016/j.euroecorev.2016.11.009

Martínez, L.R.: How much should we trust the dictator's GDP growth estimates? Journal of Political Economy **130**(10), 2731–2769 (2022) https://doi.org/10.1086/720458

Melnik, W.: Legislative redistricting and the partisan distribution of transportation expenditure. Economics of Governance **25**(1), 1–29 (2024) https://doi.org/10.1007/s10101-024-00308-w

Mergo, T., Nimubona, A.-D., Rus, H.A.: Ethnically asymmetric political representation and the provision of public goods: Theory and evidence from Ethiopia. Economics of Governance **24**(4), 519–547 (2023) https://doi.org/10.1007/s10101-023-00302-8

Michalski, T., Stoltz, G.: Do countries falsify economic data strategically? Some evidence that they might. Review of Economics and Statistics **95**(2), 591–616 (2013) https://doi.org/10.1162/REST_a_00274

MacKinnon, J.G., Webb, M.D.: Wild bootstrap inference for wildly different cluster sizes. Journal of Applied Econometrics **32**(2), 233–254 (2017) https://doi.org/10.1002/jae.2508

Rezki, J.F.: Political competition and economic performance: Evidence from Indonesia. Economics of Governance **23**(2), 83–114 (2022) https://doi.org/10.1007/s10101-022-00272-3

Roth, J.: Pretest with caution: Event-study estimates after testing for parallel trends.

American Economic Review: Insights **4**(3), 305–322 (2022) https://doi.org/10.1257/aeri.20210236

Rambachan, A., Roth, J.: A more credible approach to parallel trends. Review of Economic Studies **90**(5), 2555–2591 (2023) https://doi.org/10.1093/restud/rdad018

Rajan, R.G., Subramanian, A.: Aid and growth: What does the cross-country evidence really show? Review of Economics and Statistics **90**(4), 643–665 (2008) https://doi.org/10.1162/rest.90.4.643

Roth, J., Sant'Anna, P.H.C., Bilinski, A., Poe, J.: What's trending in difference-in-differences? A synthesis of the recent econometrics literature. Journal of Econometrics **235**(2), 2218–2244 (2023) https://doi.org/10.1016/j.jeconom.2023.03.008

Sun, L., Abraham, S.: Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics **225**(2), 175–199 (2021) https://doi.org/10.1016/j.jeconom.2020.09.006

Strange, A.M., Dreher, A., Fuchs, A., Parks, B., Tierney, M.J.: Tracking underreported financial flows: China's development finance and the aid–conflict nexus revisited. Journal of Conflict Resolution **61**(5), 935–963 (2017) https://doi.org/10.1177/0022002715604363

Wenz, L., Carr, R.D., Kögel, N., Kotz, M., Kalkuhl, M.: DOSE – Global data set of reported sub-national economic output. Scientific Data **10**, 425 (2023) https://doi.org/10.1038/s41597-023-02323-8

Webb, M.D.: Reworking wild bootstrap-based inference for clustered errors. Canadian Journal of Economics **56**(3), 839–858 (2023) https://doi.org/10.1111/caje.12661

Young, A.: Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. Quarterly Journal of Economics **134**(2), 557–598 (2019) https://doi.org/10.1093/qje/qjy029