Did China's Poor-County Designation Move Counties Across a Development Threshold?

Weifeng Zhao

Abstract Area-based antipoverty policy rests on the premise that a transfer large enough can change a poor place's trajectory and not merely its level. This proposal tests that premise using China's national poor-county programme, which still covered 832 counties when it was abolished in 2020. Existing evaluations measure its effects while transfers were still flowing, and so cannot say whether the programme temporarily enlarged the capital stock of economies that will return to their original path or pushed them past an unstable threshold in an S-shaped accumulation function; the two separate only once support ends. Identification therefore comes from exit rather than entry: counties graduated in cohorts between 2016 and 2020, and outcomes are analysed in event time around each county's own graduation year. Three hypotheses—continued divergence after graduation (threshold crossed), a gap that decays toward zero (level effect only), and no gap at any horizon—are tested with a staggered event study using the Callaway and Sant'Anna (2021) estimator, a border discontinuity across adjacent designated and non-designated counties, and a secondary regression discontinuity at the 2012 income cutoff. Development is measured by two outcomes with different error sources: nighttime luminosity from DMSP-OLS and VIIRS, produced entirely outside the Chinese administrative system, and reported county income per capita, which enters as an outcome read against luminosity and never as a control, so that agreement is informative and divergence around certification is diagnostic of reporting behaviour. Persistence after graduation is read as a threshold only if it survives checks for successor support, migration, spillovers and endogenous exit timing, and is accompanied by non-linearity in the cumulative push and bimodality across counties. The answer determines whether a one-off push can permanently change a place, or whether area- based support must be permanent to have a permanent effect.

Keywords poverty traps; big push; place-based antipoverty policy; national poor counties; China; nighttime luminosity; staggered event study; border discontinuity design

1. Introduction

The premise of area-based antipoverty policy is that a transfer large enough can change the trajectory of a poor place and not merely its level. That premise remains largely untested. China's national poor-county programme, which still covered 832 counties when it ended in 2020, was the largest test of it ever attempted. Existing evaluations agree on one point: while the money was flowing, designation raised incomes. What they cannot establish is why. A transfer can raise output for two reasons. It can enlarge the capital stock of an economy that will return to its original path, or it can push that economy past a tipping point beyond which growth is self-sustaining. While transfers continue, the two mechanisms look identical.

They separate only when the transfers stop. This proposal therefore asks whether poor-county designation moved counties across a development threshold, or produced a level effect that decays

once support ends. The question is the empirical counterpart of a theoretical one: is the accumulation function facing a poor Chinese county unique and stable, as in Solow (1956), or is it S-shaped, with an unstable threshold separating a low from a high steady state? The two answers imply different policies. If thresholds exist, a one-off push can permanently change a place, and the size and timing of exit become the central design problems. If they do not, area-based support must be permanent in order to have a permanent effect.

The identifying variation comes from exit rather than from entry. From 2016 to 2020, groups of counties exited the designation list, and the last nine graduated on 23 November 2020 when the designation system was abolished altogether. So each county has a different graduation year, and the analysis is structured in event time around that year instead of across the calendar. What happens after

a county's own exit is informative about the structure of its local economy rather than about the size of the intervention, although persistence after exit is consistent with more than one mechanism (Section 4).

Two outcome variables measure development. The first is nighttime luminosity from satellites, which is entirely outside the Chinese administration's control. The second is income per capita: the variable that this programme was designed to shift and the variable against which graduation was certified. The design is one in which income enters as an outcome and never as a covariate: conditioning on a variable that the treatment changes would soak up part of the estimated effect, the usual bad-control problem (Angrist and Pischke 2009). The report does have an incentive problem, as the same administrative hierarchy that reports county income also designated and certified exit against some quantitative targets (Wallace 2016; Martínez 2022). That is a reason to read the two outcomes off one another, rather than outright reject one. The implication, where luminosity and reported income move together, would be stronger than either series could support alone. When they diverge—especially when reported income rises going into the year of graduation while luminosity does not—the divergence situates the reporting problem temporally and is itself a result. Motivation is presented in Section 2 and the literature in Section 3. Section 4 only conceptualizes and formulates the three hypotheses (summarised in Table 3). Section 5 goes over the data with each variable's role delineated. Our empirical strategy is described in Section 6 that presents a sharp border discontinuity, a staggered event study in event time, followed by a second regression discontinuity at the income cutoff. The contribution and limitations of this work are described in Section 7, while we conclude the paper in Section 8. Section 9 briefly states how AI tools were used; the full statement is in Appendix B. Table 1 lists the data sources and Appendix A gives an access point for each.

2. Motivation

China ran the largest area-based antipoverty programme in history between 1986 and 2020. In four rounds — 1986, 1994, 2001 and 2012 — the State Council designated "national poor counties" using a rural per capita net income threshold specific to that round, relaxed for areas with large ethnic-minority populations and for old revolutionary base areas. The 2012 round used the national poverty line of 2,300 yuan at 2010 constant prices. Designated counties received targeted investment, subsidised credit, food-for-work schemes and tax relief. Eight hundred and thirty-two counties were still on the list when the last of them graduated on 23 November 2020 and the designation system was abolished.

The existing evaluations are hard to reconcile. Using a regression discontinuity design around the income cutoff of the 1994 Seven-Year Priority Poverty Alleviation Program, Meng (2013) estimates that designation raised rural income substantially over 1994–2000. A follow-up survey of households four years after disbursements had finished for a large poor-area programme in south- west China shows that most of the significant short-run income gains were spent on saving, leaving only small and statistically unimportant long-run impacts on average consumption (Chen et al., 2009).

These two results illustrate the distinction between the two theories. With a single stable steady state, a temporary transfer raises capital per worker only while it lasts, after which the county returns to its original path: a level effect that decays. By contrast, if the accumulation function is S-shaped, a sufficiently large push can move a county above an unstable threshold and generate self- sustaining growth after transfers end. All available estimates of the average effect of the Chinese programme were calculated while the money was still flowing. None checks which of the two structures created it.

Although graduation did not always eliminate all government support, the staggered timing of exit across counties makes the test possible. At the same time, the path after graduation tells you about structure, not scale. Two measurement choices follow. The first measure of local development is satellite luminosity, which county governments do not supply. Second, since we treat reported income per capita as a separate outcome rather than discarding it, the design can indicate whether and to what extent the two measures tell the same story—something that is important for both the substantive conclusion and the evaluation of how much weight should be placed on estimates originating in administrative series used in the previous literature.

