AIDDATA
A Research Lab at William & Mary

WORKING PAPER 85

July 2019

Why does aid not target the poorest?

Ryan C. Briggs
University of Guelph
Guelph Institute of Development Studies and Department of Political Science

Abstract
Foreign aid projects typically have local effects, so if they are to reduce poverty then they need to be placed close to the poor. I show that, conditional on local population, World Bank (WB) project aid targets richer parts of countries. This relationship holds over time and across world regions. I test five explanations for pro-rich targeting using a pre-registered conjoint experiment on WB task team leaders (TTLs). TTLs perceive aid-receiving governments as most interested in targeting aid politically and controlling implementation. They also believe that aid works better in poorer or more remote areas, but that implementation in these areas is uniquely difficult. These results speak to debates in distributive politics, international bargaining over aid, and principal-agent issues in international organizations. The results also suggest that tweaks to WB incentive structures to make ease of project implementation less important may encourage aid to flow to poorer parts of countries.

Author Information

Ryan C. Briggs
University of Guelph
Guelph Institute of Development Studies and Department of Political Science
rbriggs@uoguelph.ca

Acknowledgements
The author would like to thank Ariel BenYishay and the broader team at AidData for guidance on using their datasets and Matt Winters and Zeljko Bogetic for advice on the TTL survey. I'd also like to thank participants at CASID 2019, Dan Honig, Hannes Ohler, ˙Vincent Perrot, Nicolas Van de Sijpe, Matt Winters, and a number of TTLs who asked to not be named for comments. This project received ethics approval from the University of Guelph Research Ethics Board (#19-03-018).

Foreign aid is a large transfer of resources from high to low-income countries.[1] Aid is especially important from the point of view of governments in low-income countries, where it often exceeds 20% of their total expenditure (World Bank 2018).[2] If we consider politics to be about 'who gets what?', then understanding who benefits from aid is an important component of understanding politics in aid-receiving countries.

The donor community has repeatedly pledged to use aid to reduce poverty[3] and aid very often has local effects (Briggs 2018b). Thus, if one takes seriously the public positions of donors then one would expect poorer people within countries to be the direct beneficiaries of aid. The case for targeting aid to areas of poverty is even stronger presently, as Sustainable Development Goal 10 commits the donor community to reducing income inequalities within countries.

However, research on subnational aid targeting has generally shown that, conditional on population, aid flows to richer parts of countries (Öhler and Nunnenkamp 2014; Öhler et al. 2019; Custer et al. 2017; Briggs 2017, 2018a,b). While this work has produced a consistent result, much of it has focused exclusively on Africa (Briggs 2017, 2018a,b), most of it has only examined subnational aid targeting across large spatial regions (Öhler and Nunnenkamp 2014; Öhler et al. 2019; Custer et al. 2017; Briggs 2017), all of it has ignored

---

[1] Aid tends to flow to poorer countries (Alesina and Dollar 2000; Dollar and Levin 2006; Neumayer 2003; Bandyopadhyay and Wall 2006).

[2] For example, in Uganda in 2016 aid was 53% of total government expenditure. In Nepal, it was 31% (World Bank 2018).

[3] For example, in *Assessing Aid* the WB wrote that "the main aim of aid is to reduce poverty" (World Bank 1998, p. 38).

possible changes in the relationship between aid and poverty over time, and none of it has examined a comprehensive group of aid recipients.

In the first analysis in this paper, I examine the subnational distribution of WB project aid and address all of these issues. In line with past work, I show that aid targets richer parts of countries. This finding holds over time and across all world regions, though sub-Saharan Africa shows an especially strong pro-rich bias in subnational aid allocation. I also show that this result is not simply due to aid concentrating on capital cities, though capital cities do get more aid than their economic activity or population would suggest. This result confirms a puzzle in the literature: donors that say that they are pro-poor allocate within-country aid in ways that do not seem to be pro-poor.

The second analysis in this paper uses a conjoint experiment run on WB TTLs to test five explanations for why aid flows to richer parts of countries.[4] I find no support for the idea that client governments want to target aid to cities, or other relatively well-off places, because unrest there is uniquely threatening. I also find no support for the idea that aid works better in richer parts of countries. The most supported explanation is that poorer and more remote areas are uniquely difficult places in which to implement projects. If TTLs feel pressure to produce many projects and need to access project locations in order to get them approved and implemented, or if TTLs face bureaucratic hurdles to accessing remote areas, then this issue may explain why aid tends not to flow to poorer parts of countries. The negative causal

[4] Before collecting data, I pre-registered the complete pipeline of code used in the analysis (EGAP 20190424AB).

effect of remoteness on ease of implementation is stronger in sub-Saharan Africa than in other regions, a finding that matches the descriptive result that aid is most pro-rich in Africa. TTLs working in Africa, but not elsewhere, also believe that projects in remote areas will receive worse outcome ratings than other projects. Thus, the best explanation for pro-rich subnational aid targeting is an interaction between difficult geography and bureaucratic incentives that encourage TTLs to select projects that they think are low risk and easy to implement, as this allows them to focus their time elsewhere.

Aside from directly-related research on subnational poverty targeting (Öhler and Nunnenkamp 2014; Öhler et al. 2019; Custer et al. 2017; Briggs 2017, 2018a,b), the present research is related to work testing who controls foreign aid. I find support for the idea that recipient governments would like to target aid to core voters (Briggs 2014; Jablonski 2014; Dreher et al. 2019), but not swing or opposition voters (Masaki 2018). I also produce results of interest to those concerned with bypass aid (Dietrich 2013, 2016; Shin et al. 2017), aid capture (Winters 2014), the importance of recipient ownership in the successful use of aid (Deutscher and Fyson 2008), and efficiency versus equity in resource allocations (Bardhan 1996). This paper's focus on internal incentive structures within the WB places it within the "the bureaucratic turn" in research on foreign aid (Gulrajani 2017, p. 375). Methodologically, this paper is part of a growing body of research using survey experiments to evaluate the politics of foreign aid (Findley et al. 2017; Dietrich and Winters 2015; Dietrich et al. 2018).

1 Why might aid not target the poorest?

Aid in aggregate has been shown to target wealthier areas of countries (Öhler and Nunnenkamp 2014; Öhler et al. 2019; Custer et al. 2017; Briggs 2017, 2018a,b). Further, while it is possible to find subsets of aid that do not target richer areas, these subsets are generally also not pro-poor. For example Öhler et al. (2019), show that WB aid for education and health is neither pro-poor nor pro-rich.[5] As noted in Briggs (2017), this lack of poverty targeting presents a puzzle because the donor community has repeatedly pledged to use aid to reduce poverty and project aid very often has local effects. This section presents five explanations for why aid from poverty-sensitive donors may fail to reach the poorest. I later test these explanations using a conjoint experiment run on WB TTLs.

The explanations in this section are motivated by the fact that TTLs face career pressure to get many projects approved, and less importantly, to make sure that their projects are rated well internally and by the WB's Independent Evaluation Group (IEG). The importance of these two factors is shown graphically in Figure 1, where surveyed TTLs were asked about importance of getting new projects approved and getting good IEG outcome ratings to their career.[6] The high importance of getting projects approved, and the lesser but still present pressure to get good ratings, is known within the WB. For example, an IEG report from 2016 noted that staff face "pressure

---

[5]Though Marty et al. (2017) and Custer et al. (2017) both find that health aid flows more to richer areas.

[6]The bars are weighted using the survey weights. Details about the survey and sample are presented in Section 3.1.

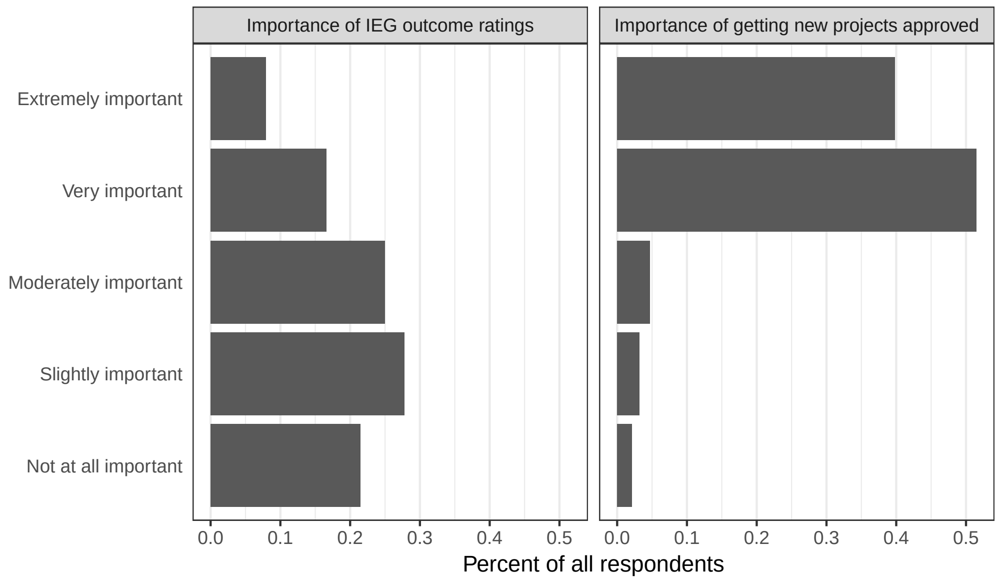

Figure 1: Importance of factors to TTL career success.

for lending volume," have a "perception that individual success depends more on obtaining new deals and ensuring timely disbursement than on quality implementation," and have an "acute focus on outcome ratings" (WB IEG 2016, p. 28). The same report noted that in the WB "prestige was perceived as coming from peer recognition of successes, particularly through getting new projects approved. Fear of damage to one's reputation and concerns about reputational risks attached to poor results was a recurrent theme" (WB IEG 2016, p. 28).

TTLs face career pressure to get projects approved and to have them rated well. In response, TTLs may learn to direct their attention away from hard-to-approve projects or projects that are likely to get bad ratings. The following paragraphs describe how these pressures could filter out hypothetical projects as they move through the project cycle, from identification to evaluation.

The first factor that matters early on in the project cycle is interest on the part of the client government. Hypothetical projects that are not of interest to the client government are very unlikely to ever become a reality, so client preferences strongly shape the set of realized projects. If clients want to spatially target aid in ways that correlate positively with income, then this initial filter could direct aid away from poorer parts of countries. For example, client governments may prefer targeting aid to cities, perhaps because political unrest in cities is both easier to organize and more threatening to the government (Bates 1981). If this is the case, then aid will flow to generally wealthier urban areas instead of poorer rural and remote areas.[7]. Importantly, this would be due to donor deference to recipient preferences and such ownership is thought to be a key ingredient in the successful use of aid (Deutscher and Fyson 2008).

Second, projects need to be approved through the WB bureaucracy. This approval process is not trivial, and staff that are interested in getting many projects approved may learn to focus on projects that are easier to get approved. It is possible that projects in poorer areas are harder to get approved, perhaps due to expected difficulties in implementation in remote areas. It could also be more difficult to get projects in poorer areas approved due to simple cost-benefit analysis, as "development programs that dig deeply

[7]People in rural areas generally have lower standards of living than people in urban areas (Young 2013; World Bank 2009) This extends beyond income. Rural children, for example, typically have higher rates of stunting, lower weight, and higher infant mortality than urban children (Smith et al. 2005; Van de Poel et al. 2007; Paciorek et al. 2013).

into the lowest-income communities [...] tend to have fewer beneficiaries and greater delivery costs" (Ajmera and Fields 2016, p. 152). For example, a clinic built in a remote area may both cost more to build than an urban clinic and may also reach fewer people.[8] Thus, TTLs trying to get projects approved quickly might learn that approval is easier when projects are placed in wealthier places.

