The Review of International Organizations
https://doi.org/10.1007/s11558-019-09365-x

COMMENT

Results from single-donor analyses of project aid success seem to generalize pretty well across donors

Ryan C. Briggs[1]

Published online: 03 September 2019

Abstract

Much research on foreign aid presents claims that apply to aid in general but tests these claims using data from one or a small number of donors. This makes it difficult to know if we have learned something about aid, or merely something about one donor. For example, the literature on project aid success has found that per capita GDP growth rates or Freedom House scores in recipient countries correlate with project success. However, these claims have been tested against data from only multilateral donors and often against data from only the World Bank. I re-examine these analyses using a dataset of harmonized project outcome scores for seven diverse donors. Most donors seem to be similarly influenced by recipient-level and project-level factors, though a few notable exceptions exist. Analyses of project aid success that focus on single donors may be able to produce knowledge about aid in general.

Keywords Foreign aid

JEL Classification F35 · O22

Many studies of foreign aid test general claims with narrow evidence. This is likely a function of limited data availability and is most acute when studies need data that

I'd like to thank Aart Kraay for comments on an earlier draft and Dan Honig for sharing data and for comments.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11558-019-09365-x) contains supplementary material, which is available to authorized users.

✉ Ryan C. Briggs
rbriggs@uoguelph.ca

[1] Guelph Institute of Development Studies and Department of Political Science, University of Guelph, Guelph, Canada

are more fine-grained than country-years flows.[1] For example, most donors do not provide project-level information about their foreign aid. One notable exception to this is the World Bank (WB), which provides project-level information for its entire set of aid projects as well as outcome ratings and locations for many projects. Unsurprisingly, scholars have responded by testing their claims about aid using WB data. While this is understandable, it is unclear if the findings produced by such analyses generalize to aid from other donors. This comment seeks to understand if past research on aid project success generalizes to donors beyond the WB.

It is reasonable to expect especially multilateral aid to be different from bilateral aid. Multilateral donors, for example, have institutional structures that encourage their aid to flow to areas where stakeholder interests overlap (Rodrik 1995). This arrangement seems to make their aid less (or perhaps differently) politicized than bilateral aid, which often closely tracks the political, economic, or military interests of donors.[2]

While there are good reasons to expect aid from multilateral donors to behave differently from bilateral aid, limited evidence suggests a high degree of similarity across donors in some domains. For example, Jablonski (2014) and Briggs (2014) both examine ethnic favoritism in aid targeting in Kenya. The former paper uses WB and African Development Bank data for all sectors and the latter uses data from all donors but is limited to projects focusing on health and roads. Both produce very similar findings of ethnic favoritism.

Donors also seem to target aid to richer places within recipient countries. This relationship exists in analyses of aid from the WB (Öhler et al. 2019; Custer et al. 2017) and the WB and African Development Bank (Öhler and Nunnenkamp 2014; Briggs 2017; 2018b). It was also shown to exist in a pooled analysis of aid from between 28–79 donors to Nigeria, Senegal, and Uganda (Briggs 2018a).

Findings of cross-donor similarity in spatial aid targeting are encouraging from the point of view of knowledge production. Our single-donor studies seem to have produced results that generalize to other donors. However, such homogeneity in effects cannot be taken for granted. China, for example, targets aid to presidential hometowns while the WB does not (Dreher et al. 2019). Chinese development flows also seems to lead to an increase in the perception of corruption, an effect which does not exist for WB aid (Brazys et al. 2017; Isaksson and Kotsadam 2018). For many other questions, we lack cross-donor results and so simply do not know if our single-donor results generalize. For example, (only) WB data has been used to examine the relationship between the degree of aid targeting and aid capture (Winters 2014), the determinants of the length of project preparation periods (Kilby 2013), or the effect of the length of preparation periods on project outcomes (Kilby 2015; Dollar and Svensson 2000).

A recent discussion of development finance research noted that a "a logical next step for future studies" was to make use of "the growing availability of project-level

[1] If country-year data suffice then the OECD DAC database has reasonable coverage for their members for many aid flows.

