Chinese vs. World Bank development projects: Insights from earth observation and computer vision on wealth gains in Africa, 2002–2013

Downloaded from: https://research.chalmers.se, 2026-09-12 23:40 UTC

Citation for the original published paper (version of record):
Daoud, A., Conlin, C., Jerzak, C. (2026). Chinese vs. World Bank development projects: Insights from earth observation and computer
vision on wealth gains in Africa, 2002–2013. World Development, 202.
http://dx.doi.org/10.1016/j.worlddev.2026.107328

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

World Development 202 (2026) 107328

Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

Regular research article

Chinese vs. World Bank development projects: Insights from earth observation and computer vision on wealth gains in Africa, 2002–2013

Adel Daoud [a,b,d,1,2,*], Cindy Conlin [a,d,1,2], Connor T. Jerzak [c,d,1,2]

[a] Linköping University, Linköping, Sweden
[b] Chalmers University, Gothenburg, Sweden
[c] University of Texas at Austin, Austin, USA
[d] AI and Global Development Lab, Linköping, Sweden

HIGHLIGHTS

- Satellite imagery controls reveal that prior studies likely overstated aid benefits in Africa.
- On average, Chinese development projects generate larger wealth gains than World Bank projects.
- Earth Observation and machine learning improve causal estimates by proxying for planners' maps.
- China's project placement is less predictable, suggesting unique political or event-driven logics.
- Both donors raise wealth, but impacts vary widely by sector, with Emergency Response having the largest effect.

ARTICLE INFO

Keywords:
Living conditions
World Bank programs
Chinese programs
Impact evaluation
Earth observation
Computer vision

ABSTRACT

Debates about whether development projects improve living conditions persist, partly because observational estimates can be biased by incomplete adjustment and because reliable outcome data are scarce at the neighborhood level. We address both issues in a continent-scale, sector-specific evaluation of Chinese and World Bank projects across 9899 neighborhoods in 36 African countries (2002-2013), representative of ~88% of the population. First, we use a recent dataset that measures living conditions with a machine-learned wealth index derived from contemporaneous satellite imagery, yielding a consistent panel of 6.7 km square mosaics. Second, to strengthen identification, we proxy officials' map-based placement criteria using pre-treatment daytime satellite images and fuse these with tabular covariates to estimate funder- and sector-specific ATEs via inverse-probability weighting. Incorporating imagery often shrinks effects relative to tabular-only models. On average, both donors raise wealth, with larger and more consistent gains for China; sector extremes in our sample include *Trade and Tourism (330)* for the World Bank (+12.29 IWI points), and *Emergency Response (700)* for China (+15.15). Assignment-mechanism analyses also show World Bank placement is often more predictable from imagery alone (as well as from tabular covariates). This suggests that Chinese project placements are more driven by non-visible, political, or event-driven factors than World Bank placements. To probe residual concerns about selection on observables, we also estimate within-neighborhood (unit) fixed-effects models at a spatial resolution about 67 times finer than prior fixed-effects analyses, leveraging the computer-vision-imputed IWI panels; these deliver smaller but, for Chinese projects, directionally consistent effects. Methodologically, we extend recent EO–ML causal inference frameworks by fusing pre-treatment satellite imagery with tabular covariates to estimate treatment propensities, and by systematically benchmarking image-augmented estimators against tabular-only and unit fixed-effects designs using new assignment-mechanism diagnostics. Empirically, we provide a continent-wide, sector-specific comparison of the neighborhood-level wealth effects of Chinese and World Bank projects across 9899 African neighborhoods.

* Corresponding author at: Linköping University, Linköping, Sweden.
Email address: adel.daoud@liu.se (A. Daoud).
[1] All authors contributed equally.
[2] www.aidevlab.org

https://doi.org/10.1016/j.worlddev.2026.107328
Accepted 23 January 2026
Available online 10 February 2026
0305-750X/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

The United Nations' establishment of the Millennium and Sustainable Development Goals, along with China's Belt and Road Initiative, has sparked growing interest since 2000 in examining how aid and development programs help nations overcome poverty. Indeed, around 2005, China switched from being a net aid recipient to a net donor (Chin, 2012). Scholars estimate China now spends twice as much as the U.S. and other major donors, like the World Bank (Daoud et al., 2022; Malik et al., 2021), generating considerable interest among international actors about China's increasing influence in developing nations, particularly Africa, and about how Chinese activities "affect social, economic, environmental, and governance outcomes in [these] low- and middle-income countries" (Dreher et al., 2022a, p. 7). Between 2011 and 2020, an average of 82 papers per year examined aid effectiveness (Asatullaeva, 2021).

However, this large and expanding literature has not reached consensus on whether development programs are effective at improving the living conditions of African people, producing economic growth and fostering institutional development in recipient countries (Ahmed, 2022; Asatullaeva, 2021; Daoud and Johansson, 2024; Daoud et al., 2017; Halleröd et al., 2013; Mandon and Woldemichael, 2023; McGillivray et al., 2006). One way to improve the evidence base of development-programs efficiency is to conduct randomized controlled trials (Banerjee et al., 2015; Duflo, 2015), but these are challenging to conduct at the scale of the African continent. Thus, most aid studies are observational and face significant methodological challenges in identifying unbiased causal estimates that apply to the target population.

There are two factors limiting such large-scale analyses. First, limited capacity in recipient countries often results in missing or low-quality outcome data, especially at the sub-national level (Burke et al., 2021; Daoud et al., 2021). Such data would cost millions of USD to collect. With the potential dismantling of USAID and, consequently, the Demographic and Health Surveys (DHS), alternative methods are needed to collect data on people's health and living conditions in Africa and beyond.

Second, this limited capacity also hampers the availability of geo-temporal control variables that would enlarge the adjustment set for causal identification (Jerzak et al., 2023). For example, one such covariate set would be to access the maps that aid and development program officials used when allocating their programs. For example, China considers its foreign aid and loan activities a state secret. Official data on these activities are not publicly available, so studies of Chinese aid use a dataset collected from media reports, recipient governments, researchers, non-governmental organizations, and Chinese embassy websites (Strange, 2017).

This study contributes to the literature by applying two innovations that address data limitations and methodological challenges in evaluating the effectiveness of foreign aid from two distinct aid donors: China and the World Bank. We study the aid they provided to 36 African countries (representing 88% of the continent's population) between 2002 and 2013, based on a representative sample of 9899 neighborhoods as the pool of potential treatment and control units of analysis.

The first innovation of our study is that we use a new data source for subnational wealth at a resolution of 6.7 km over all African human settlement areas, generated by Pettersson et al. (2023), who trained a computer vision machine learning algorithm to learn the relationship between daytime and nighttime satellite images and on-the-ground wealth measures from DHS from the same times and places. Once the algorithm learned the association between what it observed in satellite images and wealth measures, it imputed wealth for areas without DHS wealth measures to create a continent-wide IWI measure for the entire African continent, from 1990 to 2020. This would not have been possible using DHS data alone, since surveys were not conducted repeatedly for the same neighborhoods and are available only in limited timeframes for each country; see Table 1.

The second innovation of our article is that we apply a causal inference method, based on the planetary-causal-inference paradigm (Jerzak, 2023a,b), based on the hybrid modeling culture, which combines *causal* and *predictive* logics (Daoud and Dubhashi, 2023). This paradigm aims to advance research on global development by developing earth observation (EO) and machine learning (ML) methods for causal analysis and measurement—denoted *EO-ML methods* as a shorthand. To enhance our identification approach, we use daytime satellite imagery over treated and control communities to adjust for factors visible to remote sensors (such as physical geography, topography, temperature, and infrastructure) and their spatial arrangement. These images will likely enhance identification because they provide additional information on why Chinese and World Bank officials place their programs in some communities rather than others (Bedi et al., 2007b). Decision makers may look for specific patterns in their maps—proxied by our satellite images—when making sub-national project site placement decisions. These image-based confounders identify features that are hard to encode in tabular data.

We provide our EO-ML causal method with daytime satellite images centered over treated and control communities, one period before the aid project commitment, along with a large set of tabular-format variables representing political, economic, and other factors that empirical research has shown to be relevant to aid allocation and effectiveness. We employ EO-ML methods to identify features visible in satellite imagery (along with tabular covariates) to estimate each community's treatment propensity. These propensities are used in an inverse probability weighting procedure to estimate the effect of aid from each funder and sector on neighborhood wealth.

This EO-ML approach targets the average treatment effect (ATE) estimand for 36 African countries over the period 2002 to 2013. Many policy evaluation studies on Chinese and World Bank programs target the local average treatment effect (LATE) estimand, because they use an instrumental variable approach. Although LATE is considered to provide a more credible identification approach, this estimand may struggle with external validity because the variation it targets is much narrower than the ATE, as this variation relies on the compliers in the instrument. As our goal is to estimate an Africa-wide effect over the given time period, we target the ATE. Our expectation is that we will improve our identification by relying on an additional adjustment set—satellite images to capture selection-relevant covariates.

Following recent contributions (Chai and Tang, 2023), our study analyzes aid by sector (e.g., *Health*, *Education*, *Energy Generation & Supply*), which is important due to the heterogeneity in placement logics, political economy factors, expected timeframes for results, and expected geographic reach across sectors. Earlier studies analyzed sector groups, such as early- vs. late-impact aid (Dreher and Lohmann, 2015) or Economic, Social, and Production groups (Bitzer and Gören, 2018; Dreher et al., 2021; Xu et al., 2020). In contrast, our study analyzes individual aid sectors within the same framework; this enhances the comparability of aid across funders who prioritize different sectors.

The remainder of the paper proceeds as follows. The next section examines the factors that influence the placement and effectiveness of foreign aid, with a focus on the processes of development projects from China and the World Bank. Following this is a detailed description of the data and methods, accompanied by the results, discussion, opportunities for future research, and conclusion. Three appendices—Data, Results, and Code—are included separately and provide more detailed information.

1. The political economy of development projects and human development

While a large literature examines the political and economic determinants of development-project allocation, a parallel and often contentious debate centers on a more fundamental question: Does aid, or other forms of development projects, actually improve the living conditions

of people in recipient countries? (Asatullaeva, 2021; McGillivray et al., 2006). Answering this question is complicated by the diverse strategies of donors and the micro-macro paradox that aid can be locally effective yet hard to detect in aggregate outcomes. This study focuses on Africa's two most significant development financiers—China and the World Bank—whose contrasting approaches provide a critical lens for examining the impact of aid on poverty, wealth, and local economic activity. China has recently become the single largest financier of African infrastructure, while the World Bank remains one of the largest traditional donors with a broad development mandate. Their sharply different models of aid delivery offer a unique opportunity to assess whether and how foreign aid improves living conditions on the ground.

1.1. The impact of Chinese development finance on living conditions

China's emergence as a major development partner in Africa has been characterized by a focus on "hard" infrastructure projects—roads, railways, ports, and power plants (Brautigam, 2009; Dreher et al., 2022a). The primary mechanism through which this form of aid is believed to impact living conditions is by stimulating local economic activity. Sub-national studies using satellite nightlights as a proxy for economic growth consistently find that Chinese-funded infrastructure projects generate significant localized economic booms (Bluhm, 2020; Dreher et al., 2021). This reflects China's strategy of "connective financing," which reduces transportation costs, links remote areas to markets, and can spur economic agglomeration, creating measurable spillovers in the local economy (Donaldson and Storeygard, 2016).

However, whether these localized growth spurts translate into broad-based improvements in household welfare is less clear. Some micro-level studies find that Chinese aid projects are associated with increased household consumption, asset ownership, and employment, suggesting a direct positive impact on poverty reduction (Brazys et al., 2021; Leiderer, 2021; Martorano et al., 2020). Yet, the benefits may not be evenly distributed. The political economy of Chinese aid allocation—which can favor the birth regions of political leaders (Dreher, 2019) and often bypasses local governments and civil society channels—may concentrate economic gains in specific locales or among politically connected elites (Gehring, 2022; Lee et al., 2021). Moreover, although Chinese projects create employment opportunities, debates continue over the extent to which they rely on local versus imported Chinese labor and the quality of jobs created for African workers (Warmerdam and van Dijk, 2013).

China also funds numerous projects in social sectors, but the impact of these projects on living conditions is less studied and potentially more mixed. China's demand-driven, state-led model may enable the rapid construction of schools and clinics, but some research suggests that these projects may be less effective at improving long-term outcomes compared to those of traditional donors (Martina, 2020). The reason often cited is a focus on "hardware" over "software"—Chinese projects tend to emphasize the provision of physical infrastructure without an equivalent investment in the staffing, training, and institutional capacity needed to deliver services effectively.

1.2. The impact of World Bank development projects on living conditions

In contrast to China's infrastructure-heavy portfolio, the World Bank pursues a much broader development mandate that includes poverty reduction, human capital development, and governance reform (Scott, 2009). Its theory of change relies on more indirect and long-term mechanisms, aiming to build sustainable local capacity by working through recipient government systems—a principle formalized in the Paris Declaration on Aid Effectiveness (OECD, 2005).

Consequently, the measured short-run impact of World Bank aid on local economic activity is often smaller and less immediate than that of Chinese aid. Studies using nightlights data, for example, have difficulty detecting any significant uptick in luminosity around World Bank project sites (Bitzer and Gören, 2018; Dreher and Lohmann, 2015). However, this does not necessarily mean the Bank's projects have a weaker impact on living conditions; rather, they often operate through channels that nightlights fail to capture. The World Bank allocates a significant portion of its portfolio to social sectors, where investments in health and education yield long-term gains in human capital (Krueger and Lindahl, 2001). Empirical evidence bears this out: providing safe water has been shown to directly improve health outcomes by reducing disease incidence and child mortality (Duflo, 2015; Galiani et al., 2005). Yet, working through government ministries can slow implementation and create opportunities for elite capture (Platteau, 2004), and the Bank's priorities are not immune to the influence of its most powerful shareholders (Andersen, 2006; Kim and Kim, 2021; Martel, 2021). These biases can blunt the poverty-reducing impact of projects, despite the Bank's explicit pro-poor mandate (Briggs, 2017).

2. Two key gaps

Two key challenges emerge from this literature. First is the problem of measuring neighborhood-level socioeconomic outcomes (Daoud et al., 2021; Jean, 2016; Kino et al., 2021; Sakamoto et al., 2025). Most sub-national impact studies have relied on satellite nighttime lights, a proxy that is better suited to capturing the effects of large-scale infrastructure than improvements in household well-being. Our study addresses this gap by using a novel wealth index derived from machine learning and high-resolution satellite imagery (Pettersson et al., 2023), thereby providing a more holistic view of living conditions.

Second is the persistent challenge of causal identification. Development programs are not randomly assigned (Bedi et al., 2007b). While past studies have used instrumental variables or fixed-effects models, these approaches may not fully account for fine-grained confounders related to geography and the built environment. Our study adapts and applies an EO–ML causal adjustment approach to further enhance identification (Sakamoto et al., 2025): we use the satellite images themselves as control variables, which also enable new kinds of substantive insights to be gained about the assignment mechanism surrounding aid (Daoud and Dubhashi, 2023; Rubin, 1991)—capturing what kinds of information do and do not predict aid decisions. By conditioning on satellite images—which proxy for the maps used by aid officials in project allocation (Bedi et al., 2007a)—we compare treated and untreated areas with greater sensitivity, thereby enhancing the credibility of our causal estimates and addressing limitations in prior literature (Easterly, 2006). In this way, we can more robustly identify causal impacts of aid interventions while controlling for and explicitly characterizing the funder targeting strategies.

2.1. Measuring strategic allocation with satellite images & computer vision

To work toward closing these two gaps, we proceed as follows. We begin with identification: we formalize the project placement problem and show how pre-treatment satellite imagery can proxy the map-based information donors use when deciding where to locate projects, strengthening causal adjustment in an observational setting. We then turn to measurement, introducing the satellite-derived wealth panel that makes neighborhood-level evaluation feasible at continental scale.

We start by laying out the treatment-assignment problem and clarifying the role of imagery as a proxy for otherwise-unobserved spatial selection factors in observational studies of development programs. We will estimate the effects of World Bank and Chinese projects on neighborhood-level living conditions in Africa, drawing on the potential outcomes framework to define estimands and invoking graphical models to clarify assumptions. Our strategy leverages satellite imagery not merely as a covariate but as a proxy for the geospatial features that likely inform donors' assignment decisions, thereby clarifying aspects of the program allocation mechanism itself (Bedi et al., 2007b).

Here, a note on the unit of analysis is warranted. As shown in Fig. 1, the outcome Y_{it} is a raster of 6.7 km side lengths. The outcome represents socioeconomic well-being (i.e., health and living conditions broadly

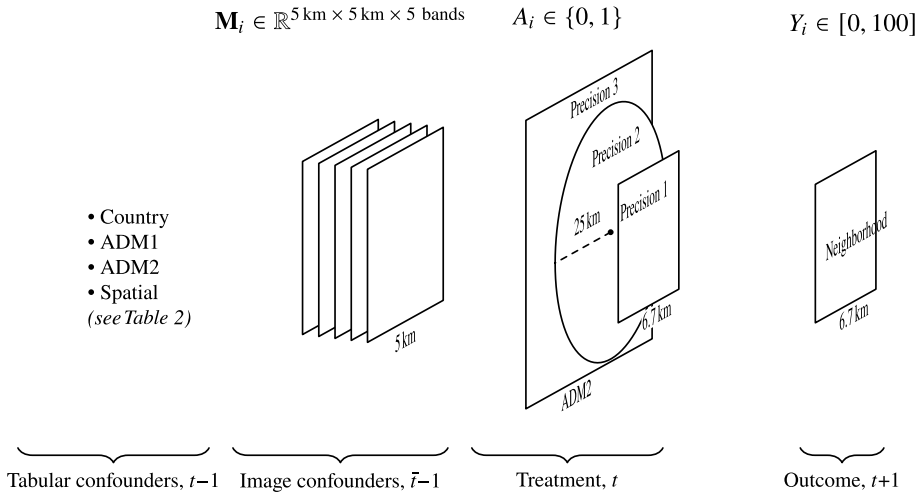


Fig. 1. Visualizing the shape of the data objects.

defined) at time t, measured via the International Wealth Index derived from satellite imagery as detailed in Pettersson et al. (2023) and further discussed in the data section. Substantively, this square i measures approximately a *neighborhood*—a block in an urban area or a village in a rural area. The outcome is measured one 3-year-period after the treatment has been deployed, for simplicity, denoted $t+1$. The treatment A_{it} denotes the presence or absence of a development project.

Each neighborhood has geocoordinates $\mathbf{l}_i \in \mathbb{R}^2$ (centroid of the square) and binary treatment indicators $A_{i,k,WB} \in \{0, 1\}$ for World Bank projects and $A_{i,k,China} \in \{0, 1\}$ for Chinese projects, where k indexes sector-specific interventions (indices for actors and sectors are dropped for brevity where unambiguous). Sector definitions appear in the data section.

A unit is in the treated group if at least one of the following holds for projects in its vicinity: (i) for precision-1 (exact) locations, the project point lies inside the 6.7 km neighborhood square to capture direct local exposure; (ii) for precision-2 ("near") locations, the neighborhood centroid is within 25 km of the reported project point to account for greater spatial uncertainty; or (iii) for precision-3 locations, where AidData only reports the second administrative division (ADM2), the neighborhood falls inside that ADM2. In other words, we only treat entire ADM2s when project locations are intrinsically coarse (precision 3); for projects with more precise coordinates, we do not treat ADM2s as a whole. Treatment status is determined contemporaneously with the commitment period.[3]

Under the potential outcomes framework (Rubin, 1990), each unit possesses counterfactuals $Y_{it}(1)$ and $Y_{it}(0)$. Our analysis targets sector-specific average treatment effects (ATEs) for World Bank and Chinese projects:

$$\text{ATE} = \mathbb{E}[Y_{it}(1) - Y_{it}(0)].$$

For a given sector, there will be two ATEs, one for World Bank and another for Chinese projects.

Identification commonly proceeds via backdoor adjustment, as depicted by the directed acyclic graph (DAG) in Fig. 2(b). Confounding emerges when unobserved factors $U_{i(t-1)}$ influence both assignment A_{it} and outcome Y_{it}. In aid allocation, such $U_{i(t-1)}$ often encompass geospatial attributes—terrain texture, infrastructure layouts, proximity to natural features, and other strategic reasons—that donors consult via maps during site selection. These maps, unavailable to researchers, encode spatiotemporal selection logics that drive why certain neighborhoods receive aid over others. Satellite images $\mathbf{M}_{i(\bar{t}-1)}$, which capture both observable and latent geospatial traits (e.g., inferred historical condition on the ground), serve as proxies for these maps. Conditioning on $\mathbf{M}_{i(\bar{t}-1)}$ thus blocks confounding paths by approximating the assignment mechanism donors employ. The image $\mathbf{M}_{i(\bar{t}-1)}$, as discussed in Data, is measured as a composition of the previous three-year images, prior to the treatment (hence, the bar $\bar{t}$ over the time indicator in Fig. 1).

While our image-based adjustment improves identification by recovering elements of $U_{i(t-1)}$, residual confounding $R_{i(t-1)}$—political, cultural, or economic factors beyond imagery or measured tabular covariates—may persist (Pearl, 2009). If $R_{i(t-1)}$ contributes minimally relative to $\mathbf{M}_{i(\bar{t}-1)}$ and observed covariates $\mathbf{X}_{i(t-1)}$, bias remains small.

For intuition, contrast the standard tabular-confounding DAG (Fig. 2(a)) with our image-augmented version (Fig. 2(b)). In the former, unobserved $U_{i(t-1)}$ biases estimates absent adjustment. In the latter, $\mathbf{M}_{i(\bar{t}-1)}$ precedes $U_{i(t-1)}$, as map features shape donor decisions (Öhler et al., 2019). These features are extracted with the help of a computer-vision function $f_U(\mathbf{M}_{i(\bar{t}-1)})$ which estimates what visible and latent structures are relevant for the presence or absence of an aid project. This estimation then enhances backdoor closure when conditioned alongside $\mathbf{X}_{i(t-1)}$.

Adjusting for high-dimensional images $\mathbf{M}_{i(\bar{t}-1)}$ demands specialized methods. We employ a computer vision method to yield $f_U(\mathbf{M}_{i(\bar{t}-1)})$, integrated with $\mathbf{X}_{i(t-1)}$ for propensity estimation and inverse probability weighting. This not only facilitates ATE recovery but elucidates assignment by revealing geospatial drivers of aid placement. This approach is further detailed in Methods.

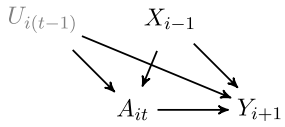
Building on this foundation, we will examine variation across funders and sectors in the relative influence of neighborhood-level factors (proxied by imagery) versus structural or institutional factors (proxied by country-level features in $\mathbf{X}_{i(t-1)}$) on assignment decisions. Such analyses promise clearer insights into donor-specific allocation logics, sharpening our understanding of aid's causal pathways.

We will later benchmark image-based adjustment against simpler specifications (difference-in-means, tabular-only, and ADM2 fixed effects) as well as unit-level fixed effects.

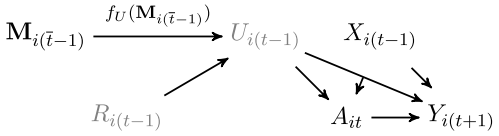
2.2. Outcome: Pettersson et al. (2023)'s satellite-derived wealth measure

Our second contribution addresses measurement of socioeconomic well-being (our outcome) by using the IWI data generated by Pettersson et al. (2023). The unit resolution is 6.7 km square and is measured in three-year periods (1990-1992, 1993–1995, etc.) over all African human settlement areas, which we call *neighborhoods*. This dataset used

[3] Investigators who wish to apply alternative treatment-footprint rules (e.g., different buffer radii or point-in-polygon criteria) and assess the sensitivity of estimates can do so using our replication package.



(a) Classical back-door confounding.



(b) Image-based confounding: $f_U(\mathbf{M}_{i(\tilde{t}-1)})$ recovers latent $U_{i(t-1)}$.

Fig. 2. (a) Traditional DAG with unobserved confounder U_i. (b) With satellite image $\mathbf{M}_{i(\tilde{t}-1)}$, a function $f_U(\cdot)$ maps $\mathbf{M}_{i(\tilde{t}-1)}$ into the latent confounder $U_{i(t-1)}$, blocking the back-door path when conditioned upon. $X_{i(t-1)}$ are observed covariates and $R_{i(t-1)}$ are confounders not present in the image.

Table 1
IWI by year group for `dhs_id = 48,923`: With vs. without CV imputations.

DHS ID	Year group	Treated	IWI
(a) Raw data without CV imputations			
48,923	2002:2004	1	?
48,923	2005:2007	0	19.6
48,923	2008:2010	0	?
48,923	2011:2013	1	?
(b) With CV imputations			
48,923	2002:2004	1	15.1
48,923	2005:2007	0	19.6
48,923	2008:2010	0	20.3
48,923	2011:2013	1	21.0

three-year intervals to handle clouds and missing images. Pettersson et al. trained a residual neural network machine learning algorithm to associate daytime and nighttime satellite imagery with IWI wealth measures from 138 Demographic and Health Surveys conducted in African countries between 1990 and 2020; Data of 57,195 survey clusters were used for training, which are representative of about 1.2 million households (Pettersson et al., 2023).

The IWI measure, developed by Smits and Steendijk (2015) and building further on the DHS original wealth index (Rutstein, 2015), is a continuous measure from 0 to 100 computed by taking the first principal component from answers to the DHS questionnaire about household access to a TV, refrigerator, phone, bike, car, utensils, and electricity. The measure also includes answers to questions about the quality of water, toilet, and floor facilities in the household, as well as the number of bedrooms (Pettersson et al., 2023; Smits and Steendijk, 2015).

To select neighborhoods, we use the locations of survey clusters from a round of the DHS for each country, as they are a nationally representative random sample of census enumeration areas. They approximate the size of a rural community or urban neighborhood, selected within urban and rural strata (Croft et al., 2018). At the time surveys were conducted, each cluster's GPS coordinates were collected and then randomly displaced for privacy purposes—by up to 2 km in urban areas and 5 km in rural areas, with an additional 1% of rural clusters displaced up to 10 km (Burgert et al., 2013). Excluding 11 DHS cluster locations that lacked adequate data, such as those with duplicate latitude and longitude coordinates or no population density estimates, we were left with 9899 units of analysis. More details and descriptive statistics are provided in Tables DA1 and DA2 in the Data Appendix; see Table 1 for illustration.

Our baseline outcome definition links each DHS cluster to the corresponding 6.7 km × 6.7 km IWI tile. This resolution is broadly comparable to the DHS GPS displacement used for confidentiality (up to 2 km in urban areas and up to 5 km in rural areas), but it does not guarantee that the undisplaced cluster location falls within the same tile. To assess sensitivity to this potential spatial misalignment, we also re-estimate our main models using a coarser outcome that averages the four adjacent IWI tiles (an effective 13.4 km × 13.4 km neighborhood; see Section 5.4).

Table 2
Comparison of China and World Bank aid in Africa from 2002 to 2013.

	China	World Bank
Countries hosting projects count	50	44
Sectors funded	22	13
Aid project count	722	513
Aid project location count	1,373	7,115
Exact locations available (precision 1)	987	4,149
Near (< 25km) locations available (precision 2)	169	258
ADM2 locations available (precision 3)	217	2,708
Portion lacking end date	68%	15%
Portion lacking full funding information	100%	28%
Portion with concurrent, co-located, multi-sector projects	3%	2%

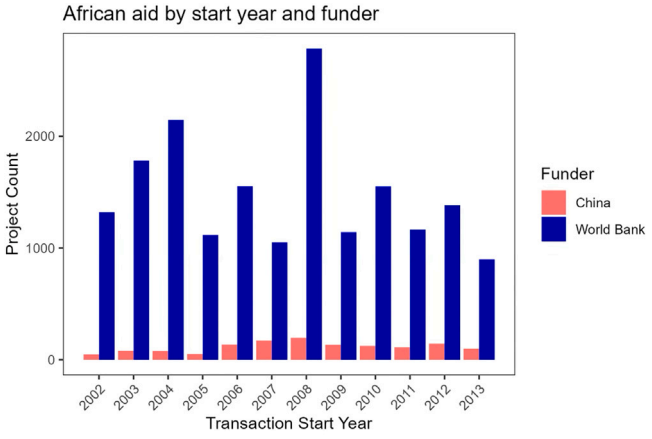


Fig. 3. African aid by start year and funder from 2002–2013.

3. Study setting, data, and measurement

3.1. Treatments: World Bank and China development projects

Having discussed our outcome data in the previous section, we now discuss the treatment. Data on both World Bank and China aid projects were sourced from (Dreher et al., 2021) and curated by AidData. Table 2 shows descriptive statistics, Fig. 3 illustrates annual project counts, and Fig. 4 depicts the sector distribution for the two funders over the study period.[4] The general takeaway is that China funds more projects overall (Table 2), but the World Bank's portfolio spans many more distinct project types and locations.

[4] Aid projects spanning multiple sectors are treated by allocating the full project count to each relevant sector or subsector, without weighting or fractional division based on the number of sectors involved.

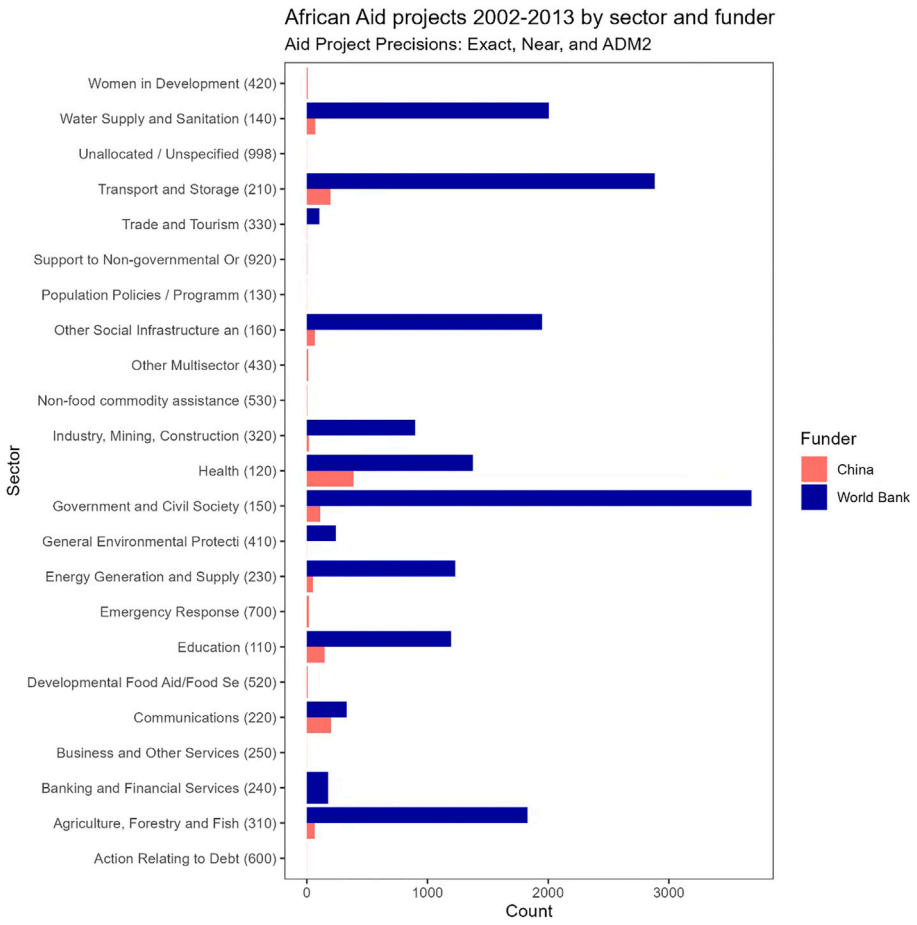


Fig. 4. African aid projects 2002–2013 by sector and funder.

World bank projects. The World Bank makes its activities publicly available; however, to obtain the latitude and longitude of project locations along with a location precision code, we use AidData's version 1.4.2 dataset (AidData, 2017). We use only projects started between 2002 and 2013, as image and confounder data are missing for earlier years, and project data are missing for 2015–2016 in the 3-year group, including 2014 projects. For projects labeled with multiple sectors, we create an observation for each parent sector (e.g., "110-Education"). We limit the data to projects in African countries hosting DHS surveys.

Chinese aid projects. China does not publicly share data about its aid projects, but AidData used the Tracking Underreported Financial Flows (TUFF) methodology to gather this data from four types of open sources: media reports, official statements from Chinese officials and organizations, aid and debt information systems in recipient countries, and research from scholars and non-governmental organizations (Research and Unit, 2017; Strange, 2017). We use version 1.1.1 of their Global Chinese Development Finance dataset.

We exclude pre-2002 and post-2013 projects due to missing satellite and covariate data for earlier periods, as well as missing project data in 2015–2016 for the 3-year group, which includes projects from 2014. We limit projects to African locations using a process described in the Data Appendix. We include projects labeled as "year uncertain" by AidData because they used the year of the first source mentioning the project. We include only projects in completed or implementation stages at the time of dataset publication, whose funding types and objectives met the OECD definition for Official Development Assistance (ODA), making them comparable to World Bank aid. Just like with World Bank aid, we include only locally targeted aid (location precisions 1–3).

Defining treatment/Control and creating panel data. As discussed previously and shown in Fig. 1, to define treatment and control groups and create panel data for each funder and sector, we define a neighborhood unit as part of the treatment group if any of the following three conditions are met: a precision 1 (exact) project lies within the 6.7 km neighborhood boundary, the neighborhood lies within a 25 km buffer around a precision 2 (near) project boundary, or the neighborhood lies within an all-ADM2 (precision 3) project area.[5] If none of these conditions are fulfilled, the unit is assigned to the control group. To avoid confounding at the country selection level, we exclude neighborhoods from the control group in countries where the funder did not conduct any projects in the sector of interest during the study period.

Focusing on these higher-resolution locality resolutions, rather than country- or ADM1-level resolutions, is better aligned with how local aid projects are deployed and their efficiency compared to country assignments (Dreher and Lohmann, 2015; Winters, 2010). Additionally, focusing on local aid enables us to leverage the proxy maps that World Bank and Chinese officials may have been using in allocating aid—unobserved by researchers—and proxied by satellite imagery.

After identifying the treatment status for all neighborhoods in each period, we create panel data of control and treated groups with three-year observations by funder and sector. Fig. 5 maps treated and control neighborhoods for World Bank and Chinese projects. Maps for all funders by sector are available in the Data Appendix.

[5] Note that the ADM2-based rule applies only to precision-3 locations; when AidData reports precision-1 or precision-2 coordinates, we rely solely on the 6.7 km and 25 km square rules and do not treat entire ADM2s.

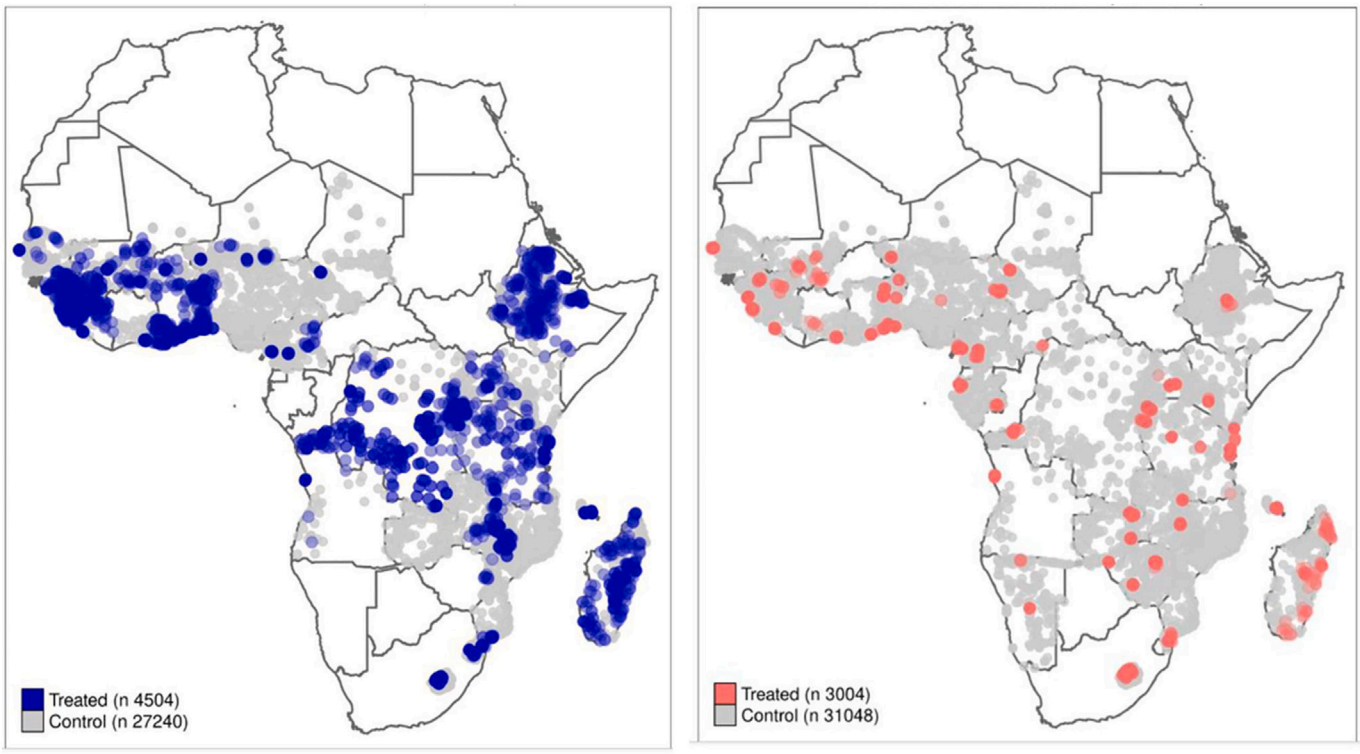


Fig. 5. Geographical distribution of World Bank and Chinese projects. Note that some aid projects cannot be distinguished visually at the continental scale.

We exclude funder-sector combinations with fewer than 80 treated locations, as they lack sufficient data for training and test splits. For the remaining sector/funder combinations, we check for and remove variables that lack variation across the treatment and control groups, as they add little statistical value.

Lastly, we estimate the ATE of each sector and funder's aid one 3-year lag after the 3-year project commitment period. This lag is important because subnational Chinese aid studies found the strongest effects during that period: Dreher et al. (2021) in year 3 and (Xu et al., 2020) two years after commitment.

3.2. Tabular-covariates adjustment set

We provide a large set of tabular confounders that the literature suggests influence both aid placement and effectiveness, variables for year fixed effects to account for common shocks, and ADM2 variables for fixed effects to account for numerous time-invariant factors at that level, including regional history, ethnic composition, and culture. There may be additional unobserved confounders that threaten inference. To avoid introducing post-treatment bias, covariates are measured pre-treatment where possible, making it more likely that they function as good controls (see Cinelli et al. (2024) for a detailed discussion). Tabular variable measurement is summarized in Table 3 and the section below. The Data Appendix contains more details.

Per-capita nightlights. Because previous literature suggests both World Bank and China place projects in the non-poorest sub-national locations (Briggs, 2021; Dreher, 2019) and that projects are most successful in areas with brighter night lights (conditional on population) and above-average education (Briggs, 2021; Greßer and Stadelmann, 2021), we calculate a per-capita nightlight variable to represent a pre-project neighborhood wealth confounder. We use nightlight data from the Defense Meteorological Satellite Program Operational Linescan System (DMSP) and population data from Columbia University's WorldPop Global Project (WorldPop, 2023). For each three-year period, we average the annual per capita nightlights and then add one, taking the log because the data are highly right-skewed.

Because nightlights are persistent and also enter the Pettersson et al. (2023) outcome-construction pipeline, we also conduct robustness checks that drop lagged per-capita nightlights from the adjustment set. Results are reported in Section 5.4.

Population density. Since empirical research suggests that both China and the World Bank allocate aid to more populous subnational regions (Dreher et al., 2022a; Dreher, 2019; Öhler et al., 2019) and population density can facilitate economic opportunity and market access (Scott, 2009), we create an annual average population density variable for the square areas around each unit of observation neighborhood from the WorldPop Population Density dataset (WorldPop, 2023). Since the average population density is highly skewed, we add one and take the log after averaging the value over 3-year periods.

Distance to gold, gems, diamonds, and petroleum. Past studies have found a relationship between project locations and natural resources (Broich, 2017; Guillon and Mathonnat, 2020), and although one study did not (Dreher and Fuchs, 2015), we include as covariates each DHS cluster location's distance to known deposits of four natural resources: gold, gems, diamonds, and petroleum, which are updated over the study period as new deposits are discovered.

We calculate each neighborhood's distance in kilometers from resource deposits known in each period using the following datasets: GOLDDATA 1.2 (Balestri, 2012), GEMSTONE (Flöter et al., 2005), DIADATA (Gilmore, 2005), and PETRODATA (Lujala, 2007). Because the data is highly skewed, we add one and take the log.

Travel minutes to city of over 50,000 population in the year 2000. Because empirical studies of both World Bank and China aid in Africa suggest that funders avoid areas with high travel times to a city (Dreher et al.,

Table 3
Tabular covariates.

Variable	Resolution	Time
Time-varying country-level covariates		
· UN Security Council Temporary Member aligned with U.S. indicator	Country	1 if was a temporary UNSC member and always voted with the U.S. in any of the 3 years prior to the project commitment period, 0 otherwise.
· UN Security Council Temporary Member non-aligned with U.S. indicator	Country	1 if was a temporary UNSC member and voted against the U.S. at least once in any of the 3 years prior to the project commitment period, 0 otherwise.
· Gini inequality measure	Country	3-year average, one period prior to project commitment
· Control of corruption score from annual Worldwide Governance Indicators	Country	3-year average, one period prior to project commitment
· Government effectiveness score from annual Worldwide Governance Indicators	Country	3-year average, one period prior to project commitment
· Political stability and absence of violence/terrorism score from annual Worldwide Governance Indicators	Country	3-year average, one period prior to project commitment
· Regulatory quality score from annual Worldwide Governance Indicators	Country	3-year average, one period prior to project commitment
· Rule of law score from annual Worldwide Governance Indicators	Country	3-year average, one period prior to project commitment
· Voice and accountability score from annual Worldwide Governance Indicators	Country	3-year average, one period prior to project commitment
Time-varying covariates of source spacecraft for composite daytime images		
· Landsat satellite spacecraft category	3-year period	1 for periods 2011 and later, when images are a composite of Landsat 5, 7, and 8. Earlier periods are assigned 0, when images are composites of Landsat 5 and 7.
Fixed effects covariates		
· Year Group	3-year period	First year of group included in each 3-year observation
· Neighborhood's second administrative division (ADM2)	ADM2	Time invariant
Covariates for each neighborhood location		
· Nightlights, per capita, logged	6.7 km	3-year average, one period prior to project commitment
· Population density, average of 1 km squares within neighborhood, logged	6.7 km	3-year average, one period prior to project commitment
· Distance to nearest known deposits of Gold, Gems, Diamonds, and Oil, kilometers, logged	Point-to-point distance	3-year average, one period prior to project commitment
· Travel minutes to city of over 50,000 population in the year 2000, average of ~1 km squares in neighborhood, logged	6.7 km	Time invariant
Time-varying covariates, inherited from neighborhoods' administrative 1, 2, or 3 area		
· Concurrent aid project count (both funders / any sector) in adjacent ADM2s, logged	ADM2	During project commitment period
· Current funder's concurrent project count in other sectors, logged	ADM2	During project commitment period
· Other funder's concurrent project count in all sectors, logged	ADM2	During project commitment period
· Chinese loan-based project count	ADM1 and ADM2	During project commitment period
· Birthplace of country's executive leader in power indicator	ADM1	1 if was leader birthplace in any of the 3 years in pre-project commitment period, 0 otherwise.
· Conflict death count, logged	ADM1	3-year sum, one period prior to project commitment
· Natural disaster count, logged	ADM1, ADM2, and ADM3	3-year sum, one period prior to project commitment
Time-varying covariates, inherited from country		
· Election year for national executive indicator	Country	1 if any of the 3 years in pre-project commitment period were an election year, 0 otherwise.

2022a), and access to an urban area can provide economic opportunities that increase aid effectiveness (Scott, 2009), we calculate each neighborhood's travel minutes to a city. To operationalize this, we use data on the estimated travel time to the nearest city of 50,000 or more inhabitants in the year 2000 (Weiss et al., 2018) to calculate the travel time in minutes for each point within the 6.7 km square area surrounding each unit of observation. We take the average for the square area, add one, and take the log.

Birthplace of leaders in power. Because previous research suggests China gives more aid to provinces in years when they are the birthplace of the country's leaders than in other years (Dreher, 2019), we use tabular, geocoded data on the birth region of political leaders and the years they were in power available from the Political Leaders' Affiliation Database (PLAD) (Bomprezzi et al., 2024). We set the variable to 1 if the national executive leader in power during the annual or 3-year period was born in the neighborhood's ADM1 and to 0 otherwise.

Nearby Chinese loan-based projects. Because China co-locates aid and loan projects to achieve agglomeration effects to increase the economic impact of its aid, we create a count of annual Chinese loan projects in the same ADM1 or ADM2 as the DHS cluster location, from AidData data about Chinese Other Official Flows (OOF) available in their version 1.1.1 dataset (Strange, 2017). Because the data is skewed, we add one and take the log.

Treated other funder count. Because simultaneous aid activities from China influence economic growth in locations of active World Bank projects, and vice versa, we create a variable to count the other funder's simultaneous projects in all sectors within the same ADM2 as the unit of observation, the neighborhood. While it would have been ideal to have a measure of the dollar amounts invested by the other funder, this data is not available for all Chinese projects, so we use a count instead. Because the data is skewed, we add one and take the log.

Other sector project count. Because aid projects the funder does in other sectors can influence the effectiveness of the sector under study, such as transportation projects enabling easier access to Health and Education projects, and because China co-locates these projects to achieve agglomeration effects to improve economic growth, we create a variable to count the funder's projects in other sectors in the same ADM2 as the unit of observation neighborhood. For multi-sector World Bank projects, this variable is incremented once for each sector except the one being studied. While it would have been ideal to measure dollar amounts invested, this data is not available for all Chinese projects, so we use a count instead. Because the data is skewed, we add one and take the log.

Adjacent ADM2 project count. Because the effects of aid projects in adjacent ADM2 regions may spill over to the ADM2 hosting the neighborhood of analysis (Bitzer and Gören, 2018; Demir and Duan, 2024; Xu et al., 2020), we create a variable to count concurrent World Bank and Chinese aid projects in all sectors in all ADM2s adjacent to the one hosting the neighborhood of analysis, including those across country borders. Because the data are skewed, we add one and take the log.

Conflict deaths. Studies find that China reduces aid in response to conflict (Sardoschau and Jarotschkin, 2024) and report mixed results on the World Bank's response to conflict (BenYishay, 2022; Öhler and Nunnenkamp, 2014). Because conflicts can increase the need for aid and the difficulty of projects, we include battle-related deaths from the Uppsala Conflict Data Program's Georeferenced Event Dataset (Croicu and Sundberg, 2015; Gehring, 2022). The variable is a 3-year sum of conflict deaths in the neighborhood's ADM1; because it is skewed, we add one and take the log.

Natural disasters. Because natural disasters impact both the need for aid and a region's aid absorption ability (Drabo, 2021), we create a natural disaster count for each neighborhood and period that includes "floods, storms, earthquakes, volcanic activity, extreme temperatures, landslides, droughts, and (dry) mass movements" (Rosvold and Buhaug, 2021, p. 2). The disaster data is from the International Disaster Database (EM-DAT) (Delforge, 2023), and geocoding is from the Center for International Earth Science Information Network's GDIS dataset (Rosvold and Buhaug, 2021). Because the data is highly right-skewed, we add one and take the log.

Election year indicator. Since empirical research indicates that more political capture of Chinese aid occurs during election years (Dolan and McDade, 2020; Dreher et al., 2022a; Dreher, 2019), we create an indicator of whether national executive elections were held in each period, based on the Database of Political Institution's *Years left in current term* variable, which is set to 0 for election years of national executive leaders.

UN security council temporary membership and voting. Since empirical research suggests both World Bank and Chinese aid allocation are influenced by recipient countries' temporary membership in the UN Security Council and their votes there (Berlin et al., 2023; Dreher et al., 2018; Guillon and Mathonnat, 2020), we create two variables to represent UN Security Council votes and temporary membership. One variable indicates complete alignment with the U.S. position, and the other indicates at least one deviating vote during the period; non-members are assigned a value of 0. This all-or-nothing operationalization approach follows Berlin et al. (2023) and Dreher et al. (2018), who suggest the U.S. may punish a deviating vote by withholding World Bank aid. Neighborhoods inherit variables from their country. The UNSC's temporary membership data were sourced from (Dreher et al., 2009), and the UNSC's voting data were obtained from Dreher et al. (2022b).

Country-level governance indicators. A large body of literature argues that aid allocation and effectiveness depend on the quality of recipient country institutions (see Asatullaeva (2021) for a review). To capture multiple dimensions of recipient country governance and institutions, we use six country-level, annual variables from the World Bank's Worldwide Governance Indicators (Kaufmann and Kraay, 2024): Control of Corruption, Government Effectiveness, Political Stability and Absence of Violence/Terrorism, Regulatory Quality, Rule of Law, and Voice and Accountability. Each neighborhood inherits the average score of its country each period, which is roughly normally distributed between -2.5 and 2.5.

Country-level inequality. Since previous research suggests that less income inequality is associated with better aid effectiveness (Skarda, 2022), we include a measure of country-level inequality, as country-level inequality, compared to within ADM2 level inequality, captures the broader institutional and policy environment that influences both aid allocation and the outcomes of development interventions (Chong et al., 2009). Each three-year period, each neighborhood inherits the average of the three national annual Gini inequality scores from the World Bank's World Development Indicators dataset (World Bank, 2023), which takes on values between 0 and 1.

3.3. Daytime satellite image confounder details

To proxy remaining confounders, we use daytime satellite images of approximately 5 km square centered over each neighborhood's 6.7 km square center. The images came from the U.S. Geological Survey Landsat 5, 7, and 8 satellite series, but since Landsat 5 data were not saved over large portions of equatorial Africa due to lack of commercial data buyers when it was generated (NASA, 2021), we limit our study years to exclude projects started prior to 2002, where only Landsat 5 images are available.

Due to clouding and since no single Landsat series covers all study years (Landsat 7 images taken on and after May 31, 2003, also have diagonal lines across them due to a Scan Line Corrector (SLC) hardware failure (Science, 2021)), we create composite images from the median pixel of all available satellites for each period. This per-pixel median mosaic across sensors and dates mitigates the Landsat-7 SLC-off striping by infilling the missing scan lines with valid pixels from overlapping acquisitions. For each cluster, we build fixed, non-overlapping three-year composites—1999–2001, 2002–2004, 2005–2007, 2008–2010, and 2011–2013—taking the per-pixel median of all available, cloud-masked Landsat 5/7/8 scenes in the window; these windows are identical across units and do not depend on treatment timing. We use visible light bands (red, green, blue) plus near-infrared and short-wave infrared spectral bands to capture information about built-up areas. More details are available in the Data Appendix.

Causal inference methods require overlap between the characteristics of treated and control units. While adding many covariates helps control confoundedness, including too many covariates can reduce the overlap needed to make causal claims. This is especially true for high-dimensional data (D'Amour et al., 2021), such as satellite images, which are unique across different locations. Our approach addresses this by

creating a lower-dimensional representation of satellite images that captures the key features and characteristics of neighborhoods found across various geographic locations.

4. Estimation method: image-based confounding analysis

Part of our analysis relies on a computer-vision-based causal-adjustment method for processing satellite images that we developed elsewhere in Jerzak (2023a,b). We use Vision Transformers, a computer vision machine learning algorithm, for our primary results, which rely on ViT, because this model is generally considered state-of-the-art (but we also report results for tabular-only, and later, unit fixed effects, analysis).

Relative to this earlier work, our contribution in this paper is to deploy the EO-ML adjustment framework in an applied setting—a multi-funder, multi-sector evaluation—and to add diagnostic layers. In particular, we (i) estimate a set of nested specifications (diff-in-means, X + FE, M, M + X + FE) to quantify how much imagery changes estimated ATEs, and (ii) compute assignment-mechanism diagnostics (out-of-sample AUCs and tabular-covariate salience measures) that characterize how donors condition on visible and tabular features. We treat these steps as methodological validation and refinement rather than as a new estimator.

Fig. 6 depicts the key intuition behind our causal-ViT method. The modeling consists of three steps: fusing image and tabular data, processing these data via transformer encoding to estimate the propensity scores, and, finally, plugging those estimates, via cross-fitting, into an inverse probability weighted mean difference estimator of the ATE. The primary objective of the propensity estimation stage is to process images and tabular data to refine our estimation of the propensity weights, which represent the probability of selection into development programs. Our hypothesis is that these weights will better balance our units than tabular weights alone in calculating the mean difference between treated and control units.

Data-fusion step. After assembling the data, we prepare them for modeling. Besides the usual data curation, this preparation entails assigning all the tabular data in the set $\mathbf{X}_{i(t-1)}$ and the images in the set $\mathbf{M}_{i(\bar{t}-1)}$. The approach here divides each image i into smaller patches for the model to process efficiently; we use a 16×16 patch size. Then we flatten the patches, preparing them for injection into the model. The tabular data has a different dimensionality, and we must therefore project it to the embedding dimensionality. To this, we also append a summary representation vector called a *classification token* (typically denoted CLS). The data streams are now ready for fusion by concatenating the patch, CLS, and (when used) tabular representations.

To summarize, while the tabular covariates contain information about a variety of political, economic, and geographical factors not visible through remote observation, the satellite images contain visible factors relevant for confounding between aid and the wealth outcome. By fusing both sources, the model will learn how to combine the importance of all these factors jointly.

Propensity-estimation step. The goals of the propensity-estimation step are to predict treatment status using the images and the supplied tabular data. This step unfolds as follows. It scales and zero-centers the data, and then appends the image pixel values to the tabular confounding variables, creating high-dimensional embeddings where the number of covariates is significantly greater than the number of observations. It regularizes the data to avoid overfitting. Then, these embeddings pass through our computer vision method to learn informative summaries of the raw image and tabular variables, enabling comparison between key features of treated and control neighborhoods.

Our primary EO-ML method is the ViT, and thus passes these embeddings through a transformer encoder. There are many insightful introductions to transformer models such as Prince (2023), including the original article by Vaswani et al. (2017). However, here we provide a high-level intuition for this encoder.

ViTs capture global context by modeling long-range dependencies across image patches via self-attention, which allows them to effectively integrate spatially distributed features in satellite images. Readers can think of self-attention as a statistical association model that captures relationships between each image patch with all others in an embedding space. In our Fig. 6, this is visualized through the Multi-head self-attention block. It is *multi-head* because the self-attention mechanism is conducted multiple times independently, ensuring that some heads capture information that other heads might have missed. This information is then passed through a normalization block and subsequently through parallel neural-network layers, allowing for further learning on how the image and tabular data are related to each other and how well they estimate treatment versus control class probabilities. In this strategy, we employ an 8-layer model with dropout and drop path, both at 10%; (Huang et al., 2016), and balanced training to improve convergence in the context of imbalanced treatment/control classes (Francazi et al., 2023).

The end-to-end trained ViT approach uses a random sample of 90% of the ADM2 areas to train the model, employing stochastic gradient descent to maximize the model's log likelihood, updating model weights

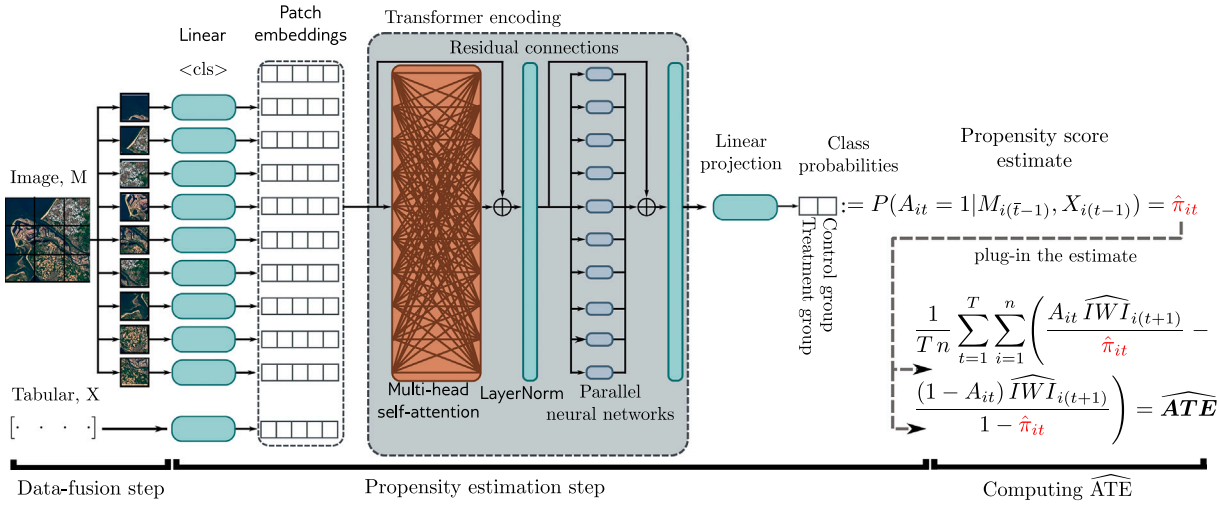


Fig. 6. A stylized visualization of the flow of data fusion, transformer architecture for propensity estimation, and the outcome model. The transformer-block visualization is inspired by Prince (2023). The plug-in estimate, $\widehat{\text{ATE}}$, uses out-of-fold indices via cross-fitting, breaking reuse of the same data for training and estimation, and limiting regularization bias.

using backpropagation. We employ a cross-fitting approach, using out-of-fold predicted treatment/control probabilities to improve inference and generate uncertainty estimates (Chernozhukov et al., 2018).

Computing $\widehat{ATE}$. Once the transformer finishes learning the class probabilities, we plug them into an Inverse Probability weighting method with the Hájek adjustment, normalizing the observation weights to sum to one. As shown in the following formulae, the methods used here rely on estimating the ATE by taking a mean difference in the expected value of the post-aid International Wealth Index estimates between neighborhoods that did and did not receive aid, adjusting for pre-project satellite imagery over treatment and control locations, along with tabular data about each location's treatment status and confounding variables. Formally, this estimation process contains two stages, as follows:

$$\hat{\pi}_{it} = \hat{P}\left(A_{it} = 1 \mid \mathbf{M}_{i(\tilde{t}-1)}, \mathbf{X}_{i(t-1)}\right) = f\left(\mathbf{M}_{i(\tilde{t}-1)}, \mathbf{X}_{i(t-1)}\right) \tag{1}$$

$$\widehat{\mathrm{ATE}} = \frac{1}{T\,n}\sum_{t=1}^{T}\sum_{i=1}^{n}\left(\frac{A_{it}\,\widehat{\mathrm{IWI}}_{i(t+1)}}{\hat{\pi}_{it}} - \frac{(1-A_{it})\,\widehat{\mathrm{IWI}}_{i(t+1)}}{1-\hat{\pi}_{it}}\right) \tag{2}$$

where

- $\hat{\pi}_{it} = \hat{P}\left(A_{it} = 1 \mid \mathbf{M}_{i(\tilde{t}-1)}, \mathbf{X}_{i(t-1)}\right)$ is the estimated treatment probability of neighborhood i at project commitment time.
- $f(\cdot)$ is the image model for the predicted treatment probabilities.
- $\mathbf{M}_{i(\tilde{t}-1)}$ is a 5 km square daytime satellite image centered around neighborhood i's location, one lag prior to project commitment. It is a cloudless composite image spanning a 3-year period; for example, project commitments between 2002 and 2004 are based on images from 1999 to 2001.
- $\mathbf{X}_{i(t-1)}$ are tabular confounders in neighborhood i, one lag prior to project commitment, including ADM2 and year fixed effects. These are summarized in Table 3 and in the Tabular Covariates section below.
- $\widehat{\mathrm{ATE}}$ is the estimated average treatment effect.
- n is the total number of unique neighborhoods for the specific funder and sector.
- T is the total number of time periods.
- A_{it} is the observed treatment status for neighborhood i in period t for the specific funder and sector.
- $\widehat{\mathrm{IWI}}_{i(t+1)}$ is the estimated International Wealth Index in community i (ranges from 0–100), measured in a 3-year composite estimate one lag after aid project commitment. For project commitments between 2002 and 2004, for example, IWI is measured between 2005 and 2007.

(Note that Hájek adjustment modifies this baseline estimator slightly by normalizing the weights for improved efficiency (Freedman and Berk, 2008).

5. Results

First, we discuss the overlap of treatment propensity scores between treated and control neighborhoods for the two funders, then we present the estimated ATEs by sector. Next, we present model quality measures and a summary of each tabular covariate's salience to aid allocation decisions across funders and sectors.

5.1. Average treatment effect by funder and sector

Using our ViT, we estimate ATEs for each funder and report four specifications that correspond to Fig. 7: (a) a difference-in-means baseline, (b) tabular covariates with ADM2 fixed effects (X + FE), (c) imagery only (M), and (d) imagery plus tabular covariates and ADM2 fixed effects (M + X + FE). This subsection also provides our main methodological benchmark: by comparing diff-in-means, tabular-only (X + FE), imagery-only (M), and imagery-plus-tabular (M + X + FE) estimators, we assess how much image-based confounding control changes estimated ATEs relative to standard approaches.

Because Fig. 7 juxtaposes several estimation regimes, a natural question is which one should be treated as the primary causal estimate. Our preferred adjustment set is the fully image-augmented specification (d), which conditions on pre-treatment satellite imagery—intended to proxy the map-based information donors use when selecting project sites—while also incorporating the standard tabular covariates and ADM2 fixed effects that absorb time-invariant local context.

Across many sectors, adding imagery (c and d) shifts ATEs downward relative to tabular-only or unadjusted models, indicating that satellite images capture additional confounding information that standard covariates may miss. In general, Chinese projects display larger effects than World Bank projects and sectoral heterogeneity. Notably, across sectors, the imagery-only estimates (c) are similar to the imagery + X + FE estimates (d), indicating that pre-treatment satellite features absorb most of the selection-relevant variation that ADM2 fixed effects and standard covariates were proxying.

Relative to prior literature on the World Bank, our image-adjusted estimates are more conservative. Under d, we see clearly positive effects in *Education (110), Transport and Storage (210), General Environmental Protection (410)*, and *Trade and Tourism (330)*. Several other sectors are small or have confidence intervals that include or are nearly zero—most notably *Communications (220), Banking and Financial Services (240), Energy Generation and Supply (230), Agriculture, Forestry and Fishing (310),* and especially *Industry, Mining, Construction (320)*. These patterns are broadly consistent with Bitzer and Gören (2018) and Xu et al. (2020) but are generally smaller after image adjustment; in contrast to Bitzer and Gören (2018), we do find positive effects for *Education* under most model specifications.

Turning to Chinese aid, results remain large and positive in social sectors such as *Education, Health, Water Supply and Sanitation,* and *Government and Civil Society*, and in economic infrastructure, with *Emergency Response (700)* yielding the largest effects. Smaller image-adjusted effects appear in *Energy Generation and Supply (230)* and *Agriculture, Forestry and Fishing (310)*, which plausibly have longer gestation and slower transmission into the IWI asset bundle.

We next look across all estimation strategies, examining the maximum and minimum ATE estimates across sectors (taking median across strategies). For the World Bank, the maximum (median) ATE is for sector *Trade and Tourism (330)*, with a value of 12.29; the minimum ATE is for sector *Government and Civil Society (150)*, with a value of −0.16. For China, the maximum median ATE is for sector *Emergency Response (700)*, with a value of 15.15; the minimum is for sector *Agriculture, Forestry and Fishing (310)*, with a value of 1.21.

These sector extremes align with differences in time-to-impact and what the asset-based outcome captures. For the World Bank, *Trade & Tourism (330)* plausibly operates through market-access and service-sector channels that can translate relatively quickly into higher earnings and durable goods (Donaldson and Storeygard, 2016; Scott, 2009). By contrast, sectors like *Government and Civil Society (150)* or *Agriculture, Forestry and Fishing (310)* might entail long build-out cycles or require significant preparatory investments before effects materialize, yielding smaller near-term IWI gains. On the China side, large *Emergency Response (700)* estimates are consistent with rapid reconstruction, injecting resources and public works (Cavallo and Noy, 2009). These interpretations are descriptive and do not rule out targeting dynamics or residual confounding.

It is also worth briefly discussing the few consistently negative estimated treatment effects—for example, World Bank's *Government and Civil Society (150)* interventions sector just discussed. Possibly, these negative or near-zero estimates might reflect how such projects may target weaker or more institutionally fragile areas, where aid is harder to translate into short-term asset gains, so our conservative image-adjusted models capture selection into more challenging environments rather than genuine harm. That said, the within-neighborhood fixed-effects

Fig. 7. ATEs by sector, funder, and modeling strategy. "FE" refers to ADM2 fixed effects; specifications: (a) Diff-in-means, (b) X + FE, (c) M, (d) M + X + FE.

robustness check in Section 5.4 yields similarly small (and sometimes negative) point estimates with wide confidence intervals for these two sectors, suggesting more limited within-unit evidence of short-run asset gains. These cases merit further investigation.

5.2. Explaining variation in ATE estimates

To summarize how different modeling ingredients affect the estimates, we pool all funder–sector results and regress the *ATE level* on simple indicators for whether a given specification includes tabular covariates (X), ADM2 fixed effects (FE), and satellite imagery (M), allowing separate intercepts by funder. This provides a compact diagnostic of the extent to which the EO–ML component contributes beyond conventional tabular adjustment. Table 4 reports two parsimonious models. We report both because Model 1 provides an interpretable marginal-effects benchmark for X, FE, and M, while Model 2 relaxes additivity by estimating distinct combination effects (e.g., $M{+}X{+}FE$), serving as a robustness check that the imagery-driven shrinkage in ATEs persists.

Several patterns emerge. First, specifications that incorporate imagery (M) often produce *smaller* ATEs than otherwise-comparable models that omit imagery. This is the strongest and most stable association in the table and is consistent with our central claim: images proxy the map-based information donors use when allocating projects, improving confounding control and shrinking observational effects toward zero. Second, adding ADM2 fixed effects (FE) is associated with slightly lower absolute ATEs on average (see robustness section for an individual-level FE analysis), whereas including X alone has little systematic association with the ATE once other ingredients are accounted for. Third, these relationships hold when we allow mean differences by funder; the China mean ATE exceeds the World Bank mean ATE in the pooled regressions, echoing the sector-level results above.

Overall fit is high (adjusted R^2 is 0.71 and 0.73), indicating that these simple model ingredients explain most of the across-specification variation. Any model variant containing M (e.g., M, $M{+}X$, $M{+}FE$, $M{+}X{+}FE$) sits below the tabular-only benchmark, reinforcing that the downward shift is broad-based rather than driven by a single

Table 4
Outcome: Absolute ATE estimates. Estimator: Weighted OLS (inverse variance weights). t-values in parentheses, $*$ indicates $p < 0.05$.

	Model 1	Model 2
Funder: China	13.39 (13.22)*	15.66 (11.55)*
Funder: WB	6.74 (8.46)*	9.05 (7.11)*
Contains X	−1.55 (−2.15)*	
Contains FE	−1.15 (−1.59)	
Contains M	−2.75 (−3.84)*	
X		−5.08 (−3.03)*
FE		−4.92 (−3.25)*
X + FE		−4.72 (−2.82)*
M		−6.89 (−4.60)*
M + X		−6.05 (−3.80)*
M + FE		−5.35 (−3.42)*
M + X + FE		−7.30 (−4.85)*
Other statistics		
Observations	200	200
Adjusted R-squared	0.71	0.73
Outcome	ATE	ATE

specification. In short, imagery-based controls consistently make the estimates more conservative, which is exactly what we would expect if they are capturing selection-relevant features that are missing from standard tabular sets.

5.3. Insights about aid assignment mechanisms

Having assessed the robustness of our impact estimates, we now turn to the assignment mechanism, using out-of-sample prediction performance to quantify how systematically each donor targets neighborhoods. Fig. 8 summarizes out-of-sample Area Under the Curves (AUC) for predicting which neighborhoods receive projects. AUC values of 0.50 indicate predictive performance no better than random guessing; 0.70 means that, in 70% of randomly drawn treated–control pairs, the model ranks the actually treated neighborhood higher; values closer to 1 indicate highly predictable placement, based on the data and modeling strategy.

Two broad patterns emerge. First, placement is meaningfully *non-random* for both funders: models that use satellite imagery alone (M) already separate treated from control units well above the random-guessing baseline in most cases, and adding tabular covariates with ADM2 fixed effects (M + X + FE) yields small, consistent improvements. This supports our core identification strategy—image features proxy the map-based information officials use when deciding where to place projects.

Second, there is sectoral heterogeneity in treatment predictability. Projects with physical requirements often show high levels of predictability using image covariates (e.g., mining), while some social sectors show more moderate separability (Education (110) is an exception). This is the type of variation we would expect if imagery and tabular variables together capture on-the-ground constraints and opportunities that are highly non-uniform across settings.

Table 5 helps quantify apparent differences between funders. In simple comparisons, World Bank sectors modestly appear more predictable than Chinese sectors. Overall, variability in AUC values is largely compositional, being driven by data volume and treatment/control balance, and possibly somewhat by intrinsic differences in assignment opacity between World Bank and Chinese funders.

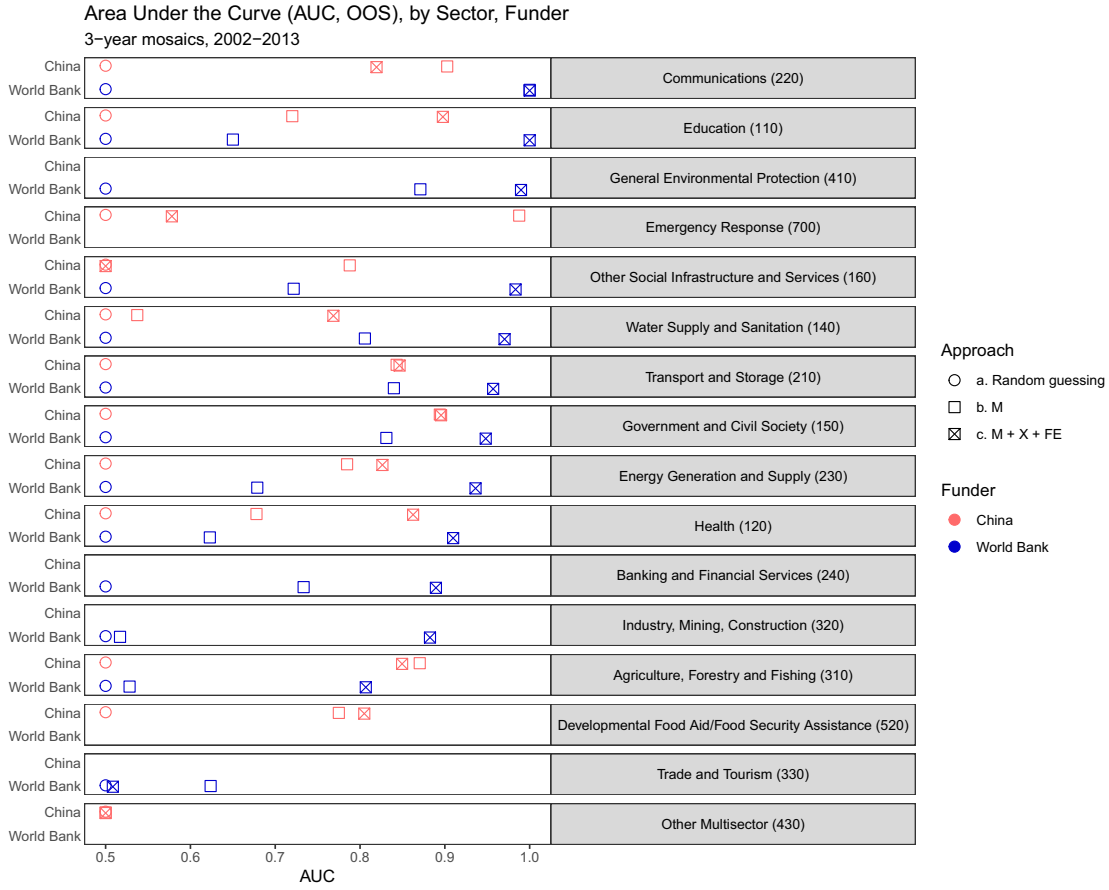


Fig. 8. Out-of-sample area under the curve (AUC), a metric for evaluating model prediction quality, for each funder, sector, and analysis backbone. Values closer to one are better. Here, FE refers to ADM2 fixed effects. Out-of-sample observations come from ADM2 areas not used in model training.

Table 5
Outcome: AUC. Baseline funder: China. Estimator: OLS. t-values in parentheses, $*$ indicates p <0.05.

	Model 1	Model 2	Model 3
WB Funder	0.08 (2.63)*		0.08 (2.40)*
log(N)		0.17 (4.85)*	0.17 (4.81)*
Class imbalance		0.40 (2.36)*	0.61 (2.97)*
Other statistics			
Intercept	0.77 (30.42)*	−1.23 (−2.75)*	−1.38 (−3.04)*
Observations	175	175	175
Adjusted R-squared	0.03	0.18	0.20
Outcome	AUC	AUC	AUC

Tabular variable salience. Fig. A.I.4 in the Appendix summarizes the relationship between tabular covariates and project placement in the tabular–only model ($X+FE$) and the image–augmented model ($M+X+FE$). As our modeling strategy involves non-linear operations (e.g., Transformers), we cannot simply look at model parameters, as we can with linear models; rather, we need to compute the sensitivity of the model's predicted treatment probability to changes in each tabular covariate, averaged across all observations, using automatic differentiation (a computational process that efficiently computes the exact derivatives of complex functions by applying the chain rule to a sequence of elementary arithmetic operations).

In these salience figures the y-axis lists the tabular variables (one row per covariate) and the x-axis lists sectors. Each cell reports a summary of how much a given covariate moves the model's predicted probability of treatment for that funder–sector. In Fig. A.I.4 (Appendix), the values are *signed* sensitivities: positive values mean that increasing the covariate is associated with a higher predicted chance of a project, while negative values mean the opposite. In Fig. 9 (here), each cell shows the *change in absolute salience* when images are added, |Salience($M+X+FE$)| − |Salience($X+FE$)|, so red cells indicate that a variable becomes more influential once images are included, blue cells indicate it matters less, and white indicates little change (or too few units for that funder–sector column). Because inputs are standardized before modeling, salience magnitudes are on a comparable scale across variables.

We highlight several observations.

First, adding images often shrinks the absolute salience of many widely used spatial proxies—population density, market access (minutes to city), and country–level governance scores—especially for the World Bank (blue panels in Fig. 9 are common). This is consistent with the identification claim: imagery carries geospatial structure that standard tabular sets only approximate, so once we condition on images, those variables matter less.

Second, changes for China are more heterogeneous across sectors. In several funder–sector cells, event- or politics-style variables (e.g., natural disasters, leader birthplace, or UNSC voting) become relatively *more* salient once images absorb geography (red cells in Fig. 9). Qualitatively, this suggests that after we control for what is visible from space, the model reallocates attention toward non-visual drivers of Chinese project placement in specific sectors.

Third, images mainly re-weight variables rather than flip their signs: the direction of association visible in Fig. A.I.4 is largely stable before vs. after imaging, but magnitudes compress. Substantively, that means imagery helps separate "where projects can feasibly go" (geography and the built environment) from "why they go there" (political and programmatic factors).

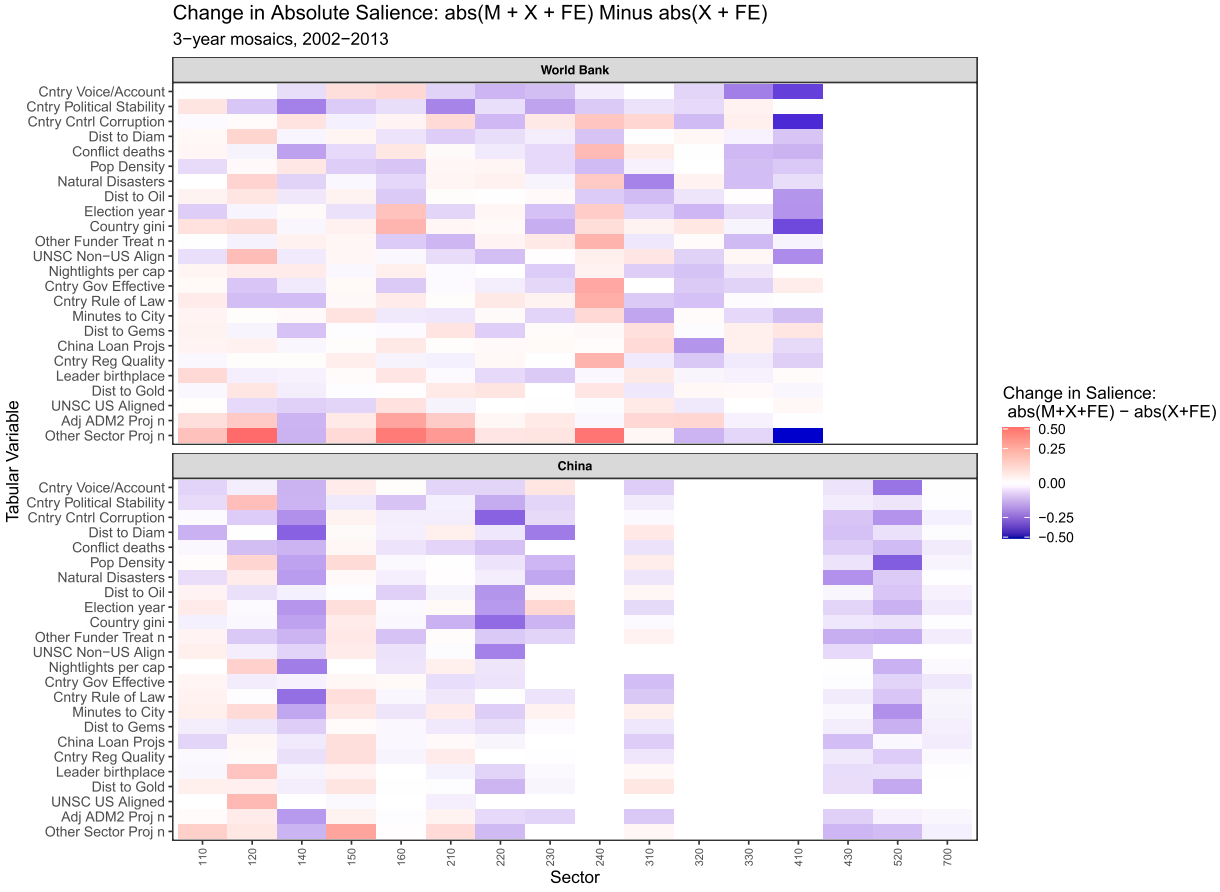


Fig. 9. Change in the absolute salience of tabular variables attributable to tabular confounders, before and after conditioning on image information, across sectors for World Bank and China. Red indicates increased salience magnitude when images are included; blue indicates decreased salience magnitude, and white indicates minimal change. Some columns are left white where there are not enough units to perform analysis for that funder-sector combination.

To gauge cross-funder similarity in these salience profiles, we perform Canonical Correlation Analysis (CCA) on the sector–by–covariate salience matrices from the image-augmented model. CCA takes two multivariate datasets measured on the same features—in our case, for each sector we have a vector of saliences over tabular variables for the World Bank and another for China—and finds pairs of linear combinations (one per funder) that are maximally correlated across sectors. We treat sectors as observations and covariate saliences as features (columns are standardized so that scale differences do not drive the results). The *leading canonical correlation* summarizes how similar the two donors' *multivariate* emphasis patterns are across sectors: values near 1 imply that the donors load on similar mixtures of covariates; values near 0 imply largely orthogonal emphasis. This method is appropriate here because it compares full profiles without forcing us to pick one covariate at a time, it is invariant to linear rescaling, and it handles collinearity by working with linear combinations rather than individual variables. The leading canonical correlation here is 0.182, indicating limited alignment between Chinese and World Bank placement logics once geography is accounted for. In short: the two donors emphasize different combinations of factors across sectors.

We do not interpret these salience scores causally; they are descriptive of the assignment mechanism under each adjustment set. Their main value here is to show *why* image adjustment reduces ATE estimates: images soak up spatial structure that tabular proxies were previously carrying, yielding more conservative (and, we argue, more credible) effect estimates.

5.4. Robustness and alternative identification strategies

The main estimates in Fig. 7 rely on selection-on-observables after conditioning on pre-treatment imagery and tabular covariates. We run four complementary robustness checks that probe whether the qualitative conclusions depend on (i) identifying variation only within neighborhoods over time, (ii) including lagged nightlights in the covariate set, (iii) the spatial scale of the IWI outcome grid (motivated by DHS coordinate displacement), and (iv) our treatment-footprint rules for imprecisely geocoded projects. Detailed figures and tables are reported in the Results Appendix.

Within-neighborhood (unit) fixed effects. To reduce reliance on selection on observables, we estimate a two-way fixed-effects model that uses only within-neighborhood changes in exposure among treatment switchers:

$$Y_{it} = \alpha_i + \lambda_t + \beta A_{it} + \varepsilon_{it},$$

where Y_{it} is IWI (measured one period after commitment), A_{it} indicates sector–funder treatment in period t, α_i are neighborhood fixed effects, and λ_t are period fixed effects. Standard errors are clustered at the ADM2 level. Across Chinese projects, point estimates are directionally consistent with the IPW results (> 0) but attenuated in magnitude and less precise (see Fig. 10), reflecting the limited number of switchers. World Bank effects more often switch signs in this unit-level fixed effects analysis. For the World Bank sector with the most consistently negative point estimates (*Government and Civil Society (150)*), the corresponding fixed-effects estimate is closer to zero and imprecise (-0.05, SE $= 0.11$). We interpret these results cautiously, pending a more rigorous investigation into switcher dynamics and within-sector heterogeneity that would complicate causal interpretation.

Robustness to lagged nightlights in the adjustment set. Because lagged per-capita nightlights are persistent and also enter the IWI construction pipeline, we re-estimate the main ViT M + X + FE specifications after removing lagged nightlights from $\mathbf{X}_{i(t-1)}$. Estimates remain qualitatively consistent: across the 25 funder–sector cells, the median absolute ATE change is 2.05 IWI points and the maximum absolute change is 7.38. The correlation between baseline and no-nightlights ATEs is 0.85.

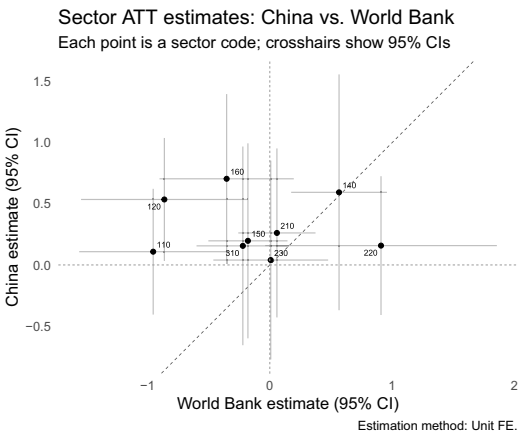


Fig. 10. Unit-level fixed effects results.

Robustness to neighborhood size. To assess sensitivity to the 6.7 km × 6.7 km outcome definition—and the possibility that DHS displacement moves some clusters toward neighborhood edges—we construct a coarser IWI outcome by averaging the four adjacent 6.7 km tiles around each cluster (effective 13.4 km neighborhoods) and re-estimate the main models. ATEs show high but slightly lower correlation with the baseline specification (correlation: 0.60) compared to the nightlight robustness; the qualitative ordering of effects across funders and sectors is broadly unchanged.

Robustness to treatment definition. Finally, we tighten the treatment definition by excluding ADM2-wide assignments that arise from imprecisely geocoded (precision-3) projects and restricting treatment to precision-1 and precision-2 locations. Fig. A.I.7 shows that effects remain broadly consistent with the main analysis; baseline and strict-precision ATEs are correlated (correlation: 0.61); see the Appendix. That said, World Bank estimates are notably higher in magnitude in this regime.

Overall, these robustness tests suggest that conclusions from the main analysis about the relative ranking of sector effects and the relative effects of Chinese vs. World Bank programs are not generally sensitive to modeling choices about nightlights, outcome resolution, or treatment-definition rules. Choice of treatment and especially identification strategy have greatest consequence for the magnitude and precision of sector-specific estimates. Broadly speaking, across identification strategy and treatment/outcome definition, Chinese programs are still estimated here to have greater effects on living conditions.

6. Discussion

Researchers continually seek to improve the causal understanding of the impact of local development projects on individuals' health and living conditions (Imbens and Rubin, 2015). In the case of the World Bank and China, there is a lack of localized tabular data on the geographical and historical maps that officials used to strategically decide on where to allocate their projects across Africa in the past two decades (Bedi et al., 2007a). Our study applies and extends a recently proposed EO–ML method that proxies these maps with the help of satellite images by Jerzak (2023a,b). The contribution here is less about inventing a new estimator and more about demonstrating, in a large-scale applied setting, how image-based confounder adjustment changes estimated impacts and what it reveals about assignment mechanisms. Another innovation here is that we use a novel subnational dataset on health and living conditions recently produced by Pettersson et al. (2023). Combining these two advances, this study analyzes the effect of locally targeted World Bank and Chinese development projects. The study includes a

sector-specific analysis, yielding findings roughly comparable to those of the subnational studies. Our use of nearly 10,000 neighborhoods as potential treated and control observations for each funder and sector analysis enables individual-sector-level analysis, making visible significant sector heterogeneity. Our findings are also robust to dropping lagged nightlights from the covariate set and to substituting alternative proxies for baseline economic conditions, alleviating concerns about mechanically re-using nightlights in both outcome construction and confounder adjustment. These results advance the aid effectiveness debate by addressing some of the challenges facing subnational development studies. Next, we discuss potential implications.

Hypothesizing potential mechanisms driving the results: why and where beneficial effects. A finding of this analysis is that, on average, Chinese aid programs demonstrate a larger positive impact on our machine-learning-derived wealth index than World Bank programs across many comparable sectors (see Fig. 7). This divergence in effect size can likely be attributed to the differing mechanisms through which each donor's project portfolio and implementation style influence local economies. The literature notes that Chinese aid is heavily concentrated in physical infrastructure sectors, such as transportation, energy, and construction (Brautigam, 2009; Dreher et al., 2018).

The mechanism for these projects is often direct: construction creates local, albeit often temporary, employment, injecting cash into the community. Upon completion, infrastructure such as roads and bridges can reduce transportation costs and better connect communities to larger markets, leading to a more widespread diffusion of economic activity (Bluhm, 2020). This enhanced connectivity can boost agricultural incomes, create new opportunities for small businesses, and increase labor mobility, all of which generate household income that can then be used to purchase assets and make home improvements, as captured by the International Wealth Index (IWI). While this pathway explains the large effects in sectors like "Transport and Storage," our study also finds significant positive effects for Chinese projects in social sectors such as "Health" and "Education" (Woods, 2008). This may be due to China's implementation model, which often bypasses recipient government systems in a "demand-driven" process that may result in the faster delivery of tangible goods and services (Martuscelli, 2020).

In contrast, the World Bank's projects yield smaller, though still broadly positive and often statistically significant, effects on the IWI. The Bank's official strategy involves working through recipient government systems to build long-term capacity, a process codified by the Paris Declaration on Aid Effectiveness (OECD, 2005). The causal mechanisms for projects in sectors like "Government and Civil Society" or "Other Social Infrastructure" are therefore more indirect. They aim to improve the quality and accountability of public services, effects that are often slow to materialize and may not translate into measurable household wealth within our study's timeframe. Even for a sector with a direct link to the IWI, like "Water Supply and Sanitation," the World Bank's process-oriented approach may be slower than China's more direct implementation (Isaksson and Kotsadam, 2018). The weaker performance of World Bank aid in "Industry, Mining, Construction" and "Agriculture," as found in our analysis, further suggests its model may be less suited to generating the kind of immediate, capital-intensive impact that our wealth measure captures effectively (Ssozi et al., 2019).

This distinction between China's more direct, often faster implementation style and the World Bank's more indirect, institution-building approach offers a compelling explanation for our findings. The "hard" and "fast" nature of Chinese aid appears to generate more immediate and tangible improvements in material living conditions as measured by satellite imagery and the IWI. The World Bank's approach, while potentially laying the groundwork for more sustainable, long-term development, produces smaller and less immediate wealth effects. Our study's ability to detect this heterogeneity underscores that the "how" of development aid is just as critical as the "how much" (Easterly, 2003).

Additionally, Pettersson et al. (2023) wealth estimates overcome measurement limitations in nightlight data used to proxy subnational wealth, address the lack of panel data in Demographic and Health Surveys, and enable the use of fixed effects at the ADM2 (county, district, city) level to control for unobserved, time-invariant confounders in subnational regions.

Moreover, our use of satellite-image deconfounding methods for causal inference yielded overall more conservative estimates compared to tabular methods (with fixed effects) only. This empirical evidence suggests that our EO-ML method effectively adjusts for confounding information not available in existing tabular sources. Consequently, this implies that previous studies are likely overestimating the beneficial effects of development projects when they fail to adjust for image features.

The image-based confounding estimation method demonstrated a wide range of AUC model quality scores in predicting treatment probability across funders, sectors, and modeling strategies. We expected image-based confounding to be more beneficial for sectors dependent on the physical conditions of treatment areas compared to those that can be placed anywhere (e.g., Energy Generation vs. Education), but did not find a clear pattern meeting those expectations.

Data resolution is a concern in studies using satellite methods and data. If the resolution of satellite images is not sufficiently high, the algorithm may miss confounding factors. The multi-resolution nature of our data raises several challenges to statistical inference (see Meng (2014)), including issues related to mixing data measured at multiple resolutions and estimating an estimand at a finer-grained resolution (6.7 km neighborhoods) than many of the covariates upon which it is based. To ensure the use of nationally representative samples in each country included in each funder/sector analysis, we use locations of neighborhoods surveyed in one round of Demographic and Health Surveys as units of analysis. We use Pettersson et al. (2023)'s wealth estimates for those locations over time. This is a non-random sample, since all these countries have a relationship with the U.S. Future research could use wealth estimates to create broader samples of neighborhoods across the African continent, including in countries never surveyed by DHS, or could include all human settlement areas across Africa instead of a sample. With the possible dismantling of funds to USAID, and its detrimental effect on future DHS data collection efforts, EO-ML methods such as Pettersson et al. (2023) will be even more critical for aid evaluation.

Future work. There are several promising angles that future research can build on. For example, future research can replicate this study incorporating novel debiasing methods in the handling of ML-generated outcomes (Angelopoulos et al., 2023; Pettersson et al., 2025; Rister Portinari Maranca et al., 2025), uncertainty quantification in the imputed-image-derived outcomes (Kakooei and Daoud, 2024), evaluating the effect of satellite image scale (Zhu et al., 2025), and incorporating large-language models for enhanced interpretability (Murugaboopathy et al., 2025).

The fact that project effects spill over into adjacent ADM2s violates the non-interference requirement of the stable unit treatment value assumption (SUTVA) and creates an ongoing challenge for this and other studies. Future work can also employ techniques (e.g., excluding neighborhoods adjacent to treated ADM2s from the control group) to account for spatial interdependence. Alternatively, studies can explicitly model the impact of spatio-network interference (Reich et al., 2021).

In conclusion, this study contributes to the ongoing discussion on the impact of local development projects and their effectiveness by demonstrating how EO-ML methods can innovate in addressing causal challenges in development research. □

CRediT authorship contribution statement

Adel Daoud: Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Cindy Conlin:**

Writing – original draft, Visualization, Formal analysis, Data curation, Conceptualization. **Connor T. Jerzak:** Writing – review & editing, Visualization, Supervision, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Additional background figures

See Figs. A.I.1 and A.I.2.

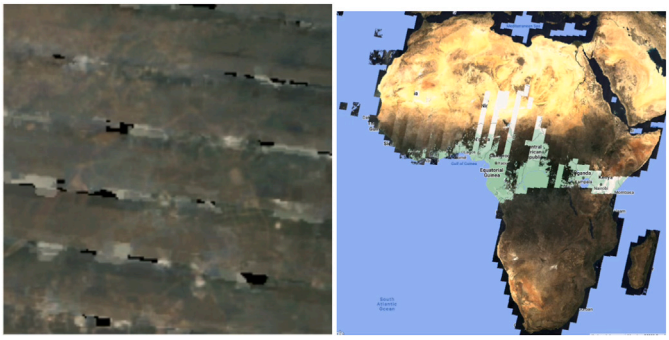
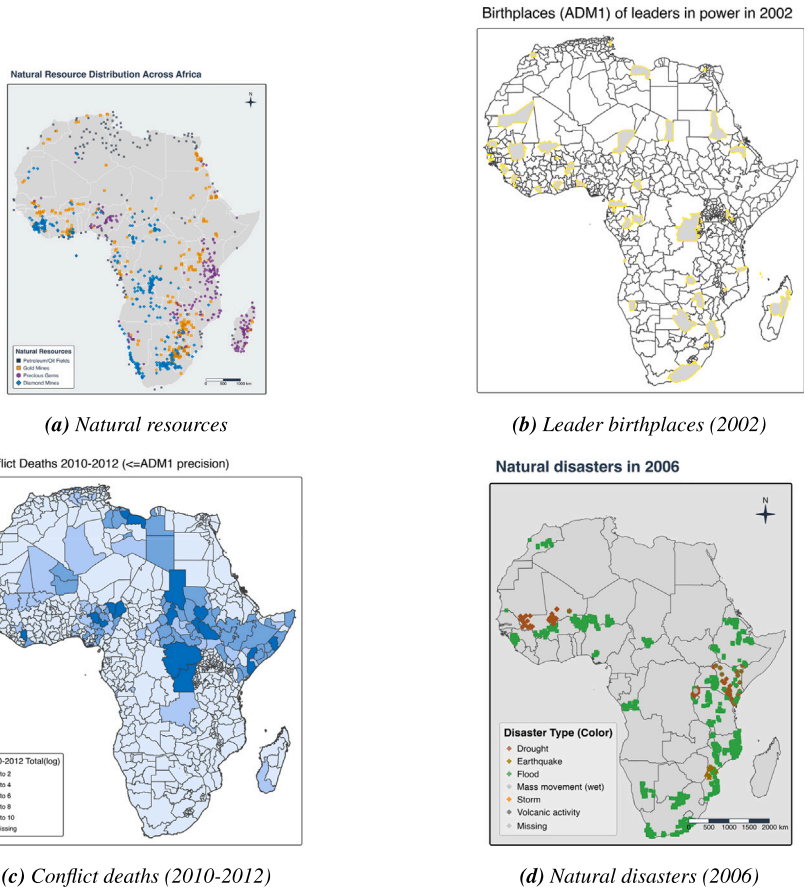


Fig. A.I.1. Missing data in daytime satellite images. LEFT: An image over Cameroon in 2005–2007 with diagonal lines due to the Landsat 7 scan line corrector hardware failure. Location: latitude 4.014, longitude 9.800. RIGHT: Landsat 5 areas of unsaved data. Source: Author queries in Google Earth Engine.



(a) *Natural resources*

(b) *Leader birthplaces (2002)*

(c) *Conflict deaths (2010-2012)*

(d) *Natural disasters (2006)*

Fig. A.I.2. Tabular covariate maps.

Appendix B. Additional empirical results

See Fig. A.I.3 and A.I.4.

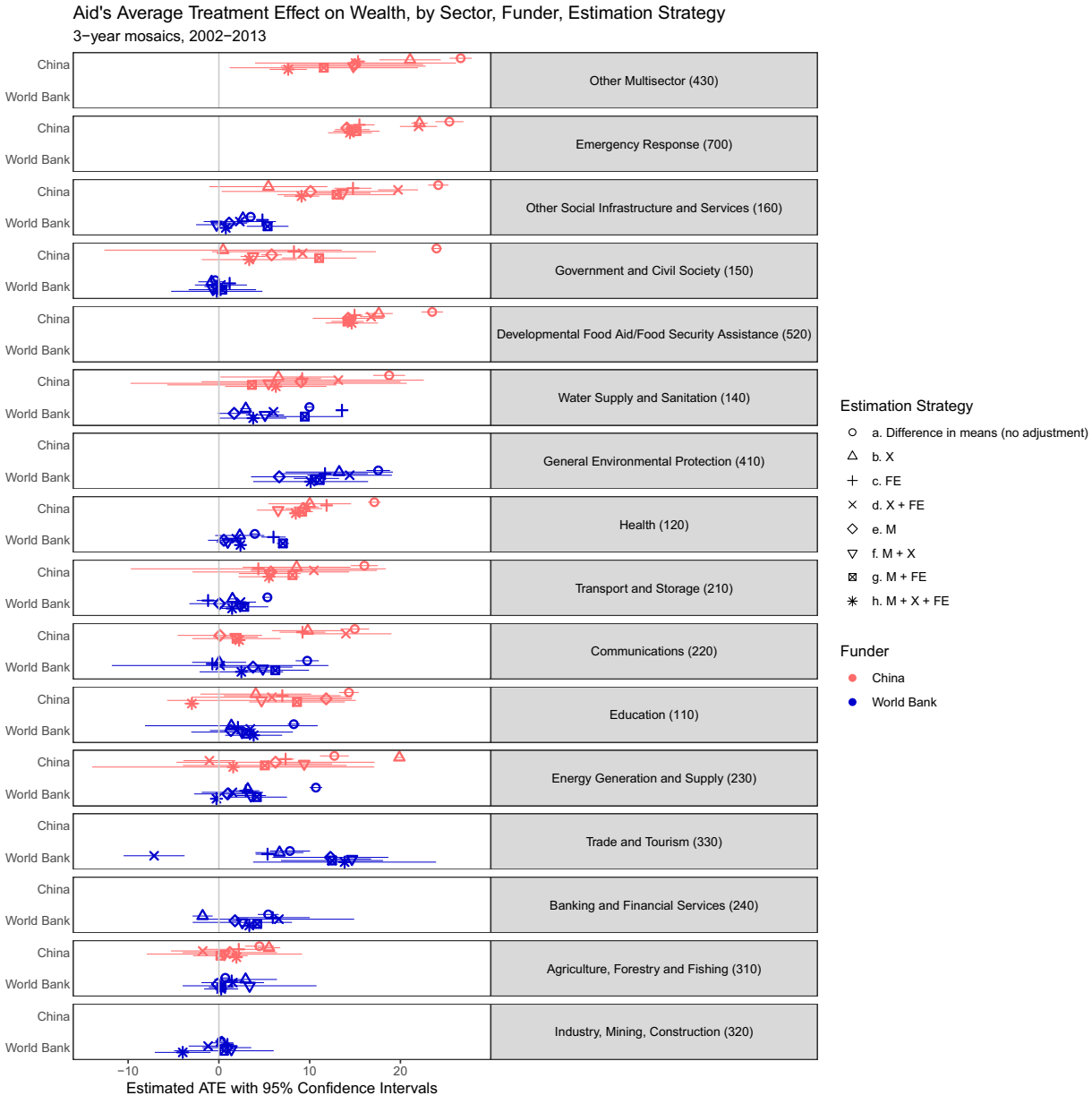


Fig. A.I.3. ATE results, in full.

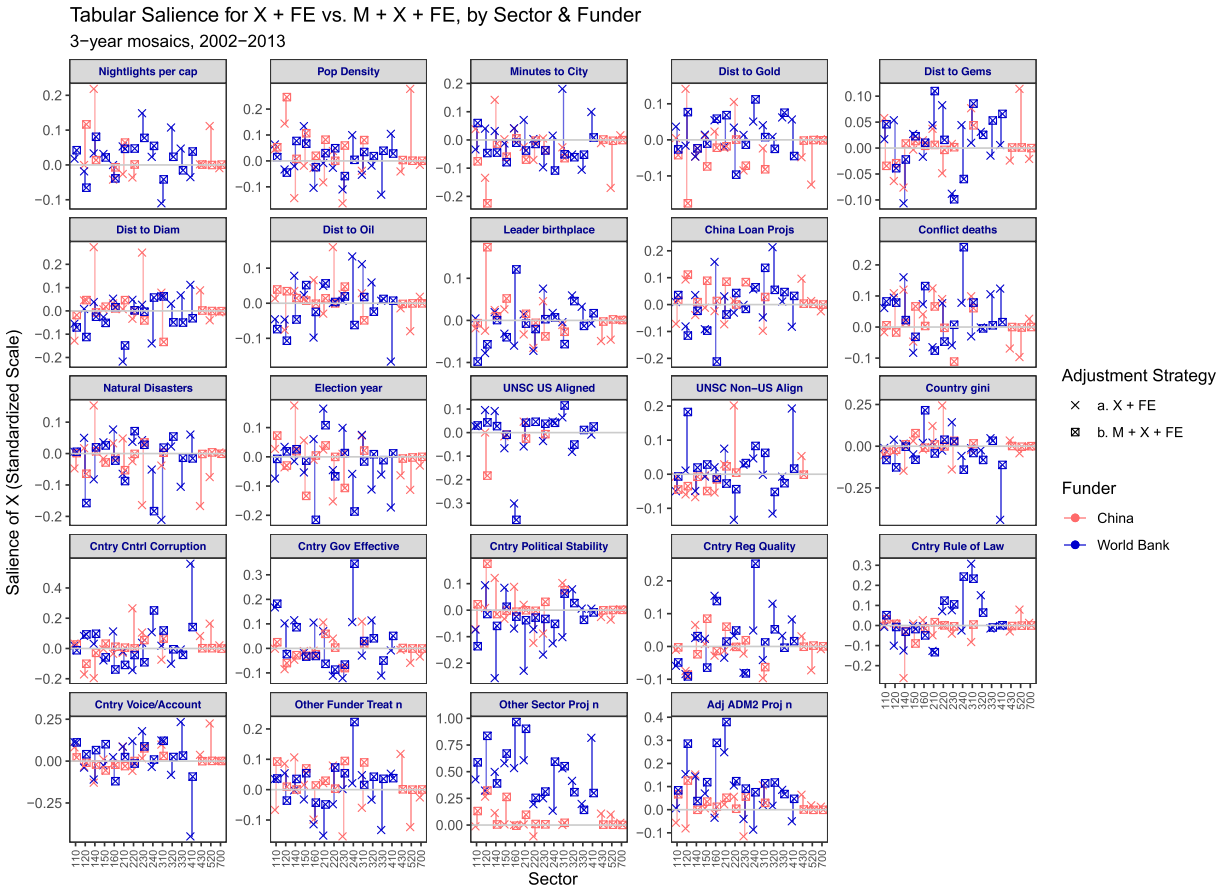


Fig. A.I.4. Salience of tabular variables to aid allocation across funders, sectors, and modeling strategy. This high-level overview indicates whether each variable had a positive, negative, or mixed effect on aid assignment for each sector (y-axis), and how the salience of each variable varied across sectors, funders, and modeling strategy.

Appendix C. Robustness checks

In this section, we present results from three robustness checks designed to assess the sensitivity of our main findings to specific methodological choices regarding covariate adjustment, outcome measurement resolution, and treatment definition.

C.1. Sensitivity to exclusion of lagged nighttime lights

Motivation: A potential concern in our analysis is the inclusion of lagged per-capita nightlights (NTL) in the covariate adjustment set. Because nightlights are highly persistent over time and are also used as features in the training of the machine-learning-derived International Wealth Index (IWI), their inclusion might introduce bias or over-control. To address this, we re-estimated our primary models after strictly excluding the lagged nightlights variable from the tabular covariate set.
Results: Fig. A.I.5 displays the Average Treatment Effects (ATEs) under this specification. The results are substantively similar to the main analysis, preserving the sectoral rankings and the general finding that Chinese projects are associated with larger wealth gains in the short term, at least using this identification strategy. This suggests that our results are not driven by the mechanical reuse of nightlights data.

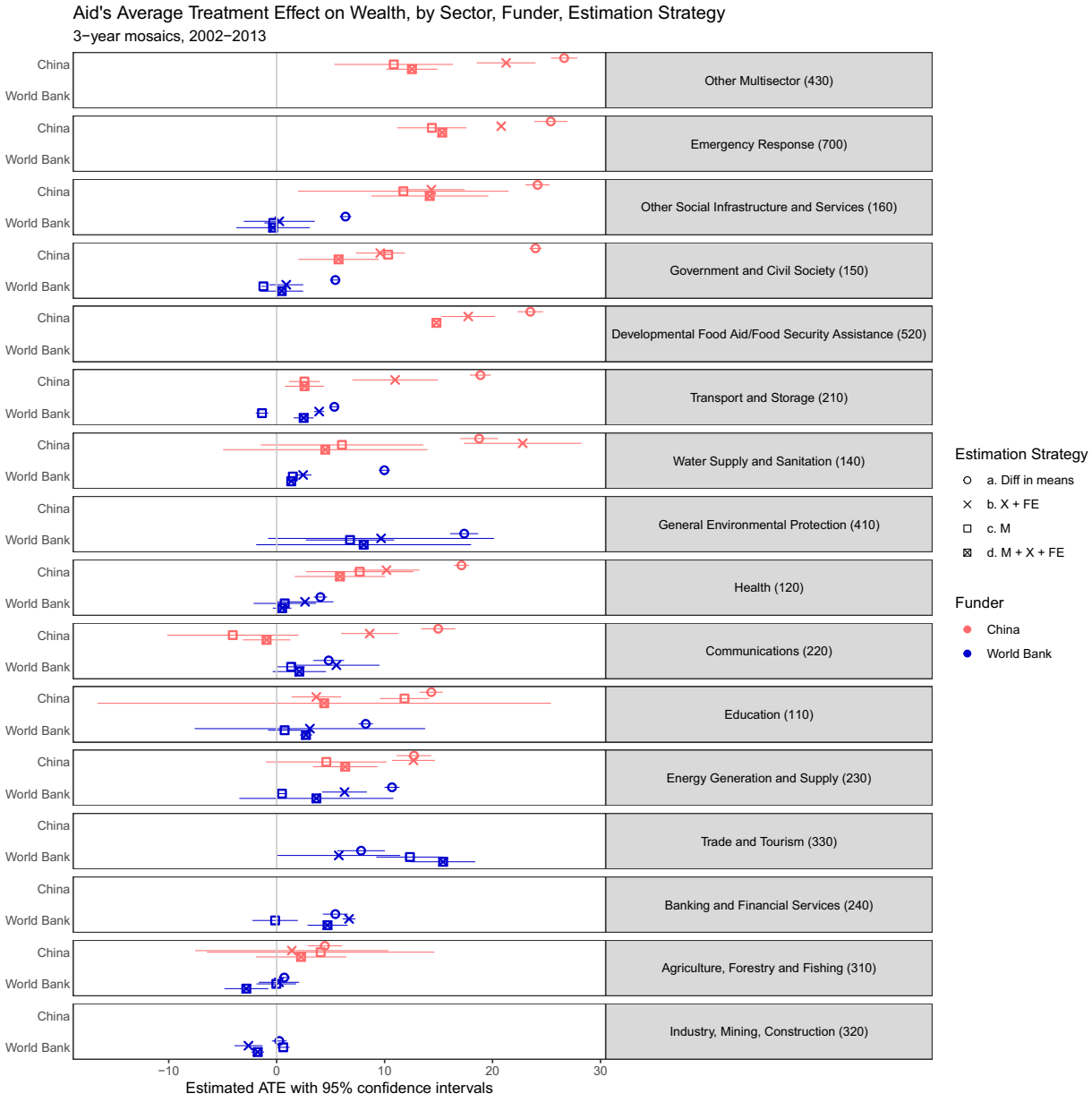


Fig. A.I.5. Robustness (no nightlights): ATEs by sector, funder, and modeling strategy, excluding lagged nightlights from the adjustment set.

C.2. Sensitivity to neighborhood size (Addressing displacement)

Motivation: The DHS cluster coordinates used to define our neighborhoods are randomly displaced by up to 2 km in urban areas and 5 km (or occasionally 10 km) in rural areas to protect respondent privacy. This displacement could misalign the treatment assignment with the outcome

measurement (the 6.7 km × 6.7 km IWI raster). To address this, we re-estimated our models using a coarser outcome grid (≈ 13.4 km × 13.4 km) created by averaging the four 6.7 km cells nearest to the cluster centroid.

Results: Fig. A.I.6 presents the ATE estimates using this coarser outcome grid. While confidence intervals naturally widen slightly due to spatial aggregation, the point estimates and overall patterns remain robust. This indicates that our findings are not artifacts of spatial misalignment caused by DHS coordinate displacement, although we do not address all forms of possible spillover.

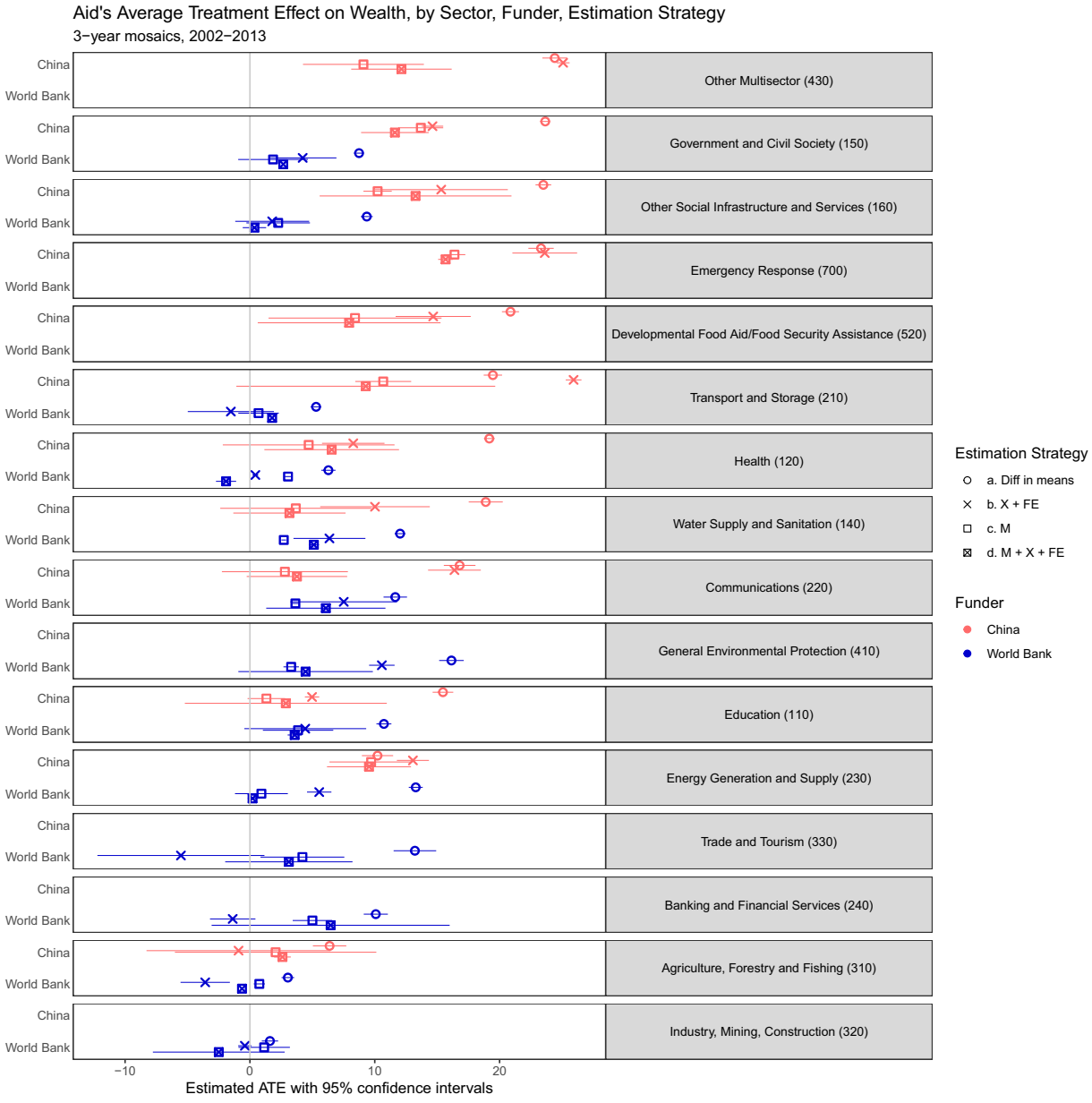


Fig. A.I.6. Robustness (Coarser Outcome Raster): ATEs by sector, funder, and modeling strategy using a ∼13.4 km outcome grid to account for DHS cluster displacement.

C.3. Sensitivity to strict treatment definition

Motivation: In our baseline analysis, we assigned treatment status based on project precision codes, including treating an entire ADM2 region if the project location was only known at the ADM2 level (Precision 3). To address concerns that this might dilute effects or introduce measurement error, we performed a "Strict Precision" robustness check. We restricted the treatment group to include only neighborhoods within 25 km of a precise project location (Precision 1 and 2), excluding those treated solely based on ADM2 boundaries.

Results: Fig. A.I.7 displays the results under this stricter treatment regime. The estimated effects remain broadly consistent with the main analysis, although World Bank estimates are notably higher in magnitude in this regime.

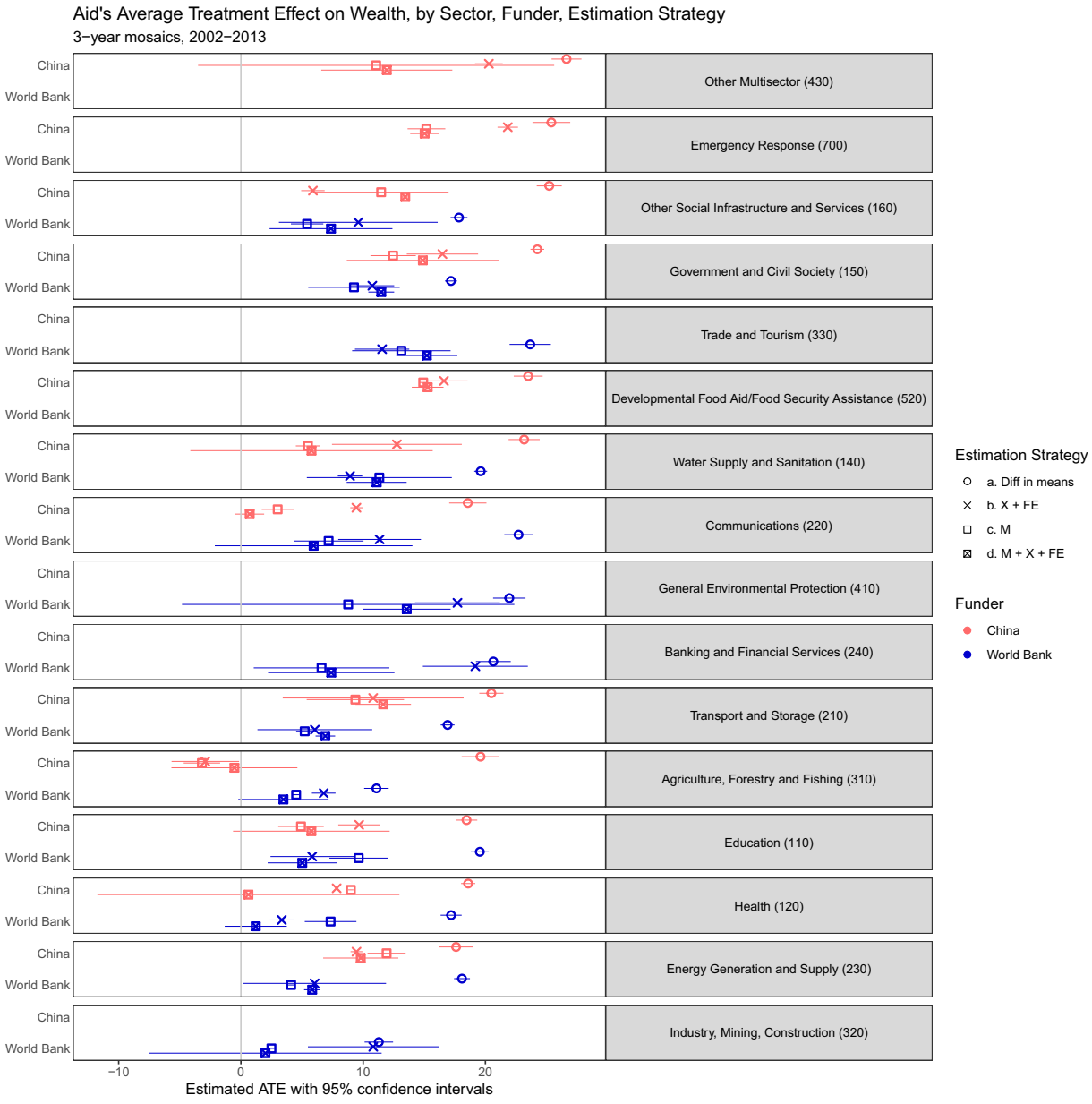


Fig. A.I.7. Robustness (Strict Precision): ATEs by sector, funder, and modeling strategy using a strict treatment definition (excluding ADM2-based assignment).

Appendix D. Supplementary data

Supplementary data for this article can be found online at doi:10.1016/j.worlddev.2026.107328.

Data availability

All code to reproduce the analyses in this paper, along with replication datasets derived from publicly available sources, will be made available at https://github.com/AIandGlobalDevelopmentLab/WorldDev.2026.1073278.

References

Ahmed, S. (2022). China's official finance in the global south: What's the literature telling us? *Review of Economics*, *73*(3), 223–252. https://doi.org/10.1515/roe-2021-0030

AidData. (2017). World Bank geocoded research release, version 1.4.2 geocoded dataset [Technical Report], https://www.aiddata.org/data/world-bank-geocoded-research-release-level-1-v1-4-2.

Hansen, H., Markussen, T., & Andersen, T. B. (2006). US politics and World Bank IDA-lending. *The Journal of Development Studies*, *42*(5), 772–794. https://doi.org/10.1080/00220380600741946

Angelopoulos, A. N., Bates, S., Fannjiang, C., Jordan, M. I., & Zrnic, T. (2023). Prediction-powered inference. *Science*, *382*(6671), 669–674.

Aghdam, R. F. Z., Ahmad, N., Tashpulatova, L., & Asatullaeva, Z. (2021). The impact of foreign aid on economic development: A systematic literature review and content analysis of the top 50 most influential papers. *Journal of International Development*, *33*, 717–751. https://doi.org/10.1002/jid.3543

Banerjee, A., Duflo, E., Glennerster, R., & Kinnan, C. (2015). The miracle of microfinance? Evidence from a randomized evaluation. *American Economic Journal: Applied Economics*, *7*(1), 22–53.

World Bank. (2023). Gini index. https://data.worldbank.org/indicator/SI.POV.GINI.

Bedi, T., Coudouel, A., & Simler, K. (2007). *More than a pretty picture: Using poverty maps to design better policies and interventions*. World Bank Publications.

Bedi, T., Coudouel, A., & Simler, K. (2007). More than a pretty picture: Using poverty maps to design better policies and interventions. Technical report, Washington, DC: World Bank.

DiLorenzo, M., Dolan, C., & BenYishay, A. (2022). The economic efficiency of aid targeting. *World Development*, *160*(106062), https://doi.org/10.1016/j.worlddev.2022.106062

Berlin, M. P., Desai, R. M., & Olofsgård, A. (2023). Trading favors? UN Security Council membership and subnational favoritism in aid recipients. *The Review of International Organizations*, *18*(2), 237–258.

Bitzer, J., & Gören, E. (2018). Foreign aid and subnational development: A grid cell analysis. Technical report, Oldenburg Discussion Papers in Economics. https://www.econstor.eu/handle/10419/175419.

Dreher, A., Fuchs, A., Parks, B., Strange, A., Tierney, M. J., & Bluhm, R. (2020). Connective financing: Chinese infrastructure projects and the diffusion of economic activity in developing countries. *Journal of Urban Economics*, *145*, https://doi.org/10.2139/ssrn.3262101

Bomprezzi, P., Dreher, A., Fuchs, A., Hailer, T., Kammerlander, A., Kaplan, L. C., Marchesi, S., Masi, T., Robert, C., & Unfried, K. (2024). Wedded to prosperity? Informal influence and regional favoritism. CESifo Working Paper.

Brautigam, D. (2009). *The dragon's gift: The real story of China in Africa*. Oxford University Press.

Brazys, S., Vadlamannati, K. C., & Dietrich, S. (2021). Does Chinese development finance raise household welfare in low- and middle-income countries? *World Development*, *140*, 105345. https://doi.org/10.1016/j.worlddev.2020.105345

Briggs, R. C. (2021). Why does aid not target the poorest? *International Studies Quarterly*, *65*(3), 739–752. https://doi.org/10.1093/isq/sqab035

Briggs, R. C. (2017). Does foreign aid target the poorest? *International Organization*, *71*(1), 187–206. https://doi.org/10.1017/S0020818316000345.

Broich, T. (2017). Do authoritarian regimes receive more Chinese development finance than democratic ones? Empirical evidence for Africa. *China Economic Review*, *46*, 180–207. https://doi.org/10.1016/j.chieco.2017.09.006

Burgert, C. R., Zachary, B., & Colston, J. (2013). Incorporating geographic information into demographic and health surveys: A field guide to GPs data collection. *Fairfax: ICF International*, https://dhsprogram.com/publications/publication-dhsm9-dhs-questionnaires-and-manuals.cfm.

Burke, M., Driscoll, A., Lobell, D. B., & Ermon, S. (Mar 2021. ISSN 0036-8075, 1095-9203. Using satellite imagery to understand and promote sustainable development. *Science*, *371*(6535), eabe8628. https://doi.org/10.1126/science.abe8628

Cavallo, E., & Noy, I. (2009). The economics of natural disasters: A survey. Technical report, IDB Working Paper Series.

Chai, Q., & Tang, Z. (2023). The World Bank and China: Comparing the impacts of their development projects in Africa. Available at SSRN 4598476.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J., 2018. *Double/debiased machine learning for treatment and structural parameters*. *The Econometrics Journal*, *21*(1), C1–C68, https://doi.org/10.1111/ectj.12097.

Chin, G. T. (2012). China as a "net donor': Tracking dollars and sense. *Cambridge Review of International Affairs*, *25*(4), 579–603.

Chong, A., Gradstein, M., & Calderon, C. (2009). Can foreign aid reduce income inequality and poverty? *Public Choice*, *140*(1), 59–84.

Cinelli, C., Forney, A., & Pearl, J. (2024). A crash course in good and bad controls. *Sociological Methods & Research*, *53*(3), 1071–1104.

Croft, T. N., Marshall, A. M. J., & Allen, C. K. (2018). Guide to DHS statistics. ICF. https://dhsprogram.com/data/Guide-to-DHS-Statistics/index.cfm.DHS Program.

Croicu, M., & Sundberg, R. (2015). UCDP georeferenced event dataset codebook version 4.0. *Journal of Peace Research*, *50*(4), 523–532.

Daoud, A., & Dubhashi, D. (Jan 2023. ISSN 2644-2353, 2688-8513. Statistical modeling: The three cultures. *Harvard Data Science Review*, *5*(1), https://doi.org/10.1162/99608f92.89f6fe66. https://hdsr.mitpress.mit.edu/pub/uo4hjcx6/release/1.

Daoud, A., & Johansson, F. D. (2025). The impact of austerity on children: Uncovering effect heterogeneity by political, economic, and family factors in low- and middle-income countries. *Social Science Research*, *118*, 102973, ISSN 0049-089X. https://doi.org/10.1016/j.ssresearch.2023.102973. https://www.sciencedirect.com/science/article/pii/S0049089X2300128X.

Daoud, A., Nosrati, E., Reinsberg, B., Kentikelenis, A. E., Stubbs, T. H., & King, L. P. (May 2017. ISSN 0027-8424, 1091-6490. Impact of international monetary fund programs on child health. *Proceedings of the National Academy of Sciences*, *114*(25), 6492–6497. https://doi.org/10.1073/pnas.1617353114. http://www.pnas.org/content/early/2017/05/09/1617353114. 00002.

Daoud, A., Jordan, F., Sharma, M., Johansson, F., Dubhashi, D., Paul, S., & Banerjee, S., Using satellites and artificial intelligence to measure health and material-living standards in India, arXiv preprint arXiv:2202.00109, 2021).

Daoud, A., Herlitz, A., & Subramanian, S. V. (Sep 2022). IMF fairness: Calibrating the policies of the international monetary fund based on distributive justice. *World Development*, *157*, 105924, ISSN 0305-750X. https://doi.org/10.1016/j.worlddev.2022.105924

Wathelet, V., Below, R., Sofia, C. L., Tonnelier, M., van Loenhout, J., Speybroeck, N., & Delforge, D., 2023. EM-DAT: The emergency events database.

Demir, F., & Duan, Y. (2024). Target at the right level: Aid, spillovers, and growth in sub-saharan Africa. *Applied Economics*, *56*(28), 3293–3333.

Dolan, C. B., & McDade, K. K. (2020). Pulling the purse strings: Are there sectoral differences in political preferencing of Chinese aid to Africa? *PLoS one*, *15*(4), e0232126.

Donaldson, D., & Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, *30*(4), 171–198.

Drabo, A. (2021). *How do climate shocks affect the impact of FDI, ODA and remittances on economic growth?* International Monetary Fund.

Dreher, A., & Fuchs, A. (2015). Rogue aid? An empirical analysis of China's aid allocation. *Canadian Journal of Economics*, *48*(3), 988–1023.

Dreher, A., & Lohmann, S. (2015). Aid and growth at the regional level. *Oxford Review of Economic Policy*, *31*(3–4), 420–446.

Dreher, A., Sturm, J.-E., & Vreeland, J. R. (2009). Development aid and international politics: Does membership on the UN Security Council influence World Bank decisions? *Journal of Development Economics*, *88*(1), 1–18.

Dreher, A., Fuchs, A., Parks, B., Strange, A. M., & Tierney, M. J. (Feb 2018). Apples and dragon fruits: The determinants of aid and other forms of state financing from China to Africa. *International Studies Quarterly*, *62*(1), 182–194, ISSN 0020-8833. https://doi.org/10.1093/isq/sqx052.

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

Dreher, A., Fuchs, A., Parks, B., Strange, A., & Tierney, M. J. (2022a). *Banking on Beijing: The aims and impacts of China's overseas development program*. Cambridge University Press. https://doi.org/10.1017/9781108564496

Dreher, A., Lang, V., Rosendorff, B. P., & Vreeland, J. R. (2022). Bilateral or multilateral? International financial flows and the dirty-work hypothesis. *The Journal of Politics*, *84*(4), 1932–1946.

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

Greenstone, M., Guiteras, R., Clasen, T., & Duflo, E. (2015). Toilets can work: Short and medium run health impacts of addressing complementarities and externalities in water and sanitation (working paper 21521). *National Bureau of Economic Research*, https://doi.org/10.3386/w21521

D'Amour, A., Ding, P., Feller, A., Lei, L., & Sekhon, J. (2021). Overlap in observational studies with High-Dimensional covariates. *Journal of Econometrics*, *221*(2), 644–654.

Easterly, W. (2003). Can foreign aid buy growth? *Journal of Economic Perspectives*, *17*(3), 23–48.

Easterly, W. (2006). *The White man's burden: Why the west's efforts to aid the rest have done so much ill and so little good*. Penguin Press.

Francazi, E., Baity-Jesi, M., & Lucchi, A. (2023). A theoretical analysis of the learning dynamics under class imbalance. In *International conference on machine learning* (pp. 10285–10322). PMLR.

Freedman, D. A., & Berk, R. A. (2008). Weighting regressions by propensity scores. *Evaluation Review*, *32*(4), 392–409.

Galiani, S., Gertler, P., & Schargrodsky, E. (2005). The causal effect of water quality on diarrheal diseases in children. *Econometrica*, *73*(5), 1505–1548. https://doi.org/10.1111/j.1468-0262.2005.00631.x

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

Gleditsch, N. P., Lujala, P., Rød, J. K., & Gilmore, E., 2005. Conflict diamonds: A new dataset.

Greßer, C., & Stadelmann, D. (2021). Evaluating water-and health-related development projects: A cross-project and micro-based approach. *The Journal of Development Studies*, *57*(7), 1221–1239.

Guillon, M., & Mathonnat, J. (2020). What can we learn on Chinese aid allocation motivations from available data? A sectorial analysis of Chinese aid to African countries. *China Economic Review*, *60*, 101265.

Halleröd, B., Rothstein, B., Daoud, A., & Nandy, S. (2013). Bad governance and poor children: A comparative analysis of government efficiency and severe child deprivation in 68 low-and middle-income countries. *World Development*, *48*, 19–31. http://www.sciencedirect.com/science/article/pii/S0305750X13000831.

Huang, G., Sun, Y., Liu, Z., Sedra, D., & Weinberger, K. Q. (2016). Deep networks with stochastic depth. In *European conference on computer vision* (pp. 646–661). Springer.

Imbens, G. W., & Rubin, D. B. (2015). *causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press. https://doi.org/10.1017/CBO9781139025751

Isaksson, A.-S., & Kotsadam, A. (2018). Chinese aid and local corruption. *Journal of Public Economics*, *159*, 146–159.

Burke, M., Xie, M., Davis, W. M., Lobell, D. B., Ermon, S., & Jean, N. (2016). Combining satellite imagery and machine learning to predict poverty. science. *Science*, *353(6301)*, 790–794. https://doi.org/10.1126/science.aaf7894

Daoud, A., & Jerzak, C. T., CausalImages: An R Package for causal inference with earth observation, bio-medical, and social science images, arXiv preprint arXiv:2310.00233, 2023a. https://doi.org/10.48550/arXiv.2310.00233

Jerzak, C. T., Johansson, F. D., & Daoud, A. (2023). Image-based treatment effect heterogeneity. In *Conference on causal learning and reasoning* (pp. 531–552). PMLR.

Johansson, F., Daoud, A., & Jerzak, C. T. (2023b). Integrating earth observation data into causal inference: Challenges and opportunities. https://arxiv.org/abs/2301.12985.

Kakooei, M., & Daoud, A. (2024). Increasing the confidence of predictive uncertainty: Earth observations and deep learning for poverty estimation. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–13.

Kaufmann, D., & Kraay, A., 2024. The worldwide governance indicators.

Kim, Y., & Kim, Y. (2021). The autonomy of international organizations? The analysis of major powers' influence over the World Bank's aid policies. *International Area Studies Review*, *24*(3), 224–240. https://doi.org/10.1177/22338659211024879

Kino, S., Hsu, Y.-T., Shiba, K., Chien, Y.-S., Mita, C., Kawachi, I., & Daoud, A. (Sep 2021). A scoping review on the use of machine learning in research on social determinants of health: Trends and research prospects. *SSM - Population Health*, *15*, 100836, ISSN 2352-8273. https://doi.org/10.1016/j.ssmph.2021.100836. https://www.sciencedirect.com/science/article/pii/S2352827321001117.

Krueger, A. B., & Lindahl, M. (2001). Education for growth: Why and for whom? *Journal of Economic Literature*, *39*(4), 1101–1136. https://doi.org/10.1257/jel.39.4.1101

Lee, J. Y., Park, J., & Kim, Y. (2021). Chinese development finance and Local-Level inequality in Africa. *The Chinese Journal of International Politics*, *14*(4), 533–564. https://doi.org/10.1093/cjip/poab015

Leiderer, S. (2021). The effects of Chinese aid on household welfare in Africa. *World Development*, *140*, 105221. 10.1016/j.worlddev.2020.105221.

Rød, J. K., Thieme, N., & Lujala, P. (2007). Fighting over oil: Introducing a new dataset. *Conflict Management and Peace Science*, *24*(3), 239–256.

Malik, A., Parks, B., Russell, B., Lin, J., Walsh, K., Solomon, K., Zhang, S., Elston, T., & Goodman, S. (2021). *Banking on the Belt and Road: Insights from a new global dataset of 13, 427 Chinese development projects* (pp. 23–36). Williamsburg, VA: AidData at William & Mary.

Mandon, P., & Woldemichael, M. T. (2023). Has Chinese aid benefited recipient countries? Evidence from a meta-regression analysis. *World Development*, *166*, 106211.

Martel, R. (2021). Can foreign aid reduce inequality? The importance of the domestic context. *The Journal of International Trade & Economic Development*, *30*(7), 1028–1053. https://doi.org/10.1080/09638199.2021.1947265

Martina, M. (2020). Does Chinese aid improve health outcomes? Evidence from a new dataset. *Journal of Development Studies*, *56*(5), 943–960. https://doi.org/10.1080/00220388.2019.1627993

Martorano, B., Metzger, L., & Sanfilippo, M. (2020). Chinese development assistance and household welfare in Sub-Saharan Africa. *World Development*, *129*, 104909, ISSN 0305-750X. https://doi.org/10.1016/j.worlddev.2020.104909. https://www.sciencedirect.com/science/article/pii/S0305750X20300358.

Martuscelli, A. (2020). The economics of China's engagement with Africa: What is the empirical evidence? *Development Policy Review*, *38*(3), 285–302.

McGillivray, M., Feeny, S., Hermes, N., & Lensink, R. (2006). Controversies over the impact of development aid: It works; it doesn't; it can, but that depends. *Journal of International Development: The Journal of the Development Studies Association*, *18*(7), 1031–1050.

Meng, X.-L. (2014). A trio of inference problems that could win you a Nobel Prize in statistics (if you help fund IT). *Past, Present, and Future of Statistical Science*, 537–562. https://doi.org/10.1201/b16720-50

Murugaboopathy, S., Jerzak, C. T., & Daoud, A. (Aug, 2025. Platonic representations for poverty mapping: Unified Vision-Language codes or Agent-Induced novelty?

NASA. (30 Nov 2021). Landsat 5. https://landsat.gsfc.nasa.gov/satellites/landsat-5/.

OECD. (2005). Paris declaration on aid effectiveness. https://doi.org/10.1787/9789264098084-en

Öhler, H., & Nunnenkamp, P. (2014). Needs-based targeting or favoritism? The regional allocation of multilateral aid within recipient countries. *Kyklos*, *67*(3), 420–446.

Öhler, H., Negre, M., Smets, L., Massari, R., & Bogetić, Ž. (2019). Putting your money where your mouth is: Geographic targeting of World Bank projects to the bottom 40 percent. *PLoS One*, *14*(6), e0218671.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Pettersson, M., Jerzak, C. T., & Daoud, A. (Aug, 2025. Debiasing machine learning predictions for causal inference without additional ground truth data: "one map, many trials" in satellite-driven poverty analysis.

Pettersson, M. B., Kakooei, M., Ortheden, J., Johansson, F. D., & Daoud, A. (2023). Time series of satellite imagery improve deep learning estimates of Neighborhood-Level poverty in Africa. In *IJCAI* (pp. 6165–6173).

Platteau, J.-P. (2004). Monitoring elite capture in community-driven development. *Development and Change*, *35*(2), 223–246.

Prince, S. J. D. (2023). *Understanding deep learning*. The MIT Press. http://udlbook.com.

Reich, B. J., Yang, S., Guan, Y., Giffin, A. B., Miller, M. J., & Rappold, A. (2021). A review of spatial causal inference methods for environmental and epidemiological applications. *International Statistical Review*, *89*(3), 605–634.

AidData Research and Evaluation Unit. (2017). Geocoding methodology, version 2.0. AidData at William & Mary. https://www.aiddata.org/publications/geocoding-methodology-version-2-0.

Rister Portinari Maranca, A., Chung, J., Hinck, M., Wolsky, A. D., Egami, N., & Stewart, B. M. (2025). Correcting the measurement errors of AI-assisted labeling in image analysis using Design-Based supervised learning. *Sociological Methods & Research*, *54*(3), 984–1016, ISSN 0049-1241. https://doi.org/10.1177/00491241251333372

Rosvold, E. L., & Buhaug, H. (2021). GDIS, a global dataset of geocoded disaster locations. *Scientific Data*, *8*(1), 61.

Rubin, D. B. (1990). Formal model of statistical inference for causal effects. *Journal of Statistical Planning and Inference*, *25*, 279–292. https://doi.org/10.1016/0378-3758(90)90077-8

Rubin, D. B. (1991). Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*, 1213–1234.

Rutstein, S. O. (2015). Steps to constructing the new DHS wealth index. *Rockville, MD: ICF International*, *6*.

Sakamoto, K., Jerzak, C. T., & Daoud, A. (Apr, 2025. A scoping review of earth observation and machine learning for causal inference: Implications for the geography of poverty.

Sardoschau, S., & Jarotschkin, A. (2024). Chinese aid in Africa: Attitudes and conflict. *European Journal of Political Economy*, *81*, 102500.

Landsat 7 | Landsat Science. 30 Nov, 2021. https://landsat.gsfc.

Scott, A. J., 2009. World development report 2009: Reshaping economic geography.

Skarda, I. (Sep, 2022. *Is income inequality important for foreign aid effectiveness?* https://doi.org/10.21203/rs.3.rs-2073414/v1 [Preprint].

Smits, J., & Steendijk, R. (2015). The international wealth index (IWI). *Social Indicators Research*, *122*(1), 65–85.

Ssozi, J., Asongu, S., & Amavilah, V. H. (2019). The effectiveness of development aid for agriculture in Sub-Saharan Africa. *Journal of Economic Studies*, *46*(2), 284–305.

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

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł, & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Warmerdam, W., & van Dijk, M. P. (2013). Chinese state-owned enterprise investments in Uganda: Findings from a recent survey of Chinese firms in Kampala. *Journal of Chinese Political Science*, *18*(3), 281–301.

Weiss, D. J., Nelson, A., Gibson, H. S., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., et al. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, *553*(7688), 333–336.

Winters, M. S. (2010). Choosing to target: What types of countries get different types of World Bank projects. *World Politics*, *62*(3), 422–458.

Woods, N. (2008). Whose aid? Whose influence? China, emerging donors and the silent revolution in development assistance. *International Affairs*, *84*(6), 1205–1221.

WorldPop. (2023). Global high resolution population denominators project: Population density. https://hub.worldpop.org/geodata/listing?id=77. Funded by The Bill and Melinda Gates Foundation (OPP1134076). Retrieved July 24, 2023.

Xu, Z., Zhang, Y., & Sun, Y. (2020). Will foreign aid foster economic development? Grid panel data evidence from China's aid to Africa. *Emerging Markets Finance and Trade*, *56*(14), 3383–3404.

Zhu, F. W., Jerzak, C. T., & Daoud, A. (07–09 May 2025). Optimizing multi-scale representations to detect effect heterogeneity using earth observation and computer vision: Applications to two anti-poverty RCTs. In B. Huang, & M. Drton (Eds.), *Proceedings of the fourth conference on causal learning and reasoning* (Vol. 275, pp. 894–919). PMLR Proceedings of Machine Learning Research. https://proceedings.mlr.press/v275/zhu25a.html.