Third, even if clients are interested in placing projects in poorer parts of countries and even if the approval process for pro-poor projects is easy, actually implementing the project in a poorer place may be difficult. This follows a similar logic to the past argument about cost-benefit analysis, in that rural areas often have lower-quality infrastructure than cities (Johns and Torres 2005). Aside from the pure difficulty in directly implementing in rural areas, monitoring can also be difficult. For example, the WB has standards around labor conditions, community relations, or environmental impacts and ensuring compliance in projects with locations spread across remote or rural areas can be both time consuming and physically challenging. Similarly, TTLs concerned with the difficulty of implementing projects may avoid poorer and remote parts of countries not because of poverty, but because accessing and working in these places is onerous. This kind of convenience factor may be one reasons why Kenyan NGOs locate themselves in places with higher

[8] A similar argument is also made in Custer et al. (2017, p. 18), who write "geographically disadvantaged regions tend to be sparsely populated and infrastructure-poor, such that development organizations might also consider it to be more efficient to focus their efforts where they can reach a greater number of poor people at a lower cost, even if those regions are relatively better off." This is also noted in Öhler et al. (2019, p. 16), who write "cost-effectiveness considerations may lead to an allocation of resources that neglects the poor in remote, difficult-to-access areas."

density of paved roads or a closer distance to Nairobi (Brass 2012).

Fourth, even if TTLs do not inherently care about the difficulty of implementation they may still avoid remote areas if they fear that projects in these places will get worse ratings. Project outcome ratings assess "the extent to which the project's major relevant objectives were achieved, or are expected to be achieved, efficiently" (WB IEG 2019). The three dimensions of the outcome score are relevance of the project's objectives, the extent to which the objectives were achieved (efficacy), and that the cost of the project was reasonable given the benefits (efficiency). Since 1995, these ratings have been on a six-point scale. Outcome ratings will be lower in poorer areas if projects in these areas are lower on any of the three dimensions of relevance, efficacy, of efficiency. If we consider efficacy, implementing in remote areas might mean leading projects that more often fail to achieve their goals due to unreliable infrastructure, a less certain security situation, or other factors that more often exist in remote areas. The IEG notes that sometimes "unsuccessful outcomes are caused by major shocks outside the control of the WB such as, for example, disasters, conflict, or economic crises" (WB IEG 2016, p. xiv).[9] Staff understandably do not appreciate projects receiving bad ratings due to factors beyond their control, and the IEG writes in a recent report that "measuring and rating project outcomes at closing against objectives stated at design years earlier has become a source of tension and perceived rigidity" (WB IEG 2016, p. xiv). Nevertheless, if projects in poorer areas

[9]In the same report they recommended reforming the review system to make it more able to accommodate course corrections and unforeseeable events.

have more downside risk then this may encourage ratings-conscious TTLs to concentrate their efforts on projects in wealthier areas where outcomes are more consistent.

Fifth and finally, TTLs and the WB in general care about promoting development. Perhaps TTLs think that projects in richer (but still poor) parts of countries are better for development than projects in the poorest places. This could again be due to the cost-benefit calculations noted above. One can plausibly help more people per dollar if aid is spent in richer places within poor countries. It is also plausible that there are strong complementarities between aid and social or physical infrastructure, and these could imply that aid projects will work much better in richer places. One could also believe that urban aid is better if one thinks that the primary goal of aid is not to ameliorate the effects of poverty but to do things that make industrialization and fast economic growth more likely. This could involve boosting education or improving the energy sector or building roads, but regardless of the sector if one wants to use aid to industrialize and boost growth then it would make sense to cluster these interventions in or near cities.[10] In this way, the debate over efficiency versus equity in subnational aid allocation mirrors broader debates in public policy (Bardhan 1996).

Each of these five explanations maps to one dependent variable in a con-

[10]The authors of the 2009 World Development Report write that their "message" is that "economic growth is seldom balanced [across rural and urban areas]. Efforts to spread it prematurely will jeopardize progress. Two centuries of economic development show that spatial disparities in income and production are inevitable" (World Bank 2009, p. 5-6). This argument will be especially persuasive if one believes that 'growth is good for the poor' (Dollar and Kraay 2002; Dollar et al. 2016) and that aid can increase the growth rate if targeted to richer parts of countries.

joint survey experiment that I ran on WB TTLs in May 2019. I randomized features of pairs of projects, such as where they are located within a recipient country, and then ask respondents to pick which project best fulfills some criteria, such as which project the client government would be more interested in. This approach allows one to see which kinds of projects *do not* get implemented. Observational data from completed project lists cannot show us the projects that did not get implemented, so testing the causal effect of some factor on project initiation with observational data is incredibly difficult.[11] This difficulty is why I use a survey experiment.

The results reveal that the most compelling explanation for pro-rich aid targeting is based on the reported difficulty of implementing projects in rural, remote, and poorer parts of countries. While these poorer places are thought to be harder places to implement aid, TTLs also think that aid is better for development when it reaches these places. The analysis of the survey data also presents a variety of results relevant to political science and development studies. For example, I show that TTLs think that client governments dislike bypass aid (Dietrich 2013) and want to target aid to core but not swing voters. Supporting the importance of local ownership of aid, I also show that implementation is thought to be harder in areas that support the opposition party and easier in places that support the party in power.

[11] In particular, it is very easy to end up with correlations that illustrate Berkson's paradox. For example, in the set of realized aid projects one may find that projects located in remote areas are not more expensive to implement than projects in urban areas. However, donor staff are likely selecting projects in part on implementation costs and so we will never see the possibly large number of remote projects that were too expensive to implement and so were passed over.

The next section presents an analysis of WB aid targeting within countries using observational data. I affirm past work showing that aid targets richer parts of countries. I then present the conjoint experiment and discuss some wider implications of the research before concluding.

2 Descriptive analysis of poverty targeting

2.1 Research design

I examine aid from the WB to all recipients over a 10 year period, and relative to past work I retain more granularity in both the spatial and temporal dimensions of my data. In the spatial dimension, I aggregate measures of aid, poverty, and population into a 0.5° latitude by 0.5° longitude grid (Tollefsen et al. 2012). This combination of better-than-regional subnational precision, universal coverage of recipient countries, and a decade of data is new in the literature.

Information on aid comes from AidData (Strandow et al. 2011) and I only look at aid projects as only they have the requisite geographic information. While not all aid is project aid, project aid made up 85% of all WB aid and 53% of all aid from OECD DAC countries in 2017, the most recent year with data (OECD DAC 2019). I use AidData's WB Geocoded Research Release, Version 1.4.2, which runs from 1995 to 2014. Following prior work (e.g. Briggs 2018b), I subset the data so that I include only projects with a precision code less than 3 (so all locations are geocoded to within 25 km of the correct location). This makes it very likely that each project is placed within the correct cell-year. I limit the data to precisely coded projects,

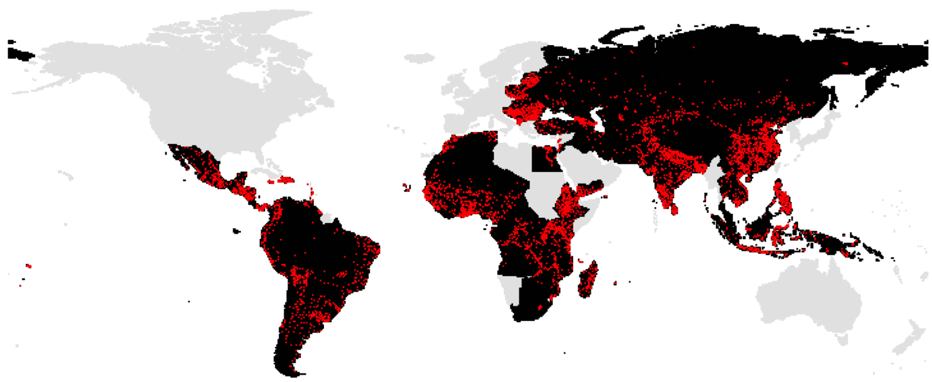

Figure 2: Red cells received at least one precisely targeted WB project between 1995 and 2005 (inclusive). Black cells received no WB project aid but were part of countries that received aid.

rather than projects that were geocoded to (centroids of) regions, because a light measure taken at a regional centroid is unlikely to generally be a good representation of the true light level in the area of the project.[12] Figure 2 gives a sense of the fineness of the spatial grid and the global coverage of WB aid.

I examine both if poorer cells are more likely to receive some aid and if

[12]This approach drops about half of the data, which was either coded with less precision or was coded with a code of 8, which is applied to aid that is assumed to go to a regional or national capital (e.g. aid for capacity building). See Briggs (2018b) for further discussion of the implications of sub-setting the data this way. In Appendix A I show that the results hold when including projects with a precision code of up to 3 (geocoded to centroids of ADM2 regions). When using this cruder geocoding, I am working with 73% of the rows in the WB dataset. Finally, it should be noted that the results of the present analysis are quite similar to other analyses that use less precisely geocoded aid and aggregate aid into accordingly larger spatial units (Briggs 2017; Öhler et al. 2019).

poorer cells are more likely to receive a greater dollar amount of aid when they do get aid. The measure of aid selection is binary. Cell-years that have a newly initiated aid project are marked with a one and are otherwise zero. The measure of aid intensity is continuous and is the natural log of the total dollar value of new aid per cell-year. When I calculate the dollar value of aid per cell-year, I follow prior work (e.g. Öhler et al. 2019; Briggs 2017) and evenly split the commitment amount of each project across its locations.

We lack grid-cell level measure of poverty that are consistent over time and countries, so I proxy for the wealth of each cell-year using mean nighttime light emission from the DMSP-OLS Nighttime Lights Time Series Version 4 (Average Visible, Stable Lights, and Cloud Free Coverages).[13] This variable is correlated with household wealth (Weidmann and Schutte 2017; Noor et al. 2008) and poverty rates (Wang et al. 2012; Proville et al. 2017; Elvidge et al. 2009). I use the natural log of the standardized light variable.

I use HYDE population data, which is available for 1990, 2000 and 2005 (Goldewijk et al. 2011, 2010).[14] I fill in the values between the anchor years using linear interpolation, which seems reasonable as population is typically slow moving.[15] I do not extrapolate past the last population estimate, so the

[13]Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by US Air Force Weather Agency. The light measure was standardized to be between zero and one.

[14]HYDE is likely preferable to GPW, as the former is has more modeling and so is less likely to spread regional counts of people over cells that are very unlikely to have anyone in them due to environmental factors (see discussion in (Briggs 2018b)).

[15]Linear interpolation is preferable to carrying the last estimate forward because if one did this then the population variable would display large and discontinuous changes when an estimate was updated. This sharp change is unlikely to be a good representation of the true (but unknown) population count. Linear interpolation is also preferable to condensing

final year of the analysis is 2005 when using the HYDE variable. One can check that the interpolation is not driving the result by noting the similarity of the relationship between light and aid in the anchor years (2000, 2005) and the interpolated years in Figure 3. In the appendix, I show that the results are robust to using GPW population data, which extends the analysis out to 2010 (CIESIN and CIAT 2005). I use the natural log of the population variable (plus one).

Moving to estimation, models 1 and 2 in Table 1 have a binary dependent variable and estimation is done using logistic regression. Only about 2% of cell-years in the analysis receive new aid projects. While this means that receiving aid is a rare event, regular concerns around analyzing rare event data with logistic regression are unlikely to apply to the present analysis because the magnitude of the bias caused by rare events is decreasing in sample size and the present sample is large (King and Zeng 2001). Beck (2015) presents results supporting the appropriateness of a conditional fixed effects logistic model when using a dataset with many groups and not overly many observations per group, which is a good description of the dataset under analysis. Standard errors are based on 1000 bootstrap replications with 618 country-year clusters. Models 3 and 4 limit the sample to cell-years that received new aid and have a continuous dependent variable. Estimation is done using OLS and standard errors are clustered on country-years.

---

the data into five or ten year panels, as this would make it impossible to untangle if aid caused an increase in light or if aid merely went to places with more light.

2.2 Results

Conditional on population, places with more light at night are more likely to receive aid (model 1) and when they get aid they get more in dollar terms (model 3). These relationships are not simply being driven by aid targeting capital cities, as they hold after adding a dummy variable marking cell-years with centroids that are within 100km from the country's capital city (models 2 and 4). These results support past research finding that aid flows to wealthier parts of countries (Öhler and Nunnenkamp 2014; Öhler et al. 2019; Custer et al. 2017; Briggs 2017, 2018a,b).

Table 1: Descriptive analysis of subnational poverty targeting

	(1)	(2)	(3)	(4)
$\ln(\text{light})_{t-1}$	1.532***	1.500***	0.235***	0.222***
	(0.110)	(0.109)	(0.043)	(0.043)
$\ln(\text{population})_t$	2.112***	2.096***	0.082***	0.080***
	(0.072)	(0.069)	(0.018)	(0.018)
$<$100 km to capital_t		1.274***		0.086**
		(0.084)		(0.035)
Dependent variable	Binary	Binary	ln(aid cost)	ln(aid cost)
Model	Logit	Logit	OLS	OLS
Country-year FEs	Yes	Yes	Yes	Yes
n country-years	618	618	638	638
n cell-years	297,343	297,343	7,526	7,526

Models 1 and 2 show odds ratios and have bootstrap standard errors based on 1000 replications and 618 country-year clusters in parentheses. Models 3 and 4 show standard errors clustered on country-years in parentheses. *** p<0.01, ** p<0.05

To test for heterogeneity in the conditional relationship between light and aid, I run the specifications used to produce models 1 and 3 in Table 1 on each annual and regional subset of the data. Figure 3 graphs the coefficients and 95% confidence intervals for $\ln(\text{light})_{t-1}$ from these analyses. The top two panels show results from the selection model (model 1 in Table 1) and the bottom two show results for the intensity models (model 3). Any estimate to the left of the vertical red lines in Figure 3 shows pro-poor aid targeting. Nearly all point estimates show pro-rich aid targeting. There is no obvious time trend in the relationship between light and aid. Sub-Saharan Africa has stronger pro-rich aid targeting than other regions, a finding consistent with past research (Öhler et al. 2019). South Asia and Latin America have lower levels of pro-rich aid targeting. In East Asia, richer places are not much more likely to receive aid but when they do receive aid they receive more than poorer places.

I subject this descriptive result to a large number of robustness tests, all of which are presented in Appendix A and only briefly described here. First, I run model 1 and 3 in Table 1 while sequentially dropping all cell-years with fewer than 10, 100, 1000, and then 10,000 people. This tests to see if the results are being driven by measurement error at the low end of the population variable. Second, to ensure that assumptions about functional form are not driving the results, I run models 1 and 3 in Table 1 with an unlogged light at night variable. Third, I run models 1 and 3 in Table 1 but rather than using a continuous measure of light at night I enter a set of dummy variables marking the light quintile (within each country-year) to which each cell-year belongs. Fourth, I reproduce the OLS results in Table 1 first using the unlogged dollar amount of aid per cell-year as the dependent variable and with a Poisson pseudo-maximum likelihood model

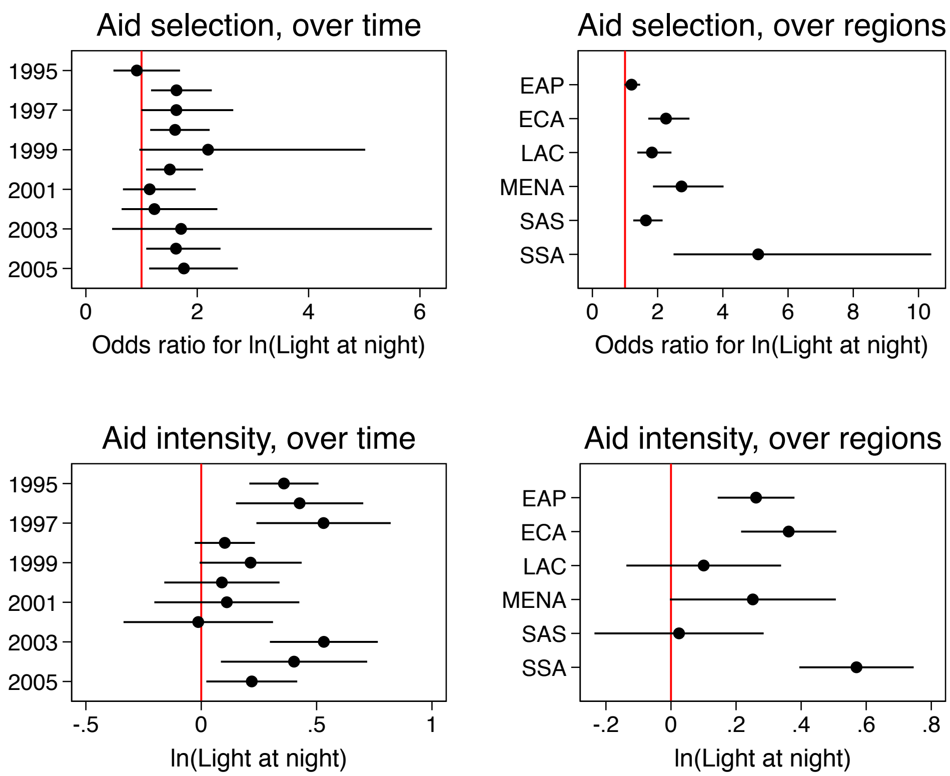

Figure 3: Heterogeneity in poverty targeting over years and regions. EAP = East Asia & Pacific, ECA = Europe & Central Asia, LAC = Latin America & Caribbean, MENA = Middle East & North Africa, SAS = South Asia, SSA = Sub-Saharan Africa

(Santos Silva and Tenreyro 2006, 2011), then clustering standard errors on countries instead of country-years, and then after re-weighting the cell-years so that each country-year has the same weight rather than each cell-year. Fifth, I replicate Table 1 and Figure 3 using a different population variable and five additional years of data. Finally, I replicate Table 1 but include projects that were geocoded to second-level regional centroids. With only very minor exceptions, all of these robustness tests show that aid flows to richer places. I find no evidence of subnational poverty targeting.

3 Why does aid not target the poorest?

The prior descriptive results tell us that aid is flowing to richer places within recipient countries, but it does not tell us why this is happening. This section tests five explanations for a lack of poverty targeting using a conjoint survey experiment run on WB TTLs. First, client governments may not be interested in poverty targeting. Second, it may be easier to get projects in richer areas approved. Third, project implementation may be more difficult in poorer areas. Fourth, expected project ratings may be lower in poorer areas. Fifth, TTLs may think that projects in wealthier areas will have a larger impact on development.

Prior to running the survey, I pre-registered a qualitative description of my design and a complete pipeline of code that moves from the raw survey data to the final presentation of the results. All of what I describe in sections

3.1 and 3.2 is based on this pre-registered analysis.[16] Section 3.3 presents an exploratory (not pre-registered) analysis of heterogeneous treatment effects.

3.1 Research design

TTLs are the Bank's main point of contact with the borrower for a project. On paper, TTLs are assigned to projects that client governments already want to complete, but in practice TTLs often have agency in discovering relevant projects and promoting them to recipient governments. TTLs have career incentives to get many projects approved and to have them be rated well, so they should be attuned to my survey questions about the factors that make projects more likely to be approved and then rated well.

On May 1, 2019, I emailed all WB staff who were listed as serving as a TTL on a project from 2008 until Feb 2019.[17] In order to not select on people who had strong opinions about poverty targeting, my invitation to take part in the survey described it honestly but generally as "a research study to understand the career incentives facing Task Team Leaders and how these relate to project selection."[18] After three weeks, I emailed a reminder to all

[16] A number of small changes to the pre-registered code are described in Appendix 7. The changes are analytically inconsequential and mostly have to do with tweaking the presentation of the graphs.

[17] I downloaded the TTL data on Feb 6, 2019, so I use all names from 2018 until then. I resorted to emailing TTLs after unsuccessfully trying to work through official WB channels for about 1 year. Directly emailing TTLs is possible because the WB project API lists the full name of the TTL associated with each project and the WB algorithmically creates email addresses based on names (first initial + last name @worldbank.org). This process created 2478 email addresses, of which 38 were not unique. I manually de-duplicated these non-unique email addresses by googling the names of each of the 80 associated TTLs.

[18] This wording should reduce concerns around experimenter demand effects, which in

TTLs who had yet to complete the survey. Data collection closed one week later, on the last day of the month.

I anticipated a low response rate because "World Bank survey response rates are generally low and tend to be declining over time" (Smets 2018, p. 4). Indeed, in 2015 the IEG's client survey (of WB staff, the Board of Directors, and external stakeholder) had a response rate of 4.7% (Smets 2018, p. 4). As I was running an unofficial survey based on cold-emailing TTLs, I predicted a response rate that was even lower. Low response rates can lead to bias if there are heterogeneous treatment effects and if selection into the survey is correlated with treatment effects (Franco et al. 2017). Further, even with no expected bias, small samples make it more likely that one gets an unlucky draw of respondents that fails offer a good representation of the population of interest. I emailed 2478 TTLs, of whom 115 completed the survey. Thus, while low, my response rate roughly matches some official WB surveys.[19]

In anticipation of a low response rate, I asked respondents the region of the world and the sector in which they had done the largest share of their work at the WB. I use the WB's official sectors and regions, and then I weight respondents so that they match the sectoral and regional distribution of all WB projects from 2008 until March 2019.[20] In practice, my sample

---

general seem weak (Mummolo and Peterson 2019).

[19] It is difficult to work out the exact response rate. Given that I algorithmically generated email addresses from names, it is probable that a few hundred of the email addresses were incorrect. Unfortunately, Qualtrics did not record bounced emails. If we assume a denominator of 2300, which is conservative, then the response rate is 5% ($115 \div 2300 = 0.05$).

[20] I downloaded the project data for weighting on March 13, 2019, so I include all projects up to that date.

approximates the population on these two dimensions and the weighting does little to alter the results.[21]

The conjoint experiment asked respondents to choose one of a pair of hypothetical aid projects that were in the sector and region of the world where they had previously stated that they had the most experience. Each project varied on five dimensions, all of which were independently randomly assigned. Projects could have an *average income of the project location* value of above national average, at national average, or below national average. Projects could be *located* in the capital city, an urban area, the outskirts of a city, a rural area, or a remote area. Projects had a *political affiliation* of the president's hometown, an area where residents favor the party in power, an area where residents favor an opposition party, an area where residents have weak partisan affiliations, or an area with no political affiliation. Each project had a *budget* that was larger than typical, typical, or smaller than typical. Finally, the *implementing partner* for each project was either the client government or a non-governmental organization. This setup yields 450 possible kinds of projects.

As previously described, aid may fail to target poorer parts of countries for at least five reasons. Each of these reasons maps to a dependent variable in the conjoint experiment. In order for a project to be implemented and so enter into the dataset analyzed in Table 1, it must first interest the client government. Accordingly, I ask respondents to select the project that would be of greater interest to the client government. Then the project must be

[21]This is demonstrated in Appendix B.

approved through the hierarchy of the WB. I thus ask respondents which project would be easier to get WB approval.[22]. Projects are then implemented and rated, and I ask respondents which project would be easier to implement and also which project they think would receive a higher outcome rating. Finally, I ask respondents which project they think would have a larger positive impact on development (using their own personal definition of development). These five dependent variables were presented in random order for every pair of profiles.

Each respondent was shown five pairs of projects, and for each pair of projects the respondent made five binary choices between the projects (one per dependent variable). This means that every respondent produced 10 observations, each with data on all five dependent variables. Thus, each panel in Figure 5 and 6 is based on an analysis of 1150 observations from 115 respondents. The graphical presentation of the results in Figure 5 and 6 shows marginal means, as ultimately I am interested in how likely it was that an aid project with a given feature level was selected, marginalizing over all other features (Leeper et al. 2018). All features of projects are independently randomly assigned, so both projects in a pair can end up with the same feature level. Thus, the marginal means will range from the probability of co-occurrence to one minus the probability of co-occurrence (Leeper et al. 2018). In the case of a variable with only two levels, this implies a possible range of 0.25 to 0.75. All other variable have a wider range. A final point

[22]The precise question was "From project concept note to Board approval, which project would be easier to get approved?"

on interpretation is that in a forced choice design if one level of a variable is selected more often then another level of the same variable will necessarily be selected less often.[23] Because of this lack of independence within variables, one should evaluate the results of each independent variable holistically.