[2] As noted by Neumayer (2003, p. 102) in a discussion of bilateral aid, "there is little doubt that economic, political, and sometimes military-strategic interests of donors play a significant and sometimes dominating role for practically all donors." For further discussion, see Briggs (2017, p. 189).

outcome data from both bilateral and multilateral development finance institutions" to test "whether, to what extent, how, and why [...] project-level factors that enable and constrain the successful implementation of development projects differ between bilateral and multilateral development finance institutions" (Nielson et al. 2017, p. 163). The present comment takes a step in this direction by testing if past research on the determinants of project outcome ratings at the World Bank generalize across six other donors.

Past work found a positive relationship between an aid recipient's Freedom House score and their WB project ratings (Isham et al. 1997; Denizer et al. 2013). It has also found a negative relationship between Freedom House scores and Asian Development Bank (AsianDB) project ratings (Bulman et al. 2017), although in many cases the statistical significance of the Freedom House scores is marginal. Denizer et al. (2013) show that WB aid projects get better outcome ratings when recipient countries are experiencing faster per capita GDP growth. They also find that projects that take longer to complete receive lower outcome scores. Bulman et al. (2017) find the same things when examining AsianDB and WB aid, thus showing these results generalize at least a little. A notable exception to the narrow focus of past work on project success is Honig (2018), which examines how different forms of control influence project success and which introduced the cross-donor data on which this comment is based.

The following analyses examine the generalizability of the above results in four ways. First, I examine the extent to which different donors agree about which recipient-time periods have more or less successful aid projects. I find much agreement across donors. For example, China in 1995–2005 stands out as a place where many aid projects were successful, while Zimbabwe in the same time period was uniquely difficult for donors (and pretty much everyone else). While this does not tell us why aid projects get better ratings in different times and places, it shows that all donors seem similarly affected by recipient country context.

Second, I examine two recipient-level factors that might explain why donors find some times and places easier for implementation. Drawing on past research, I show that project success scores from all donors respond similarly to a recipient's per capita GDP growth rate and Freedom House score. While I do not examine the causal effect of either variable, I present results with no control variables and with a set of recipient country, project start year, and aid sector fixed effects.[3] Regardless of the addition of the fixed effects, aid projects from all donors receive better outcome scores when the recipient country experiences faster growth and the magnitudes of this effect are similar across donors. This result generalizes. The Freedom House scores are more problematic. With the full fixed effects, I find that aid projects are insensitive to changes in Freedom House scores. However, this null effect is again roughly similar across all donors. The analysis without controls reproduces some of the patterns reported in past work, but it seems that these patterns are likely due to country-specific, time-invariant omitted variables rather than the actual effect of operating in a freer country.

[3]The sector fixed effects use the OECD-DAC's 5-digit purpose classification codes. This means that sectors specify something precise like 'Tuberculosis control' rather than the more aggregated 'Basic health.' The results are very similar if I use the cruder 3-digit codes for the fixed effects.

Third, for each donor I examine how much of the variation in their project outcome scores can be explained by recipient-level factors. Past work focusing on the WB (Denizer et al. 2013) and WB and AsianDB (Bulman et al. 2017) found that only about 10-20% of the variation in project outcomes exists between recipient countries, implying that most of the variation in project success takes place *within* countries. With some caveats due to data limitations, I replicate this finding in my broader dataset.

Fourth and finally, I examine if the relationship between project duration and project outcomes is similar across donors. As in my second set of tests, I do this with no controls and then with a full set of fixed effects. I find that four of the seven donors have the expected negative relationship between project duration and outcome scores. The two donors with the smallest samples sizes have insignificant results, but this may say more about the data than the donors. Finally, I find that one donor has a robust, large, and positive relationship between project duration and project outcomes. This is an unexpected result.

It seems that we can learn something about the factors that influence aid success in general from single-donor studies. Donors generally seem to find the same times and places easy or hard to implement projects. All donors have a roughly similar relationship between a recipient's per capita GDP growth and project outcomes. All donors also report no relationship between Freedom House scores and project outcomes. All donors fit a general pattern where factors that vary across recipient countries explain relatively little of the variation in their project's outcomes. Finally, most donors find that longer projects receive lower outcome ratings. I conclude by outlining a tentative framework that explains why we see cross-donor similarity in this domain. The framework offers a way to think about generating predictions for where we might see more or less similarity in donor behavior or outcomes in future research.