3. Previous literature

Traps and the big push. Solow (1956) has a unique interior steady state, so initial conditions do not bind the long run. Adding subsistence consumption makes the accumulation function S-shaped and generates multiple steady states, the formal counterpart of a one-off "big push" (Sachs 2005). The empirical record is contested. Kraay and McKenzie (2014) survey it and conclude that geographic poverty traps are hard to document. Balboni et al. (2022) reach a different conclusion: in an eleven- year panel following a randomised asset transfer in Bangladesh they find a sharp threshold, with households pushed above a critical asset level moving into higher-productivity occupations and staying there while those left below revert. Their threshold, however, is defined over household assets rather than over places. At the level of places, Kline and Moretti (2014) show that the Tennessee Valley Authority left effects on local manufacturing that outlived the programme by decades — consistent with a place-level threshold, though not identifying one. Whether a whole

ISSN 3106-5767 (Print) ISSN 3106-7026 (Online)

area, rather than a household, can be pushed across a threshold is therefore still open.

Institutions and persistence. Institutions established at one time can shape economic outcomes many decades later (Acemoglu, Johnson and Robinson 2001; Banerjee and Iyer 2005). Banerjee and Iyer supply the design this proposal borrows: they compare adjacent Indian districts assigned different colonial land-revenue systems. Neighbouring districts share geography, climate and
market access but were separated by an administrative line, so a discontinuity across that line can be attributed to the institution rather than to the terrain.

Measurement. Using nighttime luminosity as a spatial proxy for local economic activity, Henderson, Storeygard, and Weil (2012) show that nighttime lights perform well at aggregate or regional scales, while Gibson, Olivia and Boe-Gibson (2020) document weaker performance of the older DMSP series for dim rural emitters. Two additional studies motivate comparing reported income with the satellite measure. Martínez (2022) shows that the elasticity of reported GDP to nighttime lights is systematically larger in authoritarian regimes, making a lights-versus-reports comparison informative about reporting. Wallace (2016) documents how Chinese subnational economic statistics respond to political incentives at politically sensitive moments. Poor-county graduation is one such moment.

Evaluations of China's programme. Park, Wang and Wu (2002) document which counties were targeted and how much leakage the allocation rule permitted. Meng (2013) and Lü (2015) evaluate the 8–7 Plan, Park and Wang (2010) evaluate the successor village-level programme, and Chen, Mu and Ravallion (2009) address persistence most directly. Ravallion and Chen (2007) provide the national aggregate benchmark against which county-level results can be read.

Gap. Four gaps motivate this proposal. First, no existing study frames the programme as a contest between steady-state structures; the policy question (did it work?) has been kept separate from the theoretical one (was there a threshold?). Second, none uses the 2016–2020 graduation window, which is the only period in which the two hypotheses generate opposing predictions. Third, the literature measures outcomes with administrative income series that are themselves inputs into the treatment and graduation decisions being evaluated. Fourth, where income appears in this literature it is typically either the sole outcome or a control variable; it is not read against an independently produced measure of the same underlying activity.

4. Conceptual framework and testable predictions

Write per-worker capital accumulation as

$$k_{t+1} = [(1-\delta)/(1+n)]\, k_t + [s(k_t)/(1+n)]\, f(k_t) \equiv g(k_t).$$

In this case, an S-shaped g(k) can cross the 45-degree line three times: a low stable steady state, an unstable threshold with s(k) increasing in k over the poor range (if subsistence needs imply that saving at very low incomes is negative), and a high stable steady state. What matters here is the scenario shown in Fig. 1. Treating assignment as a temporary augmentation to expenditure that disappears when a county graduates in year G produces three hypotheses.

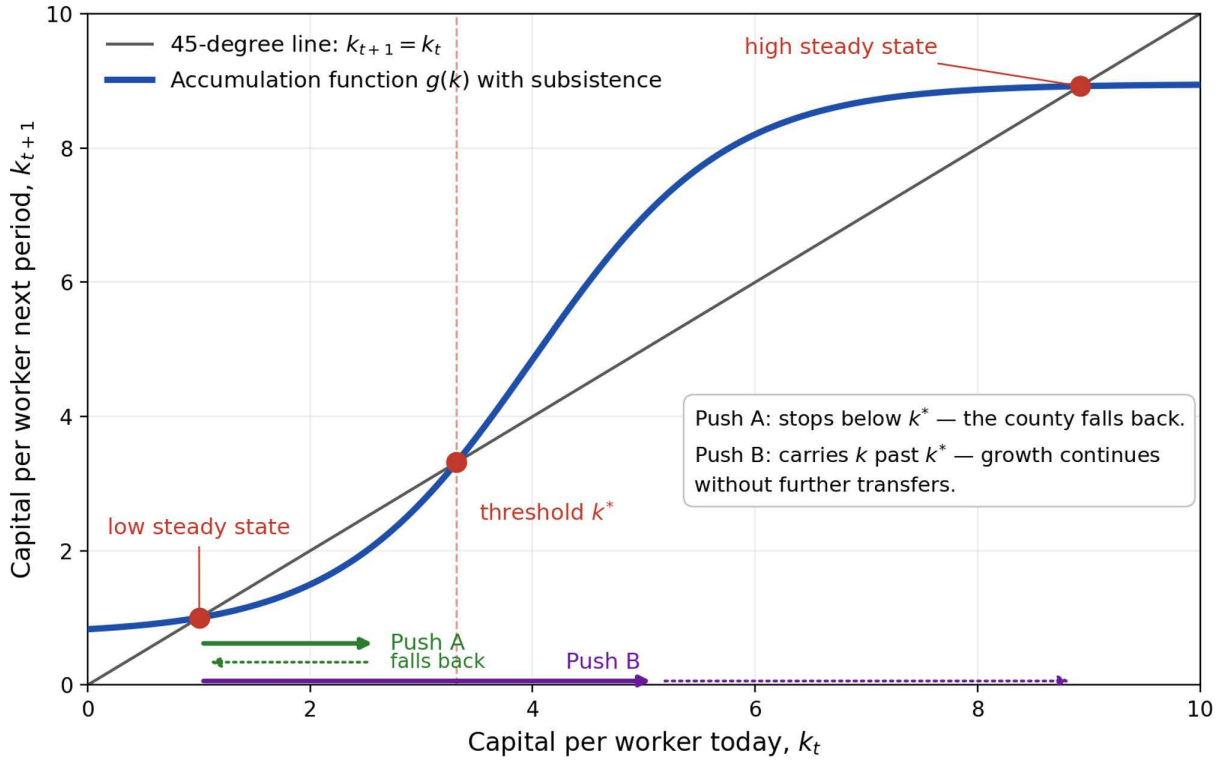


Figure 1. Conceptual illustration (theoretical diagram; no data). Schematic drawn by the author; not reproduced from any published source. An S-shaped accumulation function g(k) crosses the 45- degree line three times, generating a low stable steady state, an unstable threshold k*, and a high stable steady state. Push A leaves the county below the threshold, so it falls back; Push B carries it past the threshold, after which growth continues without further transfers.

H1 (threshold crossed). Graduated counties retain a persistent positive gap relative to comparable non-designated counties after their own graduation year; the gap is flat or rising after $\tau = 0$. This pattern is consistent with, but not sufficient to establish, a crossed threshold.

H2 (level effect only). The gap is positive while transfers are flowing and decays toward zero after
$\tau = 0$.

H3 (no effect, or offsetting reallocation). There is no gap at any horizon, either because provincial and county governments reallocated their own spending away from designated counties, or because transfers were saved rather than invested locally.

The three signatures appear in Figure 2(b) and are summarised, with the test that corresponds to each, in Table 3. The essential point is that H1 and H2 differ in the sign of the slope after graduation, which can discriminate between the two structures under the assumptions stated below, without requiring the threshold to be recovered structurally. Two auxiliary implications sharpen the test: under H1 the effect should be non-linear in the size of the cumulative push relative to baseline capital, and the cross-county distribution of the outcome should become bimodal; under H2 it should be smooth and unimodal.

Hypotheses are formulated for both outcomes. Under both H1 and H2, luminosity per capita and income per capita should show a similar qualitative pattern because they are indicators of the same local economy. A third pattern is specific to the two outcomes: reported income might increase into the graduation year and flatten immediately afterwards, while luminosity shows no corresponding movement. This pattern would not provide evidence for either structure. Instead, it would suggest that the income series responds to the certification process and should be reported as such rather than interpreted as a policy effect.

Identifying assumptions and alternative explanations. A flat or rising post-graduation path is consistent with a crossed threshold but does not uniquely identify one. The same pattern could arise from long-lived infrastructure built during the designation period, continued fiscal support under successor programmes, industrial relocation into former poor counties, population change that alters per-capita outcomes, or spillovers from neighbouring counties. H1 is therefore read as evidence of a threshold only if persistence holds in the sample without successor designation (threat iii), in totals as well as per capita (threat ii), against comparison counties that do not border graduated counties, and together with the non-linearity and bimodality described under Falsification. Otherwise, persistence is reported as a lasting effect of the programme, not as evidence of a poverty trap.

ISSN 3106-5767 (Print) ISSN 3106-7026 (Online)

Table 3. Theoretical predictions and corresponding empirical tests

Hypothesis	Before $\tau = 0$	After $\tau = 0$	Auxiliary predictions	Empirical test
H1: threshold crossed	Positive gap while transfers flow	Gap flat or rising ($\theta\tau$ non-decreasing)	Effect non-linear in cumulative push; bimodal cross-county distribution	Design 2 on restricted sample; Falsification tests
H2: level effect only	Positive gap while transfers flow	Gap decays toward zero	Effect linear; unimodal distribution	Design 2; Design 1 as cross-check
H3: no effect or offsetting reallocation	No gap	No gap	Non-linearity and bimodality separate no trap from reallocation	Designs 1 and 2; Falsification tests
Reporting response (not a structural hypothesis)	Reported income rises into $\tau = 0$; luminosity does not	Income flattens	Lights–income elasticity shifts at $\tau = 0$; bunching above certification line	Threat (iv)

Note: predictions apply to both outcomes; the gap is measured against comparison counties in event time (Section 6).

5. Data

Table 1 lists the sources and Appendix A gives an access point for each. Two principles govern selection. First, the primary outcome and all validation data are produced outside the Chinese administrative system — by U.S. federal agencies, U.S. national laboratories or U.S. research universities — so that development is measured by a process no county government controls. Second, any variable that can only be obtained from Chinese administrative records is either used for treatment assignment or reported as an outcome in its own right with the incentive problem stated, never smuggled in as a control.

5.1 The role of each variable

Before any estimation, Table 2 sets every variable's role. A key feature is classifying variables as pre- and post-treatment. Before designation status could have influenced population, rise or fall through the particular terrain and demographic conditions of a county in 1990 makes it possible to serve as controls. The policy should plausibly move income, consumption, and employment contemporaneously with population (which makes these covariates inappropriate), so none of them is used as a control; income per capita appears as Outcome 2, and CHNS consumption is used as an

additional validation measure. It becomes a borderline case when contemporary gridded population is needed as the denominator of a per-capita outcome.

Table 2. Role of each variable in the design

Role	Variable(s)	Source and note
Outcome 1	Log nighttime luminosity per capita; share of lit pixels; lit area	DMSP-OLS and VIIRS. Produced outside the Chinese administrative system.
Outcome 2	Log county income per capita (rural per capita net / disposable income)	County statistical yearbooks. Administrative in origin; read against Outcome 1, not used as a control.
Treatment	Designation status Dc; own graduation year Gc (2016–2020)	State Council lists and graduation notices. Gc defines event time $\tau = t - Gc$.
Controls (pre- treatment)	Elevation, slope, ruggedness, distance to prefecture seat; 1990 population, education, ethnic composition, urban share	SRTM (NASA/USGS); CITAS 1990 county census, archived at NASA SEDAC. Fixed before treatment.
Denominator	Gridded population, 2000–	LandScan (ORNL), GPWv4.11 (CIESIN/NASA). Reported alongside totals because migration is a response.
Validation	CHNS household income, consumption, education	Carolina Population Center, UNC. Out-of-sample check on both outcomes; not a control.
Deliberately excluded as controls	Post-treatment income, consumption, employment, fiscal transfers, contemporaneous population	All are affected by the treatment. Conditioning on them would absorb the effect (Angrist and Pischke 2009).

5.2 Outcome 1: nighttime luminosity

The pre-2013 series is the DMSP-OLS Version 4 stable-lights composite at 30 arc-seconds, covering 1992–2013. It is top-coded at digital number 63 and has no onboard calibration, so inter- satellite calibration following Hsu et al. (2015) is required. From 2012 onward the series is the VIIRS Day/Night Band annual composite (VNL V2.2), described in Elvidge et al. (2017), which is radiometrically calibrated and far more sensitive at low radiance. That sensitivity is decisive here, because rural poor counties sit close to the DMSP detection floor. Both products are generated outside China: the historical DMSP archive remains with NOAA's National Centers for Environmental Information, while the current VIIRS series is produced by the Earth Observation Group at the Payne Institute, Colorado School of Mines, which is the correct attribution for recent years. The main outcome is log radiance per capita; the share of lit pixels and total lit area are reported as secondary outcomes, since both behave better than mean radiance for dim emitters (Gibson, Olivia and Boe-Gibson 2020).

5.3 Outcome 2: income per capita

The second dependent variable is county per capita income: rural net income per person (through 2013) and rural household disposable income afterwards, deflated by the provincial average rate of living index based on rural consumers. The series is collected in the China County Statistical Yearbook and in provincial yearbooks and is available from China Data Online, licensed through
U.S. university libraries. Three points are stated plainly. First, the distribution of a U.S. system does not remove it from the Chinese administrative structure; it is birthed there, and this is the same family of series used to score designation and graduation. Second, that is why it is retained as a second outcome to be compared with luminosity, rather than being used as a control variable. Third, we absorb a break in definition around 2013 and a redefinition around rural income sampling using province-by-year fixed effects and by working with logs and differences rather than levels. 1990 and 1991 baselines are derived from the CITAS county-level economic yearbook extract archived at NASA SEDAC, anchoring the pre-period to a source held within U.S. archives.

Combining both outcomes allows for a diagnosis that neither could provide alone. Following the rationale suggested by Martínez (2022), we can estimate separate elasticities for income reported to luminosity both before and after graduation. A stable elasticity suggests reading the income results as common meaning; a jump in the certification window points to reporting behaviour, and here too the luminosity estimates do substantive work while the income estimates testify to features of the reporting process.

5.4 Treatment: designation and graduation year

Treatment has two components. Designation status comes from the State Council poor-county lists for 1986, 1994, 2001 and 2012. The graduation year is the year in which a county was certified to have exited, drawn from the 2016–2020 provincial and State Council announcements as compiled in the published literature. Both are administrative facts — a county either was or was not on a list in a given year — rather than statistics, which is why they are the only Chinese official inputs used for identification.

5.5 Geographic units and boundary changes

During the period of study, county boundaries were far from static. As prefecture-level cities expanded, counties were merged, split off and transformed into urban districts; about 1 in every 10 county-level unit underwent some administrative change during the decade. That means geography can create an "effect" by itself unless units are kept constant. Three rules address this. (i) All outcome measures are aggregated to the 1990 county polygons—the CITAS 1:1M county shape files (archived by NASA SEDAC)—physically measuring identical areas every year despite their changing names. (ii) A concordance from post-1990 units to 1990 units is constructed based on the annual division-code series; multi-unit mappings originating in a post-1990 unit are merged into single larger super-units instead of splitting (per assumption). (iii) Exclude counties for which we cannot reproduce the area defined by the 1990 Census, or for which the reconstructed area is changed by more than five per cent (two and ten per cent thresholds also reported as robustness checks). The conversion to urban district is not a neutral form of attrition either: it applies to counties that had grown, thus the drop removes some from the upper tail and biases against H1. We therefore report the number of counties which dropped for each reason, and re-estimate our main specifications retaining only those converted counties at their 1990 extent to allow the direction and size of the resulting difference to be visualised.

5.6 Population, terrain and baseline controls

The per-capita denominator is the LandScan Global annual population grid (Oak Ridge National Laboratory), with GPWv4.11 (CIESIN, Columbia University, distributed by NASA) as a robustness check. The two are built differently, and the difference matters: LandScan uses nighttime lights as an ancillary input to its dasymetric allocation, creating a mechanical link between the numerator and the denominator of the main outcome, whereas GPW allocates from census counts alone and so gives a cleaner denominator at coarser temporal resolution. Results are reported on both, and always alongside total luminosity, so that no conclusion rests on the denominator. Both series begin in 2000, so the 1992–1999 luminosity years are analysed in totals rather than per capita. Terrain covariates — elevation, slope and ruggedness — come from SRTM (NASA/USGS). Pre-treatment demographic controls come from the CITAS county-level 1990 census extract

ISSN 3106-5767 (Print) ISSN 3106-7026 (Online)

archived at NASA SEDAC, which provides population, age structure, educational attainment, ethnic composition and urban residency at the county level; this is preferred to IPUMS-International, whose harmonised geography for China does not go below the province.

5.7 Household validation

The China Health and Nutrition Survey (1989–2015), run by the Carolina Population Center at the University of North Carolina at Chapel Hill with the Chinese Center for Disease Control and Prevention and funded by the U.S. National Institutes of Health, covers provinces including Guizhou and Guangxi where designated counties are concentrated. Its income, consumption and education modules validate both outcomes at the household level. Two limits are stated in advance. First, county identifiers are suppressed in public-use CHNS files, so designation status cannot be linked at that level without a restricted geographic file—requiring an application to the Carolina Population Center—or prefecture-level data; however, the core design does not rely on this. Second, the number of sampled households in any given county-year is small, so year-to-year variation in CHNS means is largely just sampling noise. As such, one cannot interpret single-year changes as effects; hence the CHNS evidence can only be applied in pooled multi-year windows—three-year averages before and after graduation.

5.8 National context and external benchmark

National context comes from the USDA Economic Research Service International Macroeconomic Data Set and from the Penn World Table, version 11.0 (Feenstra, Inklaar and Timmer 2015). Two caveats attach to the latter: it is co-produced with the University of Groningen, so it is not solely a U.S. source, and version 11.0 switched its China series to official national accounts. It is used only as aggregate context and never as an outcome. Finally, the ERS classification of persistent-poverty U.S. counties — those with poverty rates of twenty per cent or more in the 1990 and 2000 censuses and in the 2007–11 and 2017–21 American Community Survey five-year estimates — provides an external reference point for how persistent area-level poverty is in the absence of a comparably large push.

Table 1. Data sources

Variable	Source (producer)	Coverage / unit	Role
Nighttime luminosity, 1992–2013	DMSP-OLS v4 stable lights, NOAA National Centers for Environmental Information	Annual, 30 arc-second grid, global	Outcome 1 (pre-period)
Nighttime luminosity, 2012–present	VIIRS annual composites (VNL V2.2), Earth Observation Group, Colorado School of Mines	Annual, 500 m grid, radiance-calibrated	Outcome 1 (post-period)
County income per capita	China County Statistical Yearbook and provincial yearbooks, via China Data Online (U.S. library licence)	Annual, county level	Outcome 2; administrative origin declared
Gridded population	LandScan Global (Oak Ridge National Laboratory); GPWv4.11 (CIESIN, Columbia University / NASA)	Annual (LandScan) / 5-yearly (GPW); both from 2000	Per-capita denominator; migration check
Terrain and ruggedness	SRTM elevation (NASA / USGS)	30 m and 90 m grids	Pre-treatment control
County boundaries, 1990	**CITAS 1:1M county map (University of Washington / CIESIN), archived at NASA SEDAC; GADM v4.1 for modern boundaries**	**County polygons, 31 Dec 1990**	**Fixed geographic unit; boundary- change screening**
Baseline demographics, 1990	CITAS county-level 1990 census and economic yearbook extracts, NASA SEDAC	County level, 1990–1991	Pre-treatment controls; income baseline
Household income, consumption, education	China Health and Nutrition Survey (Carolina Population Center, UNC Chapel Hill, with China CDC; NIH-funded)	10 waves, 1989–2015, 9–15 provinces	Validation of both outcomes

County boundaries, 1990	**CITAS 1:1M county map (University of Washington / CIESIN), archived at NASA SEDAC; GADM v4.1 for modern boundaries**	**County polygons, 31 Dec 1990**	**Fixed geographic unit; boundary- change screening**
National macro context	International Macroeconomic Data Set (USDA Economic Research Service); Penn World Table 11.0	Annual, country level	Aggregate benchmarking
U.S. comparison case	Persistent-poverty counties (USDA Economic Research Service)	County level, 1990–2021	External benchmark for area-based persistence
Designation and graduation status	State Council poor-county lists (1986, 1994, 2001, 2012) and 2016–2020 graduation notices, as reproduced in the published literature	832 counties; graduation year per county	Treatment assignment and event time

Note: Outcome 1 and all validation data are produced outside the Chinese administrative system. Designation status is an administrative fact rather than a statistic. Outcome 2 originates inside that system and is reported as such; Section 5.3 explains why it is retained.6. Empirical strategy

Design 1: border discontinuity (primary). Following Banerjee and Iyer (2005), compare outcomes in grid cells lying within a narrow band of the borders that separate adjacent designated and non- designated counties:
$yict = \alpha pt + \Sigma k\, \beta k\, (Dc \times 1[t = k]) + f(disti) + \gamma' Xi + \varepsilon ict$,
where αpt is a border-pair-by-year fixed effect, Dc is designation, disti is the signed distance from cell i to the boundary, and Xi collects elevation, slope, ruggedness and distance to the nearest prefecture seat. Identification requires that terrain and market access vary continuously across the border while policy does not, which is testable: covariates and pre-1994 luminosity should be continuous at the boundary. The estimation sample is restricted to cells within 5 km of a designated / non-designated boundary, standard errors are clustered by border pair, and sensitivity to spillovers is checked on narrower and wider bands.
Because administrative boundaries in China may themselves reflect historical, political or economic differences, the validity of this design is checked rather than assumed. First, balance tests re- estimate the equation with each predetermined characteristic as the dependent variable: terrain, distance to the prefecture seat and pre-1994 luminosity at the cell level, and 1990 census population, education, ethnic composition and urban share compared across each border pair. Second, continuity is shown by plotting binned means of these characteristics against signed distance on both sides of the boundary. Third, estimates are reported for bands of 2, 5, 10 and 20 km and for linear and quadratic f(dist). Finally, the design is repeated on border pairs within the same
prefecture, where both sides answer to the same prefecture government, and pairs that fail the balance tests are reported and dropped.

Design 2: Staggered event study (in event time) (key for the post-graduation question). Counties exited the designation in cohorts between 2016 and 2020, so the calendar year does not align with treatment status, and a calendar-time specification would confound counties at different points in their own post-graduation path. Denote Gc as the graduation year of county c and let event time be τ
= t − Gc, with τ = 0 being the year that county c graduated. Figure 2(a) shows the alignment. The estimating equation is
$yct = \alpha c + \lambda pt + \Sigma \tau \neq -1\ \theta\tau \cdot 1[t - Gc = \tau] + \gamma' Xc + \varepsilon ct$,
where αc are county fixed effects, λpt are province-by-year fixed effects for τ = −1, omitted as the reference period, and Xc are pre-treatment controls from Table 2. We estimate dynamic effects with the Callaway and Sant'Anna (2021) estimator instead of two-way fixed effects whenever cohorts are treated at different times, since the latter can place negative weights on comparisons under staggered adoption. For each cohort, we construct a comparison group of not-yet-graduated counties in that year, along with never-designated counties matched on baseline luminosity, terrain, and 1990 population. The θτ for τ ≥ 0 are the estimates that separate the hypotheses: non-decreasing under H1, decaying under H2, and indistinguishable from zero under H3. The whole specification is estimated separately for each of the two outcomes.
Implementation. Treatment cohorts are the five graduation years 2016–2020, with the 2020 cohort handled as in threat (vi). Group-time effects are estimated with the doubly robust Callaway and Sant'Anna estimator, conditioning on the pre-treatment controls in Table 2, and aggregated to event- time effects θτ with weights proportional to cohort size. Because every remaining county graduated in 2020, not-yet-graduated counties can serve as the comparison

ISSN 3106-5767 (Print) ISSN 3106-7026 (Online)

group only up to 2019; they are used for pre-graduation tests and short horizons, and longer horizons are identified against the matched never-designated counties. Standard errors are clustered by county, the unit of treatment, and computed by multiplier bootstrap with uniform confidence bands over τ; prefecture-level clustering is reported as a check, since graduation decisions may be correlated within prefectures. Robustness checks vary the comparison group (not-yet-graduated only, never-designated only), the inclusion of covariates, the estimator (Sun and Abraham 2021, and conventional two-way fixed effects for reference), the event window (a balanced window in which every cohort is observed), and anticipation (allowing one year before Gc for certification inspections). Figures 1 and 2 are conceptual illustrations only: estimated θτ will be shown in separate event-study plots with confidence bands, never on the same axes as the schematic paths.

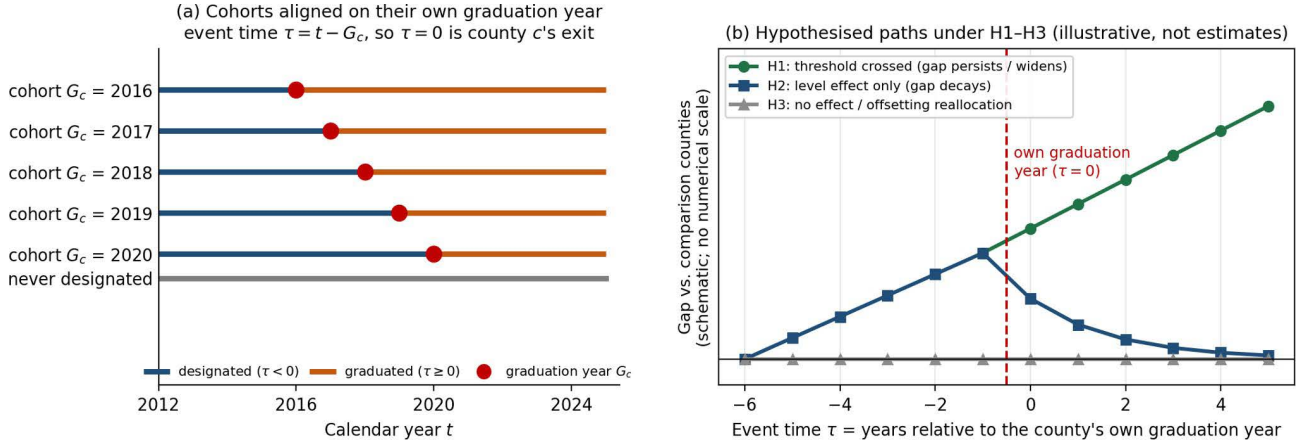


Figure 2. Conceptual illustration (no data). Schematic drawn by the author; the paths in panel (b) are hypothesised shapes, not estimates, and the vertical axis carries no numerical scale. Panel (a): counties graduated in different years between 2016 and 2020, so each cohort is aligned on its own graduation year Gc and analysed in event time τ = t − Gc; never-designated counties supply an untreated comparison at every τ. Panel (b): predicted event-study paths. Effects are positive before τ

= 0 under both H1 and H2 because transfers are still flowing; the hypotheses separate only afterwards. The vertical axis represents the effect relative to comparison counties, for either outcome.

Design 3: sharp regression discontinuity (secondary). The 2012 round applied an income threshold of 2,300 yuan, which invites a regression discontinuity. It is a cross-check rather than the main design for three reasons: the running variable exists only as a county-level figure in statistical yearbooks; the cutoff was relaxed for ethnic-minority and old revolutionary base areas, making the rule fuzzy rather than sharp; and counties had an obvious incentive to report income just below the line. A McCrary density test at the cutoff measures how much manipulation there is, and the resulting estimates are compared with those from Design 1.

Threats to validity

(i) Proxy error. The elasticity of output to light is around 0.3 across countries (Henderson, Storeygard and Weil 2012), and it is weaker for poor rural areas, which attenuates estimates toward zero and works against H1. Multiple luminosity measures and the CHNS validation mitigate but do not remove this.

(ii) Migration. Both per-capita luminosity and per-capita income can rise because people left, not because of an escape from a trap. Thus, for every per-capita output, total luminosity, total income and the population series itself are reported, and out-migration is treated as a substantive response to be measured rather than a nuisance to be controlled away.

(iii) Continued support after graduation. Graduation ended designated status, but it did not necessarily end all government support. There was a five-year transition policy from 2021–2025, and a subgroup of ex-poor counties was made national rural-revitalisation priority counties with ongoing support. In these counties, τ = 0 denotes a label change rather than an end to transfers, so incorrectly pooling them with the rest would bias the post-graduation path toward persistence and therefore towards H1. Consequently, they are identified from the published lists, excluded from the main estimation sample and analysed separately, with the core specification restricted to counties that received no successor designation. Where the two sets of estimates differ, only the restricted set can be interpreted.

(iv) Administrative reporting. An explicit numerical certification target gives county governments incentives to report income above the line in the years surrounding graduation (Wallace 2016). This threatens Outcome 2 specifically. These include three types of tests: lights-to-income elasticity estimated separately for pre- and post-τ = 0 (Martínez 2022); reported income bunching just above the certification threshold; and comparison of CHNS household estimates with reported county yield over

pooled windows. The paper finds and rests its substantive conclusion on Outcome 1: whether any reported income responds to certification.

(v) Endogenous exit timing. Counties had to demonstrate readiness to be certified, so early graduates were plausibly already on a stronger path. This biases the post-graduation trajectory toward persistence: a finding of decay is robust to it, a finding of persistence is not. The asymmetry is stated up front, and cohort-specific estimates are reported so that early and late graduates can be compared directly. Three tests address it directly. First, pre-graduation trends: with not-yet- graduated designated counties as the comparison, the coefficients $\theta\tau$ for $\tau \leq -2$ are plotted and tested jointly for equality to zero; both groups are still designated, so no gap is expected before exit. Second, placebo graduations: each cohort is assigned a false graduation year three years before its actual one and the model is re-estimated on pre-graduation years only; in addition, graduation years are randomly permuted across designated counties within province to obtain a reference distribution for the post-graduation slope. Third, cohort comparisons: 2012–2015 luminosity growth is compared across cohorts, and if early graduates were already growing faster, pre-2016 growth is added to the matching variables. If these tests fail, the post-graduation path is not read as evidence of persistence.

(vi) The pandemic. In addition, we observe in Section 2 that COVID-19 disrupted activity, mobility, and measurement, which likely disrupted the 2020 graduation cohort, with $\tau = 0$ in that same year. Thus, changes occurring in 2020 are not interpreted as policy impacts. The baseline specification excludes 2020 from the post- graduation slope estimation and begins the post-window in 2022; separate coefficients for 2020 and 2021 are presented descriptively. To check the robustness of these findings, we repeat the analysis on graduates from 2016–2018 only, who had several years free of any clear post-graduation contestation before the pandemic.

(vii) Small-sample noise. Means from a national household survey for one county-year are based on extremely few observations, and there are always some revisions in the county-level administrative series. Neither is read year by year. Inference relies on pooling pre- and post- graduation windows and the estimated path over several years, with annual deviations treated as noise unless they persist.

Falsification

If a threshold exists, effects should be non-linear in the cumulative push relative to baseline capital, and the cross-county distribution of the outcome should be bimodal; both are testable with threshold regression and a formal bimodality test. The absence of both, together with post-graduation decay, would be evidence for H2 and against area-based poverty traps in this setting.

7. Contribution and limitations

The contribution is threefold. The proposal reframes the largest area-based antipoverty programme on record as a test between steady-state structures rather than as another estimate of an average treatment effect; it organises the analysis around each county's own graduation year rather than around a single calendar date; and it measures development with two outcomes whose measurement errors have different sources, so that agreement between them is informative and disagreement is diagnostic.

Five limitations are stated pre-emptively. First, a county is not a household: Balboni et al. (2022) locate thresholds in household asset holdings, so the absence of a county-level trap would not rule out household-level traps. Second, five post-graduation years is a short window for a claim about permanence; the design can show that a gap decays but cannot establish that a gap will persist for a generation. Third, luminosity tracks electrification and built structure more closely than consumption or human capital, so a null on lights is not a null on welfare. Fourth, the border design fails wherever a boundary was itself drawn along a pre-existing economic fault line, which must be checked pair by pair rather than assumed away.

Fifth, and contributing most to our understanding of what this design recovers: the reduction in outcome is due to losing designation rather than a treatment effect on any other object that will appear over the long run of the programme. Statute is not the only force that drives a wedge between these two—continuing support under successor policies, migration between treated and comparison counties, and reporting behaviour around certification all do. None is solved, but each is responded to above. As such, one should interpret the estimates as follows: they represent a trajectory for counties that were pointed at graduation; moving from that trajectory to a long-run impact-policy claim requires interpreting the paper's result, which is made explicit and defended using the checks in Section 6, rather than assumed.

8. Conclusion

This proposal converts a question about whether a policy worked into a question about the kind of economy it acted on. China's poor-county designation ended for each county on a date that is known and that differs across counties, and the path a county follows after its own exit distinguishes a threshold that was crossed from a level effect that decays.

Did the treatment and control groups differ at baseline, as measured twice: via satellite luminosity, generated entirely externally from geographical assignments of treatment and certificates of exit; and reported income per capita, the variable targeted by the programme? The U.S. federal agency, national laboratories, and universities compile population, terrain, boundary, and baseline demographic

ISSN 3106-5767 (Print) ISSN 3106-7026 (Online)

data. Comparing the two outcomes allows the design to separate a change in the local economy from a change in what was reported about it.

The three hypotheses carry different conclusions. Continued divergence after graduation (H1), if it survives the checks on successor support, migration, spillovers and exit timing, would be consistent with an area-based push having moved local economies past a threshold, and the size and timing of exit would then become the central design questions for such programmes. Decay (H2) would place the Chinese programme alongside the transfers reviewed by Kraay and McKenzie (2014) and imply that area-based support must be permanent to have a permanent effect, whatever the graduation certificates record. A null (H3) would require the falsification tests — non-linearity in the cumulative push and bimodality in the cross-county distribution — to distinguish genuine absence of a trap from offsetting reallocation.

The limits in Section 7 bound all three readings. What the design offers that the existing evaluations cannot is a test conducted after designated status ended, on two measures at once, and a comparison the data constraint itself makes available: 832 Chinese counties that received a very large push, set

against roughly three hundred U.S. counties that have stayed above a twenty per cent poverty rate for three decades without one.

9. Statement on the use of AI

This proposal was prepared with the assistance of a large language model (Claude, Anthropic), used for discussing the research design, drafting and condensing text that I then revised, searching for candidate literature and datasets, and writing the plotting code for the conceptual Figures 1 and 2. It was not used to generate, simulate or impute any data, and every reference and data link was checked against its original source. The full statement, including the corrections made during verification, is given in Appendix B.

Appendix A. Data access

All links were checked in August 2026. Access conditions differ: the satellite, NASA SEDAC and USDA products are open; IPUMS-International and CHNS require free registration and a conditions-of-use agreement; China Data Online requires an institutional licence.

DMSP-OLS v4 stable lights, 1992–2013 — Earth Observation Group (historical archive at NOAA NCEI): https:// eogdata.mines.edu/products/dmsp/

VIIRS Day/Night Band annual composites (VNL V2.2), 2012–2024 — Earth Observation Group, Payne Institute, Colorado School of Mines: https://eogdata.mines.edu/ products/vnl/

LandScan Global annual population grids, 2000– — Oak Ridge National Laboratory: https:// landscan.ornl.gov/

Gridded Population of the World v4.11 — CIESIN, Columbia University, distributed by NASA: https://www.earthdata.nasa.gov/data/projects/gpw

SRTM elevation — NASA / USGS, via NASA Earthdata (search “SRTM”): https:// www.earthdata.nasa.gov/

China Administrative Regions GIS Data, 1:1M, county level, 1990 — CITAS (University of Washington), CASM and CIESIN, archived at NASA SEDAC: https:// sedac.ciesin.columbia.edu/ data/set/cddc-china-admin-regions-gis-1990

China county-level data on population (census) and agriculture, keyed to the 1:1M map, 1990 — CITAS and CIESIN, NASA SEDAC: https://www.earthdata.nasa.gov/ data/catalog/sedac-ciesin- sedac-cd-cty-popag-gis-1.00

China county-level data from provincial economic yearbooks, 1990–1991 — CITAS and CIESIN, NASA SEDAC: https://www.earthdata.nasa.gov/data/catalog/sedac-ciesin-sedac-cd-cty-econ- yrbk-1.00

GADM administrative boundaries, v4.1 (modern county polygons) — University of California, Davis: https:// gadm.org/data.html

County income per capita (China County Statistical Yearbook and provincial yearbooks) — China Data Online, licensed through U.S. university libraries: https://www.china-data-online.com/ — note that the University of Michigan China Data Center, the former U.S. distributor, closed in 2018, and that the underlying series originates in the Chinese statistical system

China Health and Nutrition Survey, 1989–2015 — Carolina Population Center, University of North Carolina at Chapel Hill: https://chns.cpc.unc.edu/

IPUMS-International, China census samples — University of Minnesota: https:// international.ipums.org/international/ (province-level geography only; used for context)

International Macroeconomic Data Set — USDA Economic Research Service: https:// www.ers.usda.gov/data-products/international-macroeconomic-data-set

Persistent-poverty counties — USDA Economic Research Service: https://www.ers.usda.gov/topics/ rural-economy-population/rural-poverty-well-being

Penn World Table 11.0 — University of California, Davis and Groningen Growth and Development Centre: https:// www.rug.nl/ggdc/productivity/pwt/

Designation and graduation status — State Council poor-county lists (1986, 1994, 2001, 2012) and 2016–2020 graduation notices. No consolidated public download exists; county lists are taken as reproduced in the appendices of Park, Wang and Wu (2002) and Meng (2013), and graduation cohorts from State Council Leading Group Office announcements as compiled in the published literature

Appendix B. Full statement on the use of AI

Disclosures: This document was prepared with the assistance of a large language model (Claude, Anthropic) for writing and research support, rather than as an authoritative source. The model contributed in three primary ways. First, during the refinement of the research question, it served as an interlocutor and was repeatedly prompted to challenge, rather than endorse, the identification argument. Second, it functioned as a drafting assistant: many paragraphs in Sections 2 through 7 originated from the model, after which I cut, rearranged, and rewrote them. I also used the model to help condense sentences I had previously written. Third, the model served as a search aid for candidate literature and datasets; most of the verification steps described below originated from this process. Additionally, the model generated the matplotlib code used to plot Figures 1 and 2, which are conceptual schematics of the framework rather than empirical data plots. This document is a proposal and does not contain original estimates; all numerical values in Sections 2 and 3 are drawn from the relevant literature.

The key decisions underlying this proposal were made prior to engaging the language model, which was explicitly instructed to challenge rather than reinforce these choices. These decisions include: selecting exit rather than entry designation as the distinguishing variation, thereby framing the analysis as a test between two structures rather than estimating an average effect; retaining both luminosity and reported income as outcomes and examining their relationship, rather than omitting the administration series; and excluding income from the right-hand side of the analysis. Each of these choices was revisited and refined in response to my instructor's feedback.

Model outputs were treated as provisional until verified at the original source, and this verification process was substantive rather than procedural. Because fluent drafts may persist in obscurity, I highlight four specific corrections. First, the initial version included state administrative units from Harvard's China Historical GIS; however, this resource only covers pre-1911 administrative units. Therefore, the 1990 CITAS 1:1M county map archived at NASA SEDAC and GADM v4.1 for modern polygons was used instead. Second, IPUMS-International was initially assigned a county- level role, but its census geography only provides geographic identifiers at a higher administrative level. Consequently, it appears in Table 1 solely for contextual information. Third, references to DMSP series intercalibration by Elvidge et al. (2017) were replaced by Hsu et al. (2015), because

Elvidge et al. (2017) concerns VIIRS rather than DMSP. Elvidge et al. is retained wherever VIIRS is discussed. Fourth, a draft note mischaracterized Chen, Mu, and Ravallion (2009) as 'a survey of Taiwanese households' and reversed the direction of its findings; in fact, the study surveys households in south-west China, and that section was quoted verbatim from the article. The issue of circularity raised in Section 5—namely, that LandScan uses nighttime lights as an input for population allocation, complicating its use as a denominator for lights-based outcomes—was addressed through the same verification process. All references in the list have been matched to the published article, and every link in Appendix A was checked and verified in August 2026.

Two exceptions require emphasis. The model was not used to generate, simulate, or impute any data. If information was not available in the published literature, no additional data were introduced, except where explicitly cited from authoritative sources. This approach aligns with the proposal's practical argument: the provenance of a language model must be established externally, as the model produces fluent text regardless of the underlying truth of its claims. Therefore, the considerations outlined in Section 5 also apply to county income data reported by the same administration that certified graduation. Any remaining errors are my responsibility, and the argument presented is solely my own.

References

[1] Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. American Economic Review, 91(5), 1369–1401.
[2] Angrist, J. D., & Pischke, J.-S. (2009). Mostly Harmless Econometrics: An Empiricist's Companion.
[3] Princeton: Princeton University Press.
[4] Balboni, C., Bandiera, O., Burgess, R., Ghatak, M., & Heil, A. (2022). Why do people stay poor?
[5] Quarterly Journal of Economics, 137(2), 785–844.
[6] Banerjee, A. V., & Iyer, L. (2005). History, institutions, and economic performance: The legacy of colonial land tenure systems in India. American Economic Review, 95(4), 1190–1213.
[7] Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods.
[8] Journal of Econometrics, 225(2), 200–230.
[9] Chen, S., Mu, R., & Ravallion, M. (2009). Are there lasting impacts of aid to poor areas? Journal of Public Economics, 93(3–4), 512–528.
[10] Elvidge, C. D., Baugh, K., Zhizhin, M., Hsu, F. C., & Ghosh, T. (2017). VIIRS night-time lights.
[11] International Journal of Remote Sensing, 38(21), 5860–5879.
[12] Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The next generation of the Penn World Table.
[13] American Economic Review, 105(10), 3150–3182.
[14] Gibson, J., Olivia, S., & Boe-Gibson, G. (2020). Night lights in economics: Sources and uses.
[15] Journal of Economic Surveys, 34(5), 955–980.
[16] Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. American Economic Review, 102(2), 994–1028.

ISSN 3106-5767 (Print) ISSN 3106-7026 (Online)

[17] Hsu, F.-C., Baugh, K. E., Ghosh, T., Zhizhin, M., & Elvidge, C. D. (2015). DMSP-OLS radiance calibrated nighttime lights time series with intercalibration. Remote Sensing, 7(2), 1855– 1876.

[18] Kline, P., & Moretti, E. (2014). Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority. Quarterly Journal of Economics, 129(1), 275–331.

[19] Kraay, A., & McKenzie, D. (2014). Do poverty traps exist? Assessing the evidence. Journal of Economic Perspectives, 28(3), 127–148.

[20] Lü, X. (2015). Intergovernmental transfers and local education provision — Evaluating China's 8–7 National Plan for Poverty Reduction. China Economic Review, 33, 200–211.

[21] Martínez, L. R. (2022). How much should we trust the dictator's GDP growth estimates? Journal of Political Economy, 130(10), 2731–2769.

[22] Meng, L. (2013). Evaluating China's poverty alleviation program: A regression discontinuity approach. Journal of Public Economics, 101, 1–11.

[23] Park, A., & Wang, S. (2010). Community-based development and poverty alleviation: An evaluation of China's poor village investment program. Journal of Public Economics, 94(9–10), 790–799.

[24] Park, A., Wang, S., & Wu, G. (2002). Regional poverty targeting in China. Journal of Public Economics, 86(1), 123–153.

[25] Ravallion, M., & Chen, S. (2007). China's (uneven) progress against poverty. Journal of Development Economics, 82(1), 1–42.

[26] Sachs, J. D. (2005). The End of Poverty: Economic Possibilities for Our Time. New York: Penguin Press.

[27] Solow, R. M. (1956). A contribution to the theory of economic growth. Quarterly Journal of Economics, 70(1), 65–94.

[28] Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics, 225(2), 175–199.

[29] Wallace, J. L. (2016). Juking the stats? Authoritarian information problems in China. British Journal of Political Science, 46(1), 11–29.