I calculate the marginal means using OLS with survey weights.[24] Figures 5 and 6 show point estimates and 95% confidence intervals based on standard errors clustered on respondents. While I would like readers to focus more on the probability of a project with a given feature level being picked rather than the statistical significance of differences between feature levels (the average marginal component effect), one can easily eyeball the statistical significance of the AMCE by comparing the point estimates and confidence intervals of the marginal means across values of a given independent variable.

3.2 Results

The correlations between the dependent variables are shown in Figure 4. All of the correlations are positive. Projects that are expected to have an easier approval process and also expected to get better ratings and have a higher impact on development. The weakest correlations are between ease of project implementation and development impact, though even this correlation remains positive.

[23] Appendix A in Leeper et al. (2018) has a very good discussion of related issues.

[24] For each independent variable, I run a regression of the dependent variable on every level of the independent variable (dropping the constant term). Thus, each panel in Figure 5 and 6 is made up of five unique regressions, one per independent variable. This simple approach relies on the fact that all feature levels are independently randomly assigned to profiles.

Figure 4: Correlation between dependent variables.

The full results of the conjoint experiment are presented graphically in Figures 5 and 6. My discussion of the results focuses on each dependent variable in sequence, and I discuss both the implications of the results for poverty targeting and for wider related questions in development studies and political science.

I find no support for the idea that client governments are more interested in projects that are placed in richer areas. Rather, client governments are perceived to be modestly more interested in aid that targets poorer areas. They are also thought to be more interested in aid to remote parts of countries and against aid to urban areas. Client governments are thought to be indifferent to the size of the project budget. The largest effect sizes are for the political affiliation variable. Clients are thought to be most interested in projects that are placed in the president's hometown. When shown a project located in the president's hometown, TTLs select that project to be of greater interest to the client government more than 70% of the time. This supports claims that leaders favor regions that share an ethnicity with leader (Franck and Rainer 2012) or that simply hold the leader's birthplace (Hodler and Raschky 2014). Clients also prefer to target aid to their core supporters, a finding in line with Jablonski (2014) and Briggs (2014). I find no evidence of a desire to target swing voters. Recipient governments prefer projects where they are the implementing partner to projects which bypass them and implement using NGOs.

I also find no support for the idea that projects in poorer areas are harder to get approved. In fact, TTLs think that projects in poorer areas are easier to get approved. TTLs also report that it is easier to get projects approved

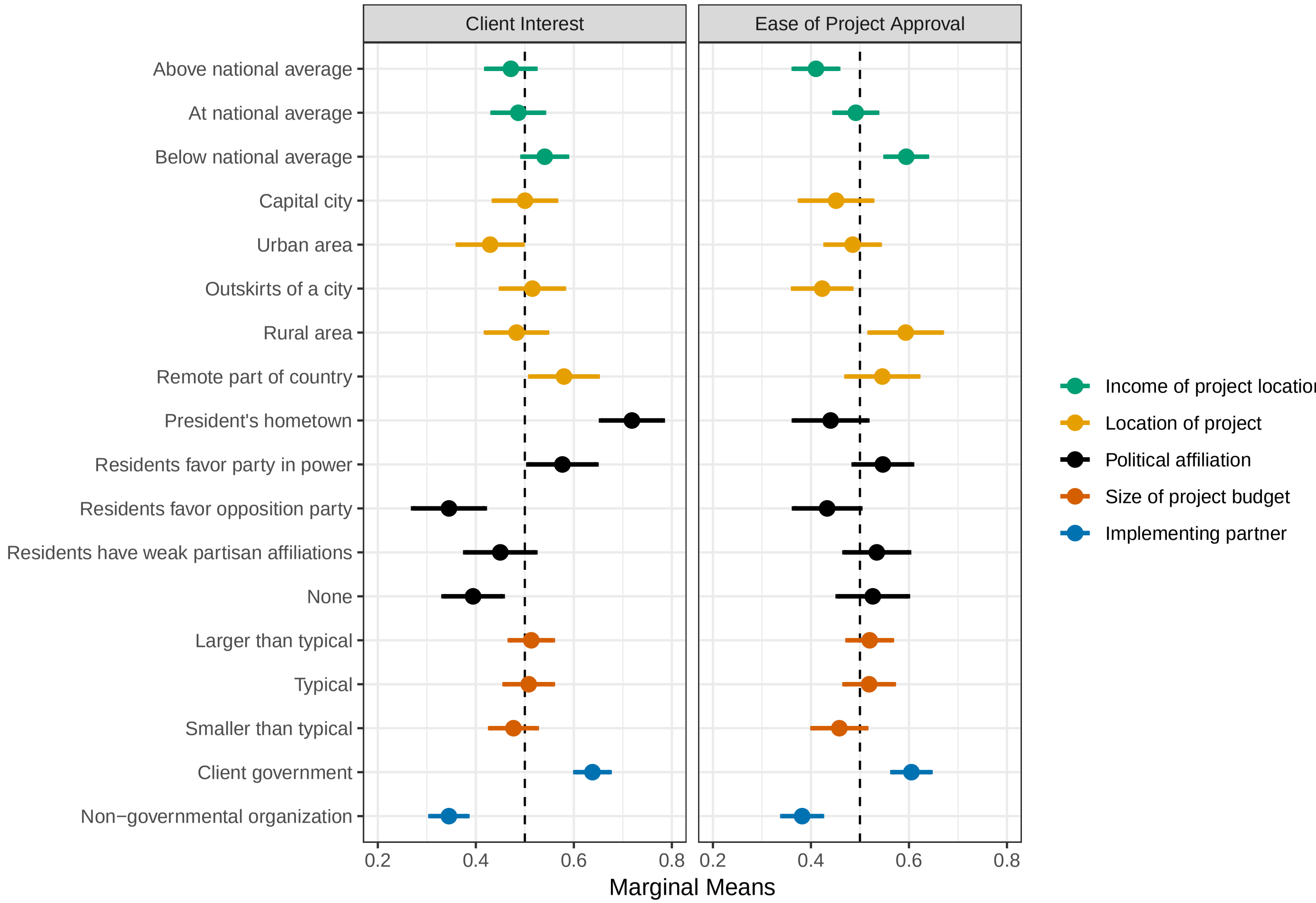

Figure 5: Effect of project attributes on probability of selection.

if they take place in rural or remote areas. While the prior results show that client governments prefer projects that are located in the president's hometown, such projects are harder to get approved. This may be one reason why the WB—unlike China—is not more likely to place aid in the president's hometown (Dreher et al. 2019). The WB is thought to have a preference for lending volume, and this is also borne out by the present results as TTLs think that projects with smaller budgets are harder to get approved. One interesting and unexpected result is that approval is fairly easy when the project's location has residents that favor the party in power (but is not the president's hometown). This suggests a tolerance for core-voter targeting in WB aid, which has been shown to exist in at least Kenya (Briggs 2014; Jablonski 2014). It is easier to get approval when the project does not bypass the client government.

Implementation is thought to be harder in poorer areas, rural areas, and remote areas. This supports explanations for pro-rich targeting that prioritize ease of access to project areas. TTLs face time and budget constraints, and so they may want to focus their attention closer to cities to economize on time and money. It should be emphasized that this dependent variable is distinct from outcome ratings or development impact, and so is capturing ease of implementation but is not not a measure of the likelihood that a project 'works'. Implementation is also thought to be easier in the president's hometown and areas of core support, and harder in areas of opposition support. This result can be read as supporting the importance of recipient ownership. When aid is targeted to places that recipients want to help, they may work harder to ensure that implementation is smooth.

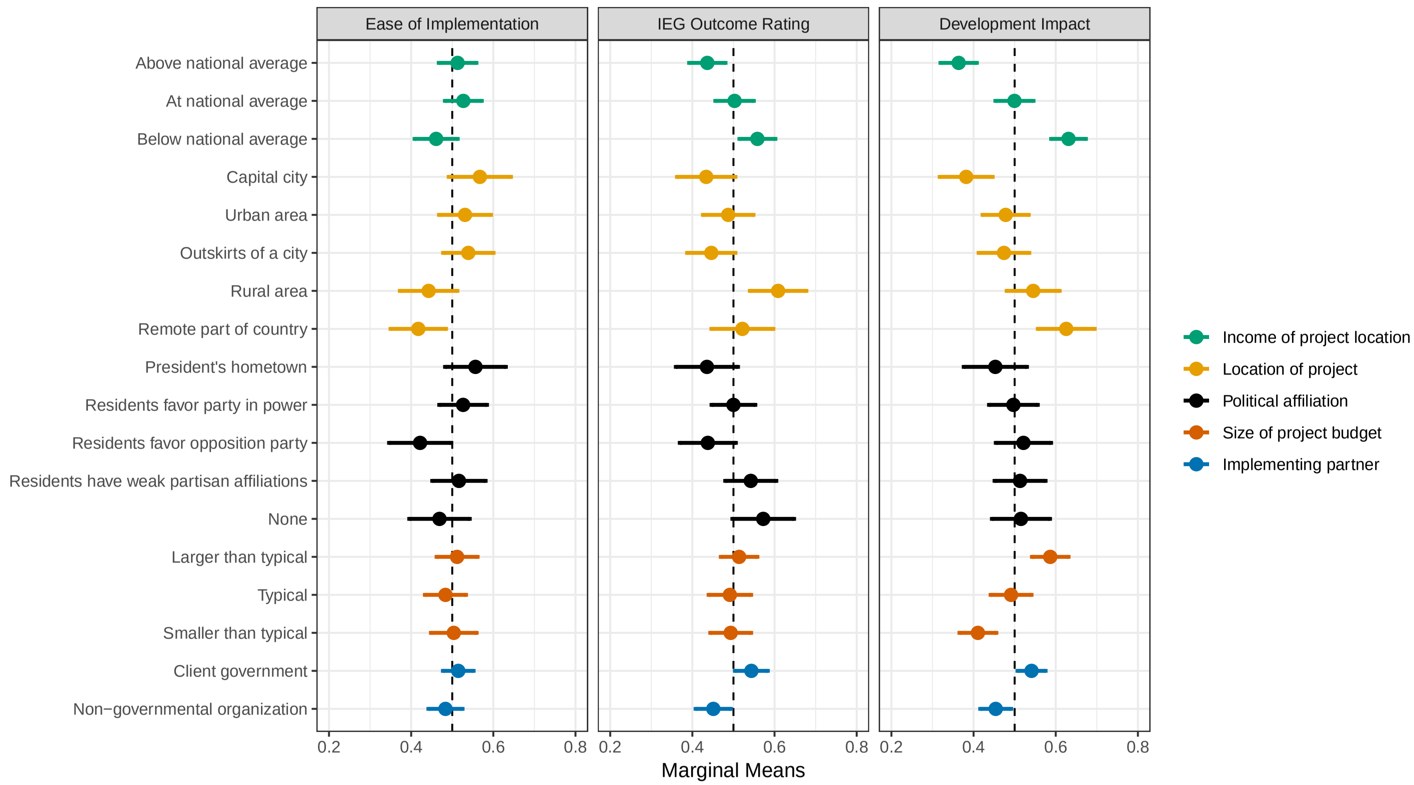

Figure 6: Effect of project attributes on probability of selection.

The outcome rating results also should enable pro-poor targeting. Outcome ratings are expected to be lower in richer areas, capital cities, and the outskirts of a city. They are expected to be higher in rural areas and poorer areas. These patterns suggests that outcome ratings are capturing development results in addition to outlays or the ability to achieve pre-determined goals (Thomas and Tominaga 2010).[25] Outcome ratings are expected to be higher in places with no political affiliation or weak partisan affiliations and lowest in the president's hometown or places that favor the opposition party. The former explanation could be due to fear of capture and the latter could be due to expected difficulties in implementation, but this is conjecture. TTLs expect that NGO-implemented aid will receive a lower outcome rating. This contradicts the results of an analysis of the effect of NGO-implementation using observational data on WB projects in Shin et al. (2017), but it is consistent the results of a similar observational analysis in Winters (2019).

Finally, TTLs think that aid has a larger positive impact on development when it is targeted to poorer areas and more remote areas. This suggests that TTLs either: view aid as primarily being about ameliorating the negative effects of poverty, or think that aid generally works better in poorer areas, or think that aid will be more likely to cause growth if targeted to poorer and more remote areas. The latter explanation seems unlikely, but the experimental data cannot rule it out. The second explanation clashes with the evidence that implementation is harder in poorer and more remote

[25]Kilby (2015, p. 117) suggests that outcome ratings are (informally) influenced by rates of return in addition to their stated goal of measuring the extent to which a project met pre-determined objectives.

places, so it seems unlikely. Thus, this probably reveals that TTLs think that the main goal of aid is to ameliorate the negative effects of poverty. TTLs think that projects with larger budgets do more good than smaller projects. Aid placed in the president's hometown has a somewhat lower development impact, which again suggests that TTLs fear capture. TTLs think that bypass aid is less effective than aid where the client government is the implementing partner.

3.3 Exploratory results

This final section discusses an exploratory (not pre-registered) analysis of heterogeneous treatment effects. Figure 3 in the descriptive analysis showed that aid from the WB had the strongest pro-rich targeting in sub-Saharan Africa. This supports similar results in Öhler et al. (2019). In response to these findings, I examine if TTLs whose work at the Bank mostly focused on sub-Saharan Africa respond differently to the conjoint experiment than those who focused on other regions.

Forty-six TTLs did the largest share of their work in Africa, while the remaining 69 did most of their work in other regions.[26] Figure 7 show the results when I replicate the analysis used to produce Figures 5 and 6 but run the analysis separately for Africa and the rest of the world.[27]

[26] I am unable to produce a similar analysis by sector because each sector has too few respondents. If I create sector groupings like a 'social sector' group with health and education, I have only 18 respondents (and 180 observations) in this group. I thus focus only on regional heterogeneity.

[27] This analysis excludes survey weights, primarily because my weights were calculated on the full sample. The top panels in Figure 7 (Africa) are based on an analysis with 460

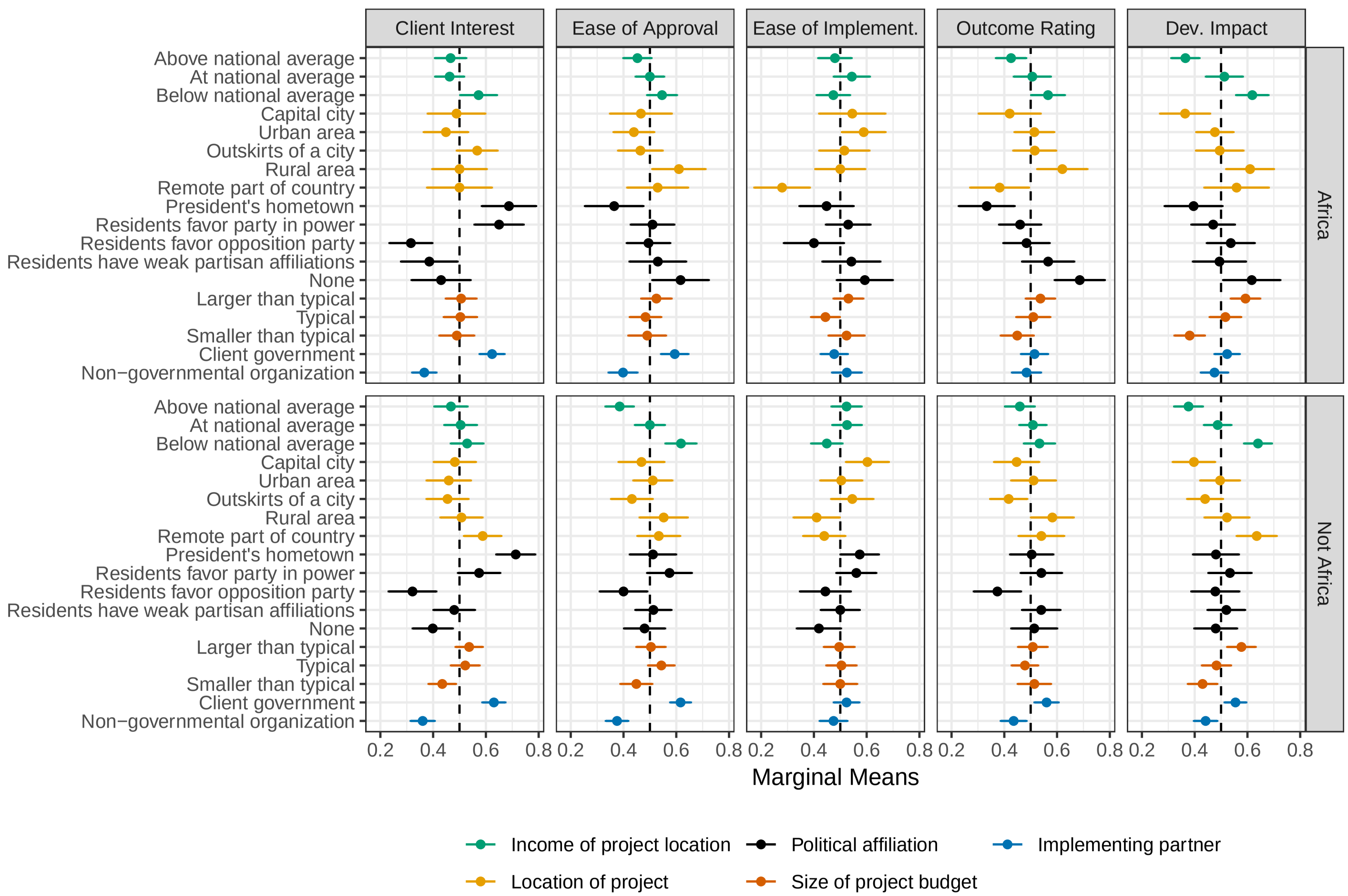

Figure 7: Effect of project attributes on probability of selection.

The client interest results are quite similar across the two groups. TTLs that work more in Africa are more likely to think that it is hard to get projects approved if the project is based in the president's hometown, and they are more likely to think that approval is easier if the location has no political affiliation. Probably the starkest result is that TTLs who work more in Africa are much more likely to think that implementation in remote areas is difficult. All TTLs think that implementing in remote areas is more difficult, but the magnitude of the African result stands out and may explain why aid to African countries is more pro-rich than aid to other world regions.

Outcome ratings are similar, though TTLs who work in Africa are more likely to expect that projects in remote areas will receive poor outcome ratings. Again, this could help explain why aid to Africa is more pro-rich than aid to other regions. TTLs who work more in Africa are also more likely to think that outcome ratings will be higher in apolitical areas or swing areas and worse in the president's hometown. TTLs working outside of Africa think that outcome ratings will be worst in opposition areas. The development impact results are generally similar, with the one exception being that only TTLs who worked more in Africa think that developmental impact is higher in areas with no political affiliation than presidential hometowns.

One plausible explanation for many of these political results is that aid capture is higher in African presidential hometowns, so only in Africa do we see presidential hometowns having more difficult implementation, lower outcome ratings, and lower expected development impact. This might explain

---

observations and the bottom panels have 690 observations. Other than not using survey weights, the analyses are the same as those used to produce Figures 5 and 6.

why only in Africa is it harder to get projects in presidential hometowns approved. Outside of Africa, approval, implementation, and outcome ratings are lower in areas that support the opposition. Outside of Africa it is also thought to be easier to implement in presidential hometowns and that such places are no worse in terms of outcome ratings or development impact. One plausible story is that outside of Africa, governments put more effort into favoring presidential hometowns with aid-funded public goods and that such aid is not more prone to capture. It may then be harder to implement in opposition areas simply because governments care less about helping them.

This section has revealed two sources of heterogeneity that may explain why aid is least likely to reach the poorest within African countries. First, implementation in remote areas is perceived to be harder in Africa than in other regions. Second, outcome ratings in remote areas are expected to be lower only in Africa. Many of the additional results in this section are consistent with a story where aid to presidential hometowns in Africa is uniquely prone to capture while aid outside of Africa is hardest to implement in opposition held areas but easier to implement in areas where government are motivated to help, such as presidential hometowns. These results come from an exploratory analysis of heterogeneity, and one avenue for future work is to confirm these results.

4 Conclusion

This paper has shown that WB aid does not flow to poorer parts of countries and has suggested that this is unlikely to be due to client governments being more interested in directing aid to richer areas, an easier approval process for projects placed in richer areas, or the belief that aid projects placed in cities or richer areas have a bigger effect on development. In fact, TTLs think that projects placed in poorer areas have an easier approval process and that the development impact of aid is higher in poorer places. The one explanation for pro-rich aid targeting to survive these tests is implementation concerns. Aid projects are thought to be harder to implement in poorer places, rural areas, and remote parts of countries. Perhaps aid is steered away from such areas because implementation there is time consuming and incentive structures within the WB encourage TTLs to select projects that are easy to implement.

I also examined heterogeneity in both the descriptive analysis of aid targeting and in the conjoint experiment. The descriptive analysis showed that Africa has the most extreme pro-rich aid targeting both in terms of aid selection and aid intensity. The conjoint survey experiment revealed that TTLs who have done most of their work in Africa think that implementation is harder in remote areas than do TTLs who worked elsewhere. TTLs who worked more in Africa also think that projects in remote areas are likely to get lower outcome ratings, a finding which does not exist in the group that mostly worked outside the continent. Both groups of TTLs agree that projects targeted to poorer and harder-to-access parts of countries have a larger positive impact on development than projects targeted to richer or more urban places.

The survey analysis also yielded a number of more broadly interesting results. First, on the politicization of aid, TTLs claim that client governments want to target aid to the president's hometown and to their core supporters. There is no evidence that client governments want to target aid to swing voters. Less surprisingly, TTLs thought that client governments were the least likely to be interested in projects that target opposition supporters. TTLs in Africa also think it is hard to get approval for aid projects that are targeted to the president's hometown, a result that offers a plausible explanation for the finding that in Africa Chinese aid but not WB aid favors the president's hometown (Dreher et al. 2019).

Second, there are a series of interesting results on the pros and cons of directing aid to presidential hometowns. Clients want aid to go to the president's hometown and, outside of Africa, projects placed in presidential hometowns are thought to be the easiest to implement. In Africa, aid in presidential hometowns is expected to have the lowest development impact and a lower outcome rating. This is plausibly explained by a story where aid to presidential hometowns in Africa is more prone to capture while such aid outside of Africa finds a partner government that is more motivated to achieve success in producing effective public goods.

Third, TTLs generally viewed bypass aid skeptically. They think that clients like it less, that it is harder to get projects that bypass the client government approved, that bypass projects will get lower outcome ratings, and that bypass projects have a smaller development impact than projects where the client government is the implementing partner. The last result especially suggests the importance in working with rather than around client governments when possible.

Finally, budget size is sometimes used to proxy for project complexity (e.g. Denizer et al. 2013). The survey results cast some doubt on the validity of this proxy. TTLs do not think that projects with larger (or smaller) budgets are any harder to implement or have worse outcome ratings. They do think that more expensive projects have a larger impact on development, but one would hope to find this result given that the question is about absolute outcomes produced and not some measure of outcomes per dollar.

The present research also has a number of limitations that could be addressed in future work. First, both analyses report average effects across the WB's entire portfolio. There may, however, be sector-level heterogeneity that is masked by averaging. I lack a sufficiently large number of respondents to my survey to examine such heterogeneity, but it plausibly exists and could be examined in future work. Second, both analyses focused on only one donor. Future work could extend the analyses in this paper to more donors. In particular, the survey analysis could be run on other donors at low cost.

Past research on subnational poverty targeting noted that positive correlations between aid and income "should not be read as showing that aid is being targeted badly [because it is] entirely possible that aid is flowing to the places where it can be used most effectively and that [these] places tend to be places of relative wealth" (Briggs 2018a, p. 908). The present paper can now go somewhat further. TTLs at the WB believe that aid works better for development when it is targeted to poorer and more remote parts of countries—but this is not happening. The only explanation for this pattern to survive the conjoint experiment was that implementation is more difficult in poorer and more remote areas, though in Africa aid to remote areas is also expected to get lower outcome ratings. Tweaks to WB incentive structures that make ease of project implementation less important, or that better condition outcome ratings on the difficulty of remote contexts, may encourage aid to flow to poorer parts of countries.

References

Ajmera, M. and Fields, G. A. (2016). *Invisible Children: Reimagining International Development at the Grassroots*. Springer.

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

Bandyopadhyay, S. and Wall, H. J. (2006). The determinants of aid in the post-cold war era. *FRB of St. Louis Working Paper No 2006-021B*.

Bardhan, P. (1996). Efficiency, equity and poverty alleviation: Policy issues in less developed countries. *The Economic Journal*, 106(438):1344–1356.

Bates, R. H. (1981). *Markets and States In Tropical Africa: The Political Basis of Agricultural Policies*. University of California Press, Oakland, CA.

Beck, N. (2015). Estimating grouped data models with a binary dependent variable and fixed effects: What are the issues? *Working Paper presented at the annual meeting of the Society for Political Methodology*.

Brass, J. (2012). Why do ngos go where they go? evidence from kenya. *World Development*, 40(2):387–401.

Briggs, R. C. (2014). Aiding and abetting: Project aid and ethnic politics in Kenya. *World Development*, 64:194–205.

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

Briggs, R. C. (2018a). Leaving no one behind? A new test of subnational aid targeting. *Journal of International Development*, 30(5):904–910.

Briggs, R. C. (2018b). Poor targeting: A gridded spatial analysis of the degree to which aid reaches the poor in Africa. *World Development*, 103:133–148.

CIESIN and CIAT (2005). Gridded population of the world, version 3 (gpwv3): Population count grid. doi:10.7927/h4639mpp. accessed 03.06.2013.

Custer, S., DiLorenzo, M., Masaki, T., Sethi, T., and Wells, J. (2017). Beyond the tyranny of averages: Development progress from the bottom up. *AidData Policy Report*.

Denizer, C., Kaufmann, D., and Kraay, A. (2013). Good countries or good projects? macro and micro correlates of World Bank project performance. *Journal of Development Economics*, 105:288–302.

Deutscher, E. and Fyson, S. (2008). Improving the effectiveness of aid. *Finance and Development*, 45(3):15–19.

Dietrich, S. (2013). Bypass or engage? explaining donor delivery tactics in foreign aid allocation. *International Studies Quarterly*, 57(4):698–712.

Dietrich, S. (2016). Donor political economies and the pursuit of aid effectiveness. *International Organization*, 70(1):65–102.

Dietrich, S., Mahmud, M., and Winters, M. S. (2018). Foreign aid, foreign policy, and domestic government legitimacy: experimental evidence from Bangladesh. *Journal of Politics*, 80(1):133–148.

Dietrich, S. and Winters, M. S. (2015). Foreign aid and government legitimacy. *Journal of Experimental Political Science*, 2(2):164–171.

Dollar, D., Kleineberg, T., Kraay, A., et al. (2016). Growth still is good for the poor. *European Economic Review*, 81(C):68–85.

Dollar, D. and Kraay, A. (2002). Growth is good for the poor. *Journal of economic growth*, 7(3):195–225.

Dollar, D. and Levin, V. (2006). The increasing selectivity of foreign aid, 1984–2003. *World Development*, 34(12):2034–2046.

Dreher, A., Fuchs, A., Hodler, R., Parks, B. C., Raschky, P. A., and Tierney, M. J. (2019). African leaders and the geography of China's foreign assistance. *Journal of Development Economics*.

Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B., and Bright, E. (2009). A global poverty map derived from satellite data. *Computers & Geosciences*, 35(8):1652–1660.

Findley, M. G., Harris, A. S., Milner, H. V., and Nielson, D. L. (2017). Who controls foreign aid? elite versus public perceptions of donor influence in aid-dependent uganda. *International Organization*, 71(4):633–663.

Franck, R. and Rainer, I. (2012). Does the leader's ethnicity matter? ethnic favoritism, education, and health in sub-saharan africa. *American Political Science Review*, 106(02):294–325.

Franco, A., Malhotra, N., Simonovits, G., and Zigerell, L. (2017). Developing standards for post-hoc weighting in population-based survey experiments. *Journal of Experimental Political Science*, 4(2):161–172.

Goldewijk, K. K., Beusen, A., Drecht, G. V., and Vos, M. D. (2011). The hyde 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecology and Biogeography*, 20(1):73–86.

Goldewijk, K. K., Beusen, A., and Janssen, P. (2010). Long-term dynamic modeling of global population and built-up area in a spatially explicit way: Hyde 3.1. *The Holocene*, 20(4):565–573.

Gulrajani, N. (2017). Bilateral donors and the age of the national interest: What prospects for challenge by development agencies? *World Development*, 96:375–389.

Hodler, R. and Raschky, P. A. (2014). Regional favoritism. *The Quarterly Journal of Economics*, pages 995–1033.

Jablonski, R. (2014). How aid targets votes: The impact of electoral incentives on foreign aid distribution. *World Politics*, 66(2):293–330.

Johns, B. and Torres, T. T. (2005). Costs of scaling up health interventions: a systematic review. *Health policy and planning*, 20(1):1–13.

Kilby, C. (2015). Assessing the impact of World Bank preparation on project outcomes. *Journal of Development Economics*, 115:111–123.

King, G. and Zeng, L. (2001). Logistic regression in rare events data. *Political analysis*, 9:137–163.

Leeper, T. J., Hobolt, S. B., and Tilley, J. (2018). Measuring subgroup preferences in conjoint experiments. *Working Paper*.

Marty, R., Dolan, C. B., Leu, M., and Runfola, D. (2017). Taking the health aid debate to the subnational level: the impact and allocation of foreign health aid in Malawi. *BMJ Global Health*, 2(1):e000129.

Masaki, T. (2018). The political economy of aid allocation in africa: Evidence from Zambia. *African Studies Review*, 61(1):55–82.

Mummolo, J. and Peterson, E. (2019). Demand effects in survey experiments: An empirical assessment. *American Political Science Review*, 113(2):517–529.

Neumayer, E. (2003). *The pattern of aid giving: The impact of good governance on development assistance*. Routledge, London.

Noor, A. M., Alegana, V. A., Gething, P. W., Tatem, A. J., and Snow, R. W. (2008). Using remotely sensed night-time light as a proxy for poverty in africa. *Population Health Metrics*, 6(1):5.

OECD DAC (2019). QWIDS: Query Wizard for International Development Statistics: About. https://stats.oecd.org/qwids/.

Öhler, H., Negre Rossignoli, M., Smets, L., Massari, R., and Bogetic, Z. (2019). Putting your money where your mouth is: geographic targeting of World Bank projects to the bottom 40 percent. *PloS one*, 14(6):e0218671.

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

Paciorek, C. J., Stevens, G. A., Finucane, M. M., Ezzati, M., Group, N. I. M. S., et al. (2013). Children's height and weight in rural and urban populations in low-income and middle-income countries: a systematic analysis of population-representative data. *The Lancet Global Health*, 1(5):e300–e309.

Proville, J., Zavala-Araiza, D., and Wagner, G. (2017). Night-time lights: A global, long term look at links to socio-economic trends. *PloS one*, 12(3):e0174610.

Santos Silva, J. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, 88(4):641–658.

Santos Silva, J. and Tenreyro, S. (2011). Further simulation evidence on the performance of the poisson pseudo-maximum likelihood estimator. *Economics Letters*, 112(2):220–222.

Shin, W., Kim, Y., and Sohn, H.-S. (2017). Do different implementing partnerships lead to different project outcomes? evidence from the world bank project-level evaluation data. *World Development*, 95:268–284.

Smets, L. (2018). Impact of phone reminders on survey response rates. *World Bank policy research working paper 8305*.

Smith, L. C., Ruel, M. T., and Ndiaye, A. (2005). Why is child malnutrition lower in urban than in rural areas? evidence from 36 developing countries. *World Development*, 33(8):1285–1305.

Strandow, D., Findley, M., Nielson, D., and Powell, J. (2011). *The UCDP-AidData Codebook on Geo-Referencing Foreign Aid. Version 1.1.* Uppsala Conflict Data Program, Uppsala, Sweden: Uppsala University.

Thomas, V. and Tominaga, J. (2010). Evaluation for better development results. *Journal of Development Effectiveness*, 2(3):371–386.

Tollefsen, A. F., Strand, H., and Buhaug, H. (2012). PRIO-GRID: A unified spatial data structure. *Journal of Peace Research*, 49(2):363–374.

Van de Poel, E., ODonnell, O., and Van Doorslaer, E. (2007). Are urban children really healthier? evidence from 47 developing countries. *Social science & medicine*, 65(10):1986–2003.

Wang, W., Cheng, H., and Zhang, L. (2012). Poverty assessment using dmsp/ols night-time light satellite imagery at a provincial scale in china. *Advances in Space Research*, 49(8):1253–1264.

WB IEG (2016). *Behind the Mirror: A Report on the Self-Evaluation Systems of the World Bank Group.*

WB IEG (2019). Ieg data and ratings. http://ieg.worldbankgroup.org/data.

Weidmann, N. B. and Schutte, S. (2017). Using night light emissions for the prediction of local wealth. *Journal of Peace Research*, 54(2):125–140.

Winters, M. S. (2014). Targeting, accountability and capture in development projects. *International Studies Quarterly*, 58(2):393–404.

Winters, M. S. (2019). Too many cooks in the kitchen? the division of financing in world bank projects and project performance. *Politics and Governance*, 7(2):117–126.

World Bank (1998). *Assessing Aid: What Works, What Doesn't, and Why.* The World Bank, Washington, DC.

World Bank (2009). *World Development Report 2009: Reshaping economic geography.*

World Bank (2018). Net ODA received (% of central government expense). https://data.worldbank.org/indicator/DT.ODA.ODAT.XP.ZS.

Young, A. (2013). Inequality, the urban-rural gap and migration. *The Quarterly Journal of Economics*, 128(4):1727–1785.

5 Appendix A: Robustness tests for descriptive analysis

This appendix presents robustness tests for the analyses shown in Table 1 and Figure 3. I first present results showing that measurement error at the low end of the population variable in unlikely to be affecting the results. I then show that the results are not sensitive to the functional form of the relationship between light and aid.

I then present three additional results for the OLS portion of Table 1. First, I replicate the results using an unlogged aid cost DV and a Poisson pseudo-maximum likelihood model, then I replicate Table 1 but cluster standard errors on countries instead of country-years, then I replicate it again after re-weighting the cell-years so that each country-year has the same weight rather than each cell-year.

I then replicate Table 1 and Figure 3 using a different population variable and five more years of data. With only very minor exceptions, all of these robustness tests show that aid flows to richer places. I find no evidence of subnational poverty targeting.

Finally, I replicate Table 1 but use a dataset that allowed projects with a precision code of less than 4 (the main text uses less than 3). This means that the dataset include the 26% of all rows that were geocoded to a regional centroid. The results are similar in direction and significance. As expected, the magnitude of pro-rich aid targeting is somewhat smaller.

Table 2 replicates model 1 in Table 1, but sequentially drops cell-years with fewer than 10, 100, 1000, and 10,000 people. Table 3 shows results from the same replication exercise for model 3 in Table 1. These tests show that measurement error at the low end of the population variable is not influencing the results.

In Table 2, the odds ratio for light is 1.5 and the odds ratio for population is 2.1. These results remain very consistent when dropping low population cell-years. In the most extreme case (model 4 of Table 2), about half of all cell-years are dropped but the results remain very similar.

Table 2: Aid Selection: Dropping low population cell-years

	(1)	(2)	(3)	(4)
$\ln(\text{light})_{t-1}$	1.517***	1.511***	1.487***	1.495***
	(0.112)	(0.112)	(0.105)	(0.103)
$\ln(\text{population})_t$	2.125***	2.131***	2.158***	2.155***
	(0.073)	(0.072)	(0.076)	(0.085)
Drop obs. with pop. below:	10	100	1000	10000
Country-year FEs	Yes	Yes	Yes	Yes
n country-years	618	618	618	614
n cell-years	258,757	240,241	198,954	138,913

Coefficients report odds ratios. Bootstrap standard errors based on 1000 replications and country-year clusters in parentheses. *** $p<0.01$

The results in Table 3 are quite similar in Table 1. This is unsurprising as aid typically targets places with higher population counts. In the most extreme case, dropping cell-years with fewer than 10,000 people (model 4 of Table 3) only reduces the number of aid-receiving cell-years by about 3%.

Table 3: Aid Intensity: Dropping low population cell-years

	(1)	(2)	(3)	(4)
$\ln(\text{light})_{t-1}$	0.231***	0.230***	0.231***	0.223***
	(0.043)	(0.043)	(0.044)	(0.045)
$\ln(\text{population})_{t}$	0.085***	0.086***	0.085***	0.102***
	(0.019)	(0.019)	(0.019)	(0.020)
Drop obs. with pop. below:	10	100	1000	10000
Country-year FEs	Yes	Yes	Yes	Yes
n of country-year	638	638	638	636
n cell-years	7,523	7,519	7,487	7,293

Robust standard errors clustered on country-years in parentheses
*** p<0.01

Table 4 replicates Table 1 but uses an unlogged light at night variable. This is done to test if the results are sensitive to assumptions about functional form. The coefficient for light at night is much larger in magnitude, as it now essentially represents the shift from total darkness (zero) to total brightness (one). Otherwise, the results are similar.

Table 4: Table 1, unlogged light at night

	(1)	(2)	(3)	(4)
light_{t-1}	5.668***	4.902***	1.343***	1.269***
	(1.947)	(1.830)	(0.271)	(0.277)
$\ln(\text{population})_t$	2.268***	2.244***	0.105***	0.102***
	(0.085)	(0.084)	(0.017)	(0.017)
$<$100 km to capital_t		1.311***		0.090**
		(0.082)		(0.035)
Dependent variable	Binary	Binary	ln(aid cost)	ln(aid cost)
Model	Logit	Logit	OLS	OLS
Country-year FEs	Yes	Yes	Yes	Yes
n country-years	618	618	638	638
n cell-years	297,343	297,343	7,526	7,526

Models 1 and 2 show odds ratios and have bootstrap standard errors based on 1000 replications and 618 country-year clusters in parentheses. Models 3 and 4 show standard errors clustered on country-years in parentheses. *** p<0.01, ** p<0.05

Table 5 also address functional form. However, in this case I break the light variable into quintiles *within each country-year* rather than using a continuous measure of light. I then replace the light variable in the analyses in Table 1 with a set of dummy variables marking the brightness quintile to which each cell-year belongs.

Two results emerge from this exercise. First, there is no significant difference in aid targeting at the low end of the light distribution. This makes sense, as the light variable is very heavily skewed and most of the dimmest 40% of cell-years have roughly (if not exactly) the same low light values.

Second, we see an increasing likelihood of receiving at least one aid project as one sequentially moves up each quintile from the 2nd dimmest until the brightest. This consistent increase tells us that aid is not merely avoiding the dimmest places or targeting the brightest places but rather that as cells get brighter they see an increasing chance of receiving at least one aid project. Within country-years and conditional on population, the odds of the brightest 20% of cells receiving aid are about three times higher than the odds of the dimmest (or 2nd dimmest) 20%. The odds of the 2nd brighest 20% are about twice as high as the dimmest 40%. The odds for the middle quintile are about 1.5 times higher. This same effect does not hold for the analyses of aid cost, where the brightest quintile clearly pulls away from the bottom four. Once a place is receiving aid, it gets more aid in dollar terms only if it is in the top 20% of the light distribution.

Table 5: Table 1, light quintiles

	(1)	(2)	(3)	(4)
2nd dimmest quintile$_{t-1}$	0.953	0.956	0.060	0.060
	(0.120)	(0.126)	(0.061)	(0.061)
Middle quintile$_{t-1}$	1.605***	1.604***	0.087	0.087
	(0.194)	(0.198)	(0.061)	(0.061)
2nd brightest quintile$_{t-1}$	2.093***	2.095***	0.078	0.081
	(0.271)	(0.277)	(0.062)	(0.062)
Brightest quintile$_{t-1}$	3.335***	3.300***	0.197***	0.192***
	(0.448)	(0.441)	(0.073)	(0.073)
ln(population)$_t$	1.979***	1.957***	0.131***	0.123***
	(0.074)	(0.074)	(0.017)	(0.017)
<100 km to capital$_t$		1.274***		0.125***
		(0.080)		(0.034)
Dependent variable	Binary	Binary	ln(aid cost)	ln(aid cost)
Estimation	Logit	Logit	OLS	OLS
Country-year Fes	Yes	Yes	Yes	Yes
n cell-years	297,343	297,343	7,526	7,526
n country-years	618	618	638	638

Models 1 and 2 show odds ratios and have bootstrap standard errors based on 1000 replications and 618 country-year clusters in parentheses. Models 3 and 4 show standard errors clustered on country-years in parentheses. *** p<0.01, ** p<0.05

Table 6 follows the guidance in Santos Silva and Tenreyro (2006) and Santos Silva and Tenreyro (2011) and replicates models 3 and 4 from Table 1 using an unlogged dependent variable and a Poisson pseudo-maximum likelihood model. More light is still positively and significantly correlated with more aid.

Table 6: Table 1, Models 3 & 4, PPML

	(1)	(2)
$\ln(\text{light})_{t-1}$	0.460***	0.459***
	(0.084)	(0.084)
$\ln(\text{population})_t$	0.082*	0.081*
	(0.045)	(0.045)
$<$100 km to capital_t		0.019
		(0.078)
Country-year FEs	Yes	Yes
n cell-years	7,504	7,504
n country-years	534	534

Robust standard errors clustered on country-years in parentheses. *** p<0.01

Table 7 replicates the OLS results (models 3 and 4) from Table 1, but clusters standard errors on recipient countries rather than country-years. The size of the standard errors barely changes.

Table 7: Table 1, Models 3 & 4, country-clustered errors

	(1)	(2)
ln(light)$_{t-1}$	0.235***	0.222***
	(0.051)	(0.052)
ln(population)$_t$	0.082***	0.080***
	(0.018)	(0.018)
<100 km to capital$_t$		0.086***
		(0.032)
Country-year FEs	Yes	Yes
n cell-years	7,526	7,526
n country-years	638	638

Robust standard errors clustered on countries in parentheses. *** p<0.01

Table 8 replicates the OLS results (models 3 and 4) from Table 1, but weights each observation so that each country-year (rather than each cell-year) has the same influence on the results. This ensures that atypical and larger countries are not driving the results. The influence of light and population is moderately stronger, but the results are similar.

Table 8: Table 1, Models 3 & 4, re-weighted observations

	(1)	(2)
$\ln(\text{light})_{t-1}$	0.297***	0.270***
	(0.065)	(0.063)
$\ln(\text{population})_{t}$	0.151***	0.148***
	(0.027)	(0.027)
$<$100 km to capital$_t$		0.109**
		(0.044)
Country-year FEs	Yes	Yes
n cell-years	7,526	7,526
n country-years	638	638

Robust standard errors clustered on country-years in parentheses.
*** p<0.01, ** p<0.05

Table 9 replicates Table 1 but uses the GPW population variable instead of HYDE. The GPW population data extend to 2010, so this also adds additional data to the dataset. The results are again somewhat stronger but similar.

Table 9: Table 1, GPW population variable

	(1)	(2)	(3)	(4)
$\ln(\text{light})_{t-1}$	1.824***	1.800***	0.264***	0.256***
	(0.145)	(0.141)	(0.038)	(0.038)
$\ln(\text{population})_t$	2.017***	1.998***	0.092***	0.089***
	(0.063)	(0.060)	(0.015)	(0.015)
$<$100 km to capital_t		1.227***		0.069**
		(0.066)		(0.030)
Dependent variable	Binary	Binary	ln(aid cost)	ln(aid cost)
Estimation	Logit	Logit	OLS	OLS
Country-year FEs	Yes	Yes	Yes	Yes
n cell-years	455,012	455,012	12,180	12,180
n country-years	969	969	982	982

Models 1 and 2 show odds ratios and have bootstrap standard errors based on 1000 replications and 618 country-year clusters in parentheses. Models 3 and 4 show standard errors clustered on country-years in parentheses. *** p<0.01, ** p<0.05

Figure 8 replicates Figure 3 but use the GPW population variable instead of HYDE. The GPW population data extend to 2010, so the annual cross-sections now also run out to 2010 (instead of 2005).

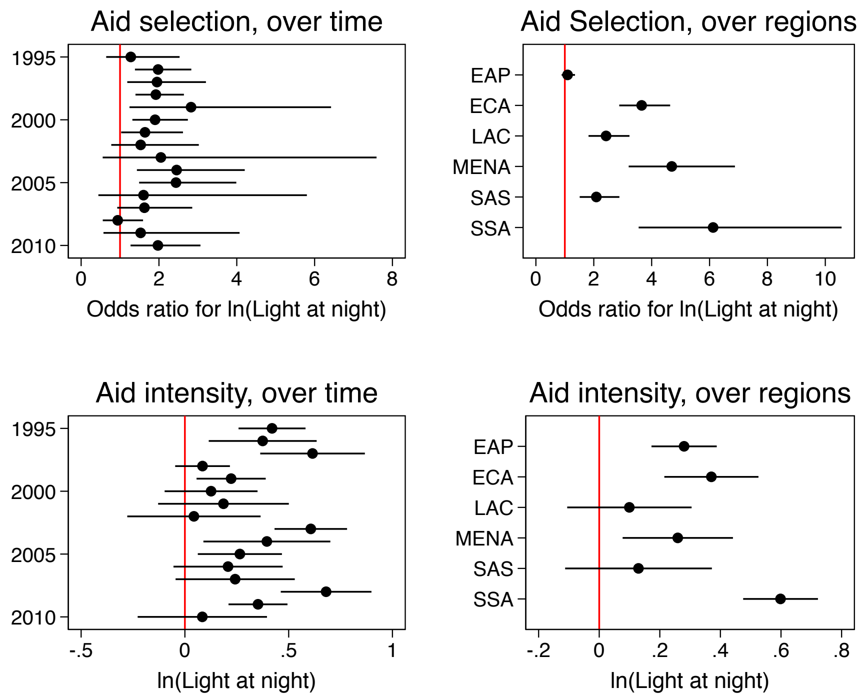

Figure 8: Heterogeneity in poverty targeting over years and regions, GPW population data. EAP = East Asia & Pacific, ECA = Europe & Central Asia, LAC = Latin America & Caribbean, MENA = Middle East & North Africa, SAS = South Asia, SSA = Sub-Saharan Africa

Table 10 replicates Table 1 but instead of limiting the dataset to projects geocoded to within 25 km of the true location it allows projects that were geocoded to the centroids of second-level (ADM2) regions. This is not a minor adjustment, as 26% of all of the rows in the dataset are geocoded at this level of precision. With this expansion of the dataset, the analysis includes 73% of the WB dataset.

It should be noted that adding projects that were only geocoded to a regional centroid very likely introduces bias into the $\ln(\text{light})_{t-1}$ coefficient. This is because centroids will often be darker, but rarely brighter, than the true aid location. Intuitively, this is because populated places are likely to be brighter than centroids, which may often have no human settlements at all. If we code projecst to centroids and take the light value there, then it will appear that aid is targeting many places with no light at all, when in fact it may often be targeting populated places within the region. Regardless, the results are similar in terms of sign and significance. The magnitude of pro-rich aid targeting is somewhat lower, and one way to read these results is that the show something like the lower plausible bound of the result.

Table 10: Table 1, Precision code < 4

	(1)	(2)	(3)	(4)
$\ln(\text{light})_{t-1}$	1.232***	1.205**	0.296***	0.288***
	(0.091)	(0.090)	(0.042)	(0.043)
$\ln(\text{population})_t$	1.981***	1.968***	0.090***	0.088***
	(0.061)	(0.057)	(0.015)	(0.015)
<100 km to capital$_t$		1.342***		0.063*
		(0.082)		(0.037)
Dependent variable	Binary	Binary	ln(aid cost)	ln(aid cost)
Estimation	Logit	Logit	OLS	OLS
Country-year FEs	Yes	Yes	Yes	Yes
n cell-years	305,508	305,508	11,557	11,557
n country-years	658	658	681	681

Models 1 and 2 show odds ratios and have bootstrap standard errors based on 1000 replications and 658 country-year clusters in parentheses. Models 3 and 4 show standard errors clustered on country-years in parentheses. *** p<0.01, ** p<0.05

6 Appendix B: Additional conjoint information

This appendix provides additional information about conjoint analysis. First, I show how the weights adjust the sample so that it closely matches the WB's portfolio.

Table 11 shows the results for regions and Table 12 shows the results for sectors. Numbers may not perfectly sum due to rounding. East Asia and Pacific region was under-represented in the sample and Europe and Central Asia was over-represented. The Public Administration sector was over-represented in the sample. Balance is very good after weighting.

Table 11: Regional Balance

Region	Population	Sample	Sample-Pop	Weighted	Weighted-Pop
SSA	0.37	0.4	0.03	0.37	0
EAP	0.16	0.1	-0.07	0.16	0
ECA	0.13	0.22	0.09	0.13	0
ALC	0.14	0.12	-0.02	0.14	0
MENA	0.08	0.07	-0.01	0.08	0
SSA	0.12	0.1	-0.03	0.12	0

Table 12: Sectoral Balance

Sector	Population	Sample	Sample-Pop	Weighted	Weighted-Pop
Agriculture	0.15	0.17	0.01	0.16	0
Education	0.09	0.1	0.01	0.09	0
Energy & Extractives	0.12	0.09	-0.03	0.12	0
Financial Sector	0.05	0.09	0.03	0.05	0
Health	0.06	0.06	0	0.06	0
Industry, Trade, Services	0.07	0.04	-0.03	0.07	0
ICT	0.04	0.01	-0.03	0.03	-0.01
Public Administration	0.14	0.19	0.05	0.15	0
Social Protection	0.08	0.11	0.03	0.08	0
Transportation	0.09	0.08	-0.01	0.09	0
Water & Sanitation	0.1	0.07	-0.03	0.11	0

While weighting can introduce a 'researcher degree of freedom,' that concern dose not exist in this paper because the weighting code was pre-registered before data collection. However, I still present Figures 5 and 6 without the use of survey weights. The results are shown in Figures 9 and 10. The results without the weights are very similar to those reported in the main text. These unweighted figures can be directly compared to Figure 7, which also does not use survey weights.

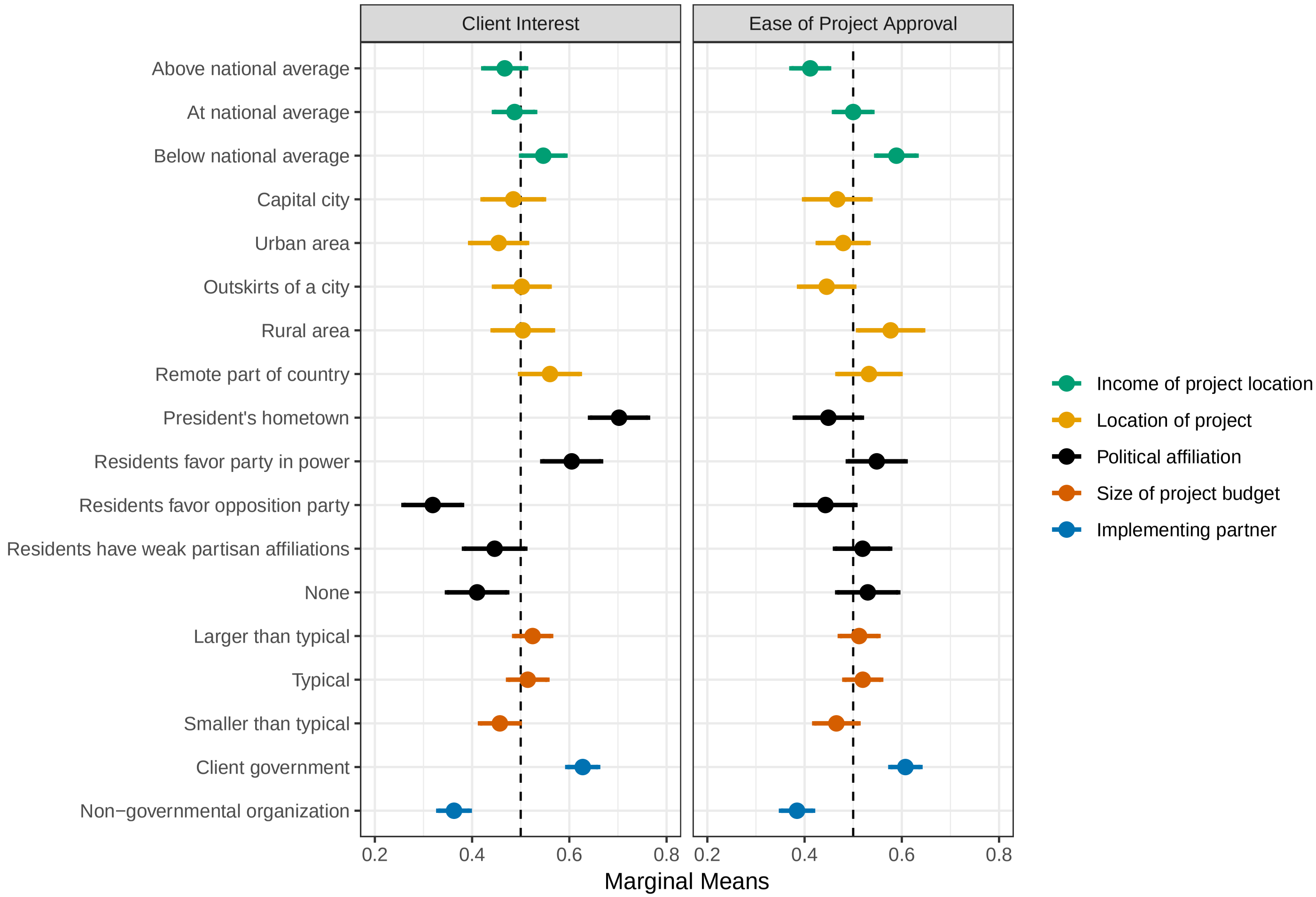

Figure 9: Effect of project attributes on probability of selection.

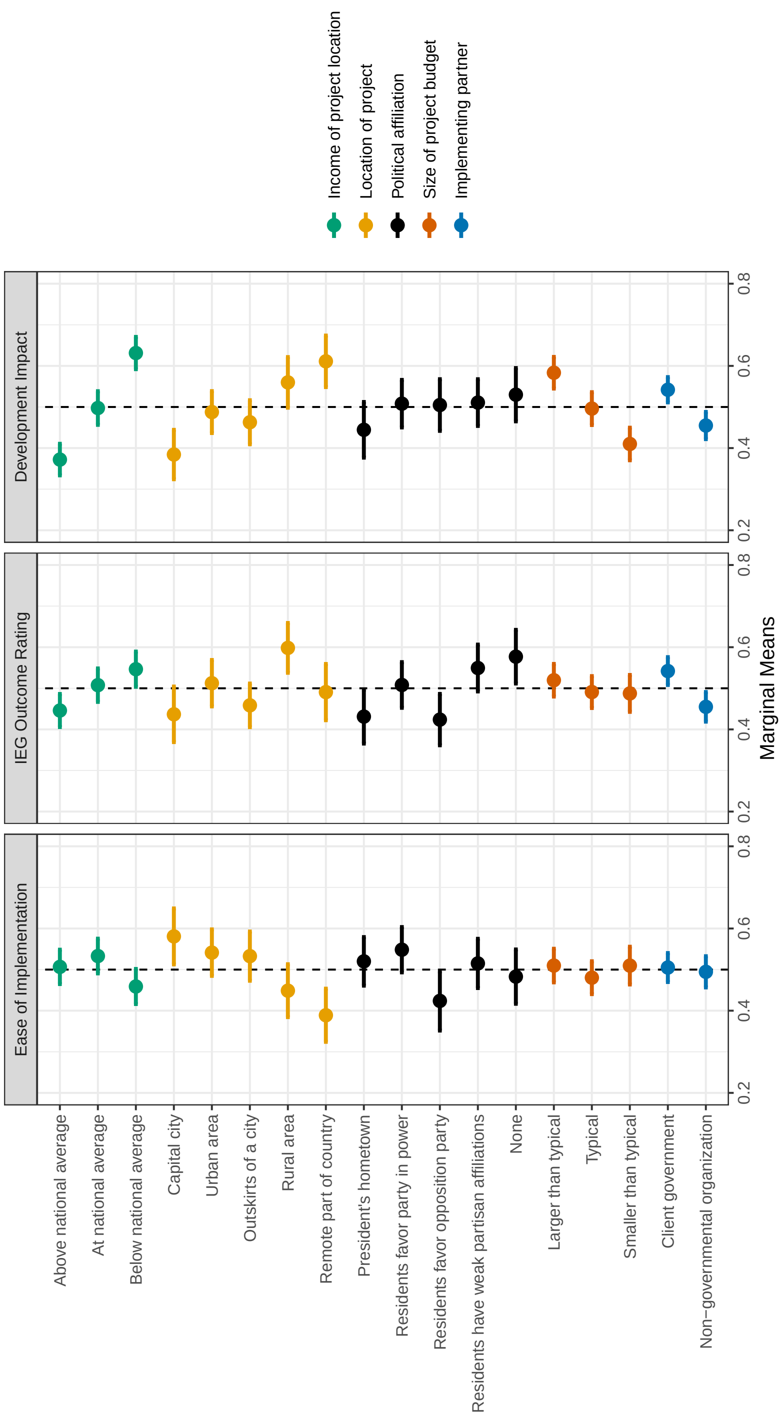

Figure 10: Effect of project attributes on probability of selection.

7 Appendix C: Deviations from pre-registered code

All survey-related code—from data cleaning to analysis to graphical presentations—was pre-registered with EGAP before data collection (20190424AB).

The present version of the analysis includes a small number of changes to the pre-registered code. All but two of these changes involve the graphical presentation of the results.

The only substantive change is that when weighting the TTLs against WB projects using regions, I dropped the "other" region. Projects with this region make up less than 0.5% of all projects, and no TTL in the survey selected this region. Technically this was not a change to the code as the list was an external .csv file, but it was a change to the analysis plan that occurred after I saw the data.

All other changes to the analysis code were cosmetic and involve modifications to graphs:

1. I adjusted the legend of the correlation plot so that it better matched the limits of the data (change to line 31 of the pre-registered code),

2. I re-ordered the graphs across the two plots. This involved changing line 122 and 134 (in the pre-registered code) and then adding code to re-order the graphs within each plot,

3. I fixed a typo on line 104 (in the pre-registered code) that prevented the levels of one of the independent variables from being re-ordered properly

4. I made the points and lines thicker in plots Figures 5 and 6. This involved adding "cex=3" to line 125 and 137 and adding "size=1" to 126 and 138 of the pre-registered code.

5. I ensured the range of the x-axis is the same in Figures 5 and 6. This involved adding "xlim(.2, .8) + " to a new line in each of the two plots.

None of these changes have any effect on the results of the analysis, and all were done purely for clarity in presentation.

Finally, the last change was to one function in my data cleaning code ("TTL Data Prep.r"). This code made use of the group_by_() tidyverse function, which was deprecated shortly after pre-registering. I re-wrote the function using the approved curly-curly operator. This involved changes to the "make_cum_per()" function. This was a minor change and, as one would expect, the function produces the same outputs from the same inputs.