1 Data

Data for project outcomes comes from the Project Performance Database (PPD) introduced in Honig (2018). The database includes information from seven donors on project outcomes, the recipient country, start dates, and end dates.[4] The database includes information on the AsianDB, the Global Fund to Fight AIDS, Tuberculosis and Malaria (GFATM), the German Society for International Cooperation (GiZ), the International Fund for Agricultural Development (IFAD), the Japan International Cooperation Agency (JICA), the German Development Bank (KfW), and the WB. The database thus includes bilateral donors, multilateral donors with global coverage, a regional multilateral donor, and a fairly new public-private partnership targeting specific diseases, among others.

While the database does not include the United States or China, the two bilateral donors whose giving is most plausibly influenced by grand strategy, it includes two bilateral donors who vary on some key attributes. Germany is one of the more

[4]The dataset includes 8 donors, but the UK's DFID lacks information on start and end dates and so I ignore it here. The first year in the database in which a project was started is 1956 and the last year is 2012, but the data are densest between 1975–2005.

poverty-sensitive donors, comparable to Sweden and the multilateral donors over some time periods, while "Japanese aid has been strongly biased against the poorest countries since the early 1990s" (Nunnenkamp and Thiele 2006, p. 1183).][5] Unlike Germany, Japan's aid also consistently tracks its exports (Nunnenkamp and Thiele 2006; Canavire et al. 2006). Japan had a somewhat larger aid budget than Germany from 1990 to 2007, but German aid has exceeded Japanese aid since then (OECD DAC 2019). Japan and Germany are similar in the role of their government in their economy and the degree to which their aid bypasses recipient governments (Dietrich 2016). They are also similar in the extent to which their agencies have organizational autonomy (Honig 2018, 2019).

Notably, the PPD includes a consistent six-point project outcome score based on donor-reported outcome data and harmonized in Honig (2018). This outcome score represents an "overall holistic success rating," is based on a shared OECD standard, and is "intended to incorporate a project's relevance, effectiveness, efficiency, sustainability, and impact" (Honig 2018, p. 59–60).[6] Despite best efforts at harmonization, it surely remains the case that donor-specific factors mean that one cannot use these scores to examine which donor has more successful aid. However, these scores allow one to make systematic comparisons of project success within the same donor. The scores also permit cross-donor comparisons after being normalized, for example after being de-meaned.[7] The combination of a consistent outcome score and coverage across seven diverse donors makes this database an excellent resource for testing the generalizability of past research on the correlates of project aid success.

After dropping projects that lack data on start year, end year, or outcome rating, the dataset includes 11,722 projects. However, as is shown in Fig. 1, coverage across donors is uneven. About 60% of all projects come from the WB. All other donors have under 2000 projects and the IFAD has the smallest number of projects at 28.

I add to this dataset the growth rate of GDP per capita and Freedom House scores, both drawn from the International Political Economy Data Resource (Graham and Tucker 2019). I do this by assigning to each project the average of each variable taken over the lifetime of each aid project. The Freedom House score for both civil and political liberties runs from 1 to 7. Following past work, I reverse the score so that higher scores mean more freedom and then I average the two scores together.[8]

[5]Japan's lack of poverty sensitivity is partially due to the fact that "Japanese aid is concentrated on (relatively advanced) Asian neighbours" (Nunnenkamp and Thiele 2006, p. 1191).

[6]One may worry that project outcome ratings are biased by, for example, political factors. Kilby and Michaelowa (2016) examine WB project outcome ratings for bias and find that most geopolitical variables do not seem to cause bias in IEG ratings. Only non-permanent membership in the UN Security Council seems to influence IEG ratings.

[7]For more information, see Honig (2018, p. 58–68) and Honig (2019). Section two of Bulman et al. (2017) also has a useful discussion of the potential challenges when comparing outcome scores across different donors. Every time that I make cross-donor comparisons I normalize the data by donor (through the addition of donor fixed effects, for example).

[8]Past work also used (private) WB CPIA scores as a measure of the institutional environment of a recipient country (Denizer et al. 2013; Bulman et al. 2017). Publicly available data on CPIA scores runs from 2005 to 2016, and only 7% of the projects in the PPD start on or after 2005. For this reason, I do not examine the relationship between CPIA scores and project outcomes.

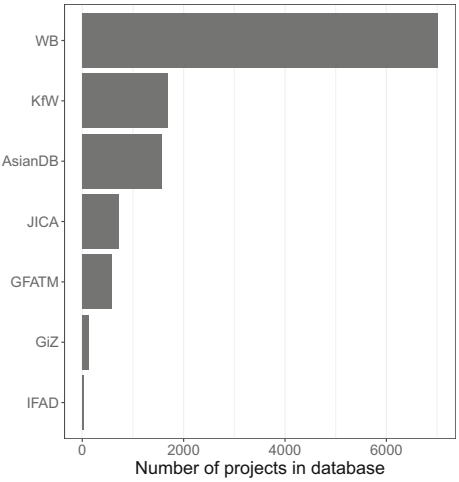

Fig. 1 Projects in PPD

2 Analysis

2.1 Do donors find the same places difficult?

To assess whether or not studies from one donor generalize, it is first useful to understand if aid projects from different donors are more likely to succeed in the same times and countries. The only existing evidence on this question comes from Bulman et al. (2017), who document a moderate correlation in the cross-national success rates of projects by the AsianDB and the WB. We thus have little sense of the degree of agreement across donors over which recipient-time periods produce better or worse project outcomes. I examine the correlation between project success in different recipients and time periods across five donors with sufficient data.[9]

To examine if donors agree on which recipients have more successful projects, I first create a set of recipient-periods. Each period is one decade long and the decades start in 1955 and run until 2015.[10] I then calculate the mean outcome rating for all projects started in each recipient-period for each donor.

Calculating mean project success scores per donor within recipient-decades allows for a balance between two competing pressures. First, one would like to compare across fine grained temporal units because some of the recipient-level features that influence project success are likely to vary over time. However, one needs to balance this concern against the fact that one needs information on at least two donors per

[9]I drop GiZ and IFAD from this analysis as both donors have very few projects and so I expect them to produce very unreliable results.

[10]The dataset runs from 1956 to 2012, so starting on years that end in five creates a better spread of projects over decadal periods than starting on years that end in zero.

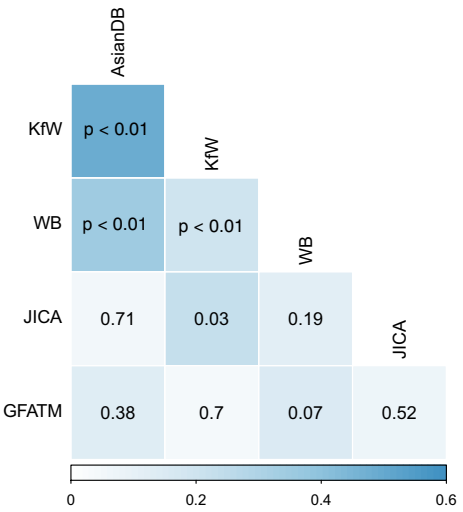

Fig. 2 Cross-donor correlations between project success per recipient-period (numbers in cells are p values testing zero correlation)

recipient-period in order to make a comparison. Only the WB has more than 2000 projects in the database, and so one quickly runs into the curse of dimensionality when trying to make cross-donor comparisons across 176 recipients and 55 years.

To balance these concerns, I make comparisons using ten year periods.[11] At this length, recipient characteristics can change over six periods but each period is long enough that I have a moderate chance of having recipient-periods where aid projects from multiple donors were started. In practice, when using periods of this length each pair of donors has between 41 (GFATM–AsianDB) and 230 (WB–KfW) shared recipient-periods.[12]

I then calculate the Pearson correlation coefficients between mean project outcome scores for each pair of donors.[13] Figure 2 presents the results. The color of each cell shows the correlation between the two donors at that intersection. The number in each cell is a p-value testing the null hypothesis that the correlation between mean project scores per recipient-decade for those two donors is zero.[14]

While the correlations between the pairs of donors could run from -1 to 1, in this instance all correlations are positive. When a recipient-period has aid projects from one donor that are more successful than that donor's average, projects from other

[11] The results are quite similar if I use five year periods instead of ten year periods.

[12] I report scatter plots of some of the correlations in the online appendix.

[13] Note that in calculating the Pearson correlation coefficients the data are demeaned by donor. Thus, I am never making a comparison between absolute levels of projects success scores across donors. This is important as different donors very likely have different baseline levels of project success ratings.

[14] The p-values are shown to two significant digits.

donors are also more successful than their average. However, the strength and statistical significance of the positive correlations varies across pairs of donors. The strongest correlations are between the WB, the AsianDB, and KfW. Not coincidentally, these are also the three donors with the most projects in the database. All three of these correlations are significant at $p < 0.01$ and the correlation coefficients run from a high of 0.49 to a low of 0.20. Many of the remaining correlations between donor-pairs are not statistically significant but all are positive.

The country-period with complete data across the five donors and the worst cross-donor average score is East Timor (1995–2005) with 3.7. If we expand our search and look for the worst average score allowing for missing data on one of the donors, then the clear 'winner' is Zimbabwe (1995–2005) with a cross-donor mean score of 2.3. It is followed by Pakistan (1975–1985), Sri Lanka (1985–1995), Kazakhstan (1985–1995), and then India (1975–1985).

The country-period that received aid from all five donors and has the best score is China (1995–2005) with 5.2. It is followed by Vietnam, Thailand, Cambodia, and India, all over the same time period. Allowing for missing data on one donor, the best country-period is El Salvador (1995–2005) with a score of 5.6. It is followed by Romania and Indonesia in the same time period, and then Vietnam (2005–2015). While impressionistic, these best and worst country-periods align well with commonsense understandings of when aid is more or less likely to work well.

2.2 Recipient-level factors

The correlation analysis above tells us that donors tend to find that the same places enable or disable project success, but they do not test explanations as to why this may be. This section offers such tests. I examine across-donor variation in the relationship between project outcome ratings and two recipient-level variables: GDP per capita growth and Freedom House scores. I first examine the growth rate of GDP per capita.

The growth rate of GDP analysis is made up of two regressions. The first is described in Eq. 1, where i indexes the aid projects and d indexes the seven donors. The regression explains project outcome ratings as a function of mean GDP per capita growth over the life of the project, a set of donor dummies, and the interaction between the set of donor dummies and mean GDP per capita growth. The second regression adds to Eq. 1 a set of recipient country, sector, and project start year fixed effects. Both regressions cluster standard errors on recipient countries.[15]

$$Outcome_i = \alpha + \beta GDPpcGrowth_i + \sum_{d=AsianDB}^{WB} \theta_d Donor_d + \gamma_d \left(\sum_{d=AsianDB}^{WB} Donor_d \times GDPpcGrowth_i \right) + \mu_i \quad (1)$$

[15] Clustering on periods and sectors generally produced similar standard errors. Clustering on (the small number of) donors produced smaller standard errors. Thus, clustering on recipients seems conservative.

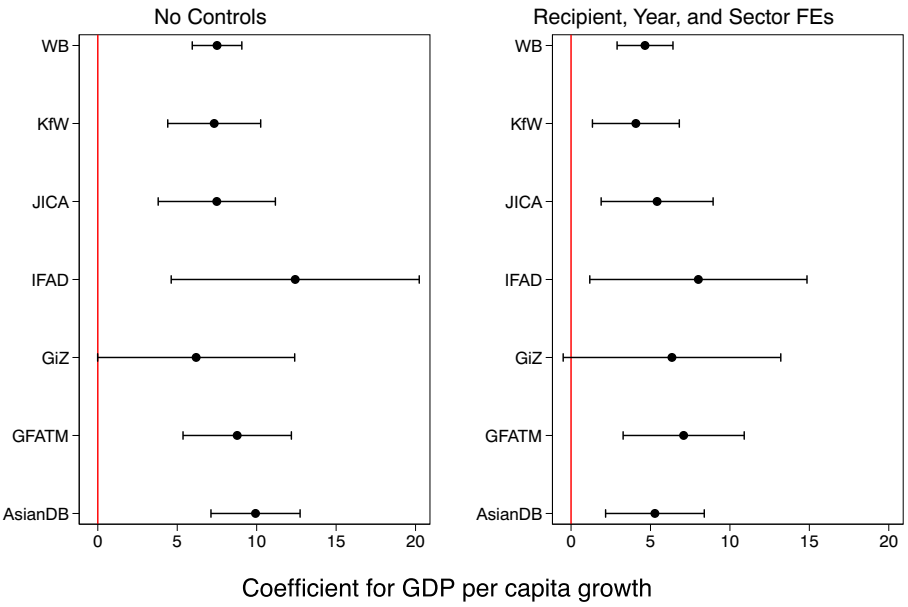

Fig. 3 Relationship between GDP per capita growth and project success

I present the results graphically in Fig. 3, where the effects for each donor are built up from the growth of GDP per capita base term (β) and the interaction terms (each donor's γ).[16]

Past work had found a statistically significant and substantively meaningful relationship between GDP per capita growth and project outcomes for both WB and AsianDB aid. I closely replicate these results. The coefficient for the WB in the left pane of Fig. 3 is about 7.5, which is very close to the 8.1 found by Denizer et al. (2013).[17] Given that the analysis in the two papers is not identical, I consider this a very close replication of the core claim that WB projects get better ratings when recipients are experiencing higher growth.

More notably, aid projects in general work better when the recipient country has higher economic growth. This effect is present—and is roughly similar in size—across the AsianDB, GFATM, IFAD, JICA, KfW, and WB. The point estimate for the only remaining donor, GiZ, is similar to those of the other donors but it has much larger standard errors and its 95% confidence interval crosses zero. It should be recalled that GiZ and IFAD have very small sample sizes and that throughout this paper their results should be viewed skeptically. The results hold regardless of the addition of the set of fixed effects.

[16]More precisely, the AsianDB is the base term in the interaction, so its estimate in Fig. 3 is β. The interaction terms (each donor's γ) then are added to β to show effect for each donor.

[17]Their coefficient comes from Table 3, Panel B, Model 6. They add a series of sector fixed effects and some related interactions, but they notably do not add country fixed effects as I do in the right pane of Fig. 3.

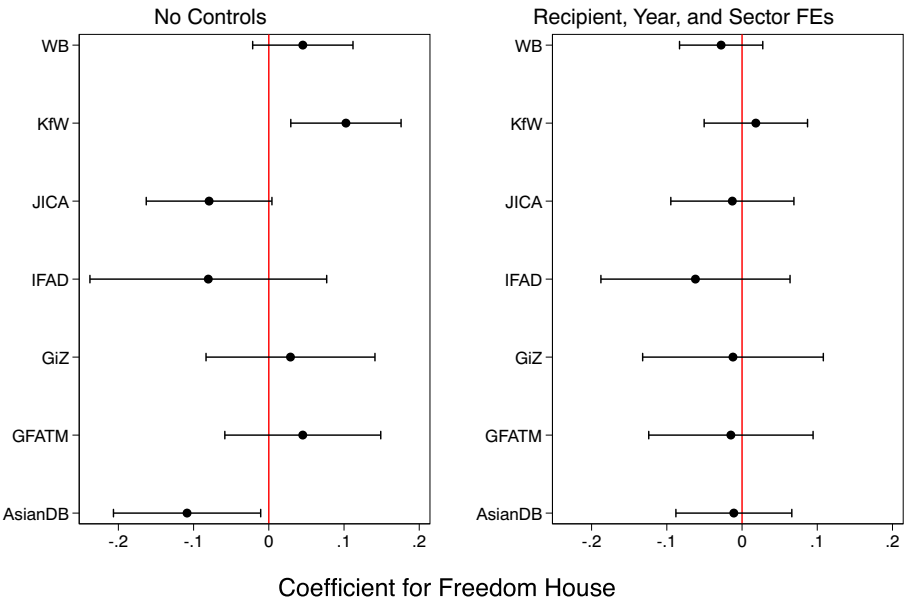

Fig. 4 Relationship between Freedom House scores and project success

I next turn to the relationship between Freedom House scores and project outcomes. The regression and presentation style follow the prior analysis. The left panel of Fig. 4 shows results from a minimal regression with simply the Freedom House score, a set of donor dummies, and an interaction between the two. The right panel adds to this the full set of fixed effects.

The analysis with no controls shows some of the patterns noticed in past work, such as a positive point estimate for the WB and a negative one for the AsianDB. However, the addition of the fixed effects removes all of these patterns and produces near-zero point estimates for all donors. This suggests that any correlation between Freedom House scores and project outcomes is due to unobserved country-specific factors.[18] While this finding differs from past work, the null effect of Freedom House scores (especially in the right panel) is quite consistent across donors. Nobody's aid does better in places with higher Freedom House scores, once we net out time-invariant recipient country-specific factors.

2.3 Bounding the influence of recipient-level factors

The motivation for the above analysis is the finding that donors agree on which places and times produce more successful projects. However, the above results do not tell us how important recipient-level factors are in explaining project success in general. In fact, past work looking at the WB (Denizer et al. 2013) and WB and AsianDB

[18] If I add the fixed effects sequentially, the point estimates all move toward zero when the recipient country fixed effects are added. The year and sector fixed effects do very little in this case.

(Bulman et al. 2017) suggests that recipient factors may not actually explain that much of the variation in project success.

Following Denizer et al. (2013) and Bulman et al. (2017), I examine how much of the variation in project-level outcomes can be attributed to recipient-level factors by regressing the six-point outcome score on a set of recipient dummies for each donor for each year in which the donor had active projects. Each year includes all active projects by a donor, and I consider a project to be active in a year if the year is on or after the project start date while also being on or before the project end date. I exclude from the analysis any year where a donor had fewer than 10 active projects. Iterating this way over cross-sections of project data by donor and year leads to 197 regressions. I am interested in the degree to which variation in project-level outcomes can be attributed to recipient factors so, following the approach in Denizer et al. (2013) and Bulman et al. (2017), I record the R-squared for each regression.

When using this method, Denizer et al. (2013) found that WB projects in the period 1995–2005 had a mean R-squared of 0.20. When using the same method and looking at projects by the AsianDB and WB, Bulman et al. (2017, p. 335) found that "only 10–25% of the variation in project outcomes is between countries."

Aside from recording the R-squared value for each regression, I also record for each donor-year the total number of the donor's projects active in the year and the number of recipients that had at least one active project from the donor. This is important when comparing the R-squared values across donors, as the R-squared will be inflated when each recipient receives few projects.[19] This issue complicates across-donor comparisons, as a high R-squared for one donor-year regression may imply that the donor had numerous projects within each recipient and that projects within the same recipient had similar outcomes. However, it could also simply imply the donor had few projects per recipient in that year. This statistical issue can be resolved by comparing the R-squared values resulting from each regression against the number of projects per recipient in each year for each donor.

Figure 5 depicts the results of the analysis. Each dot represents one of 197 regressions. The Figure is color-coded by donor and includes lowess fit lines by donor. All of the donors under study appear to be following the same trend and data sparsity explains every case where the recipient dummies explain most of the variation in project outcomes. For example, the mean R-squared value for GiZ is 0.86, but it also has very few projects per recipient. The WB or KfW also have similarly inflated R-squared values when we look only at the years when they have a similarly low number of projects per recipient.

As the number of projects per recipient increases, the R-squared decreases and the relationship is non-linear. The overall trend in R-squared values more-or-less levels out around 12 active projects per recipient, so it is probably correct to say that only about 15% of project-level outcome variation can be attributed to recipient-country

[19] In the limiting case, if each recipient receives one project then the recipient dummies will explain 100% of the variation in project outcomes. This issue is noted in footnote 5 of Bulman et al. (2017) and they respond to it by dropping some of their data.

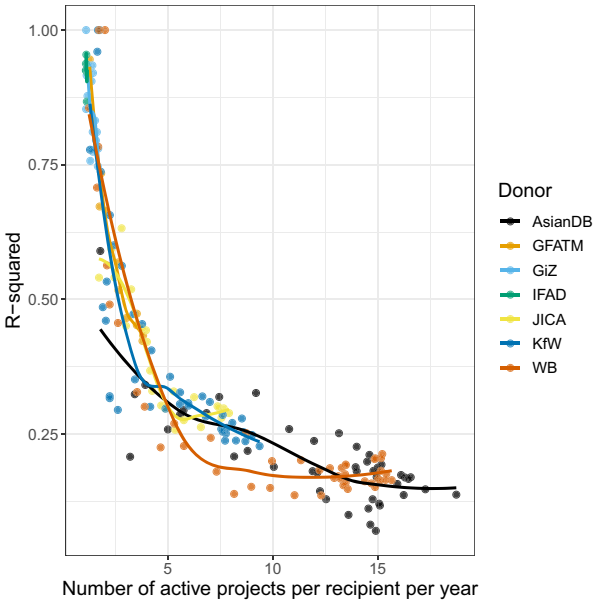

Fig. 5 The declining influence of recipient dummies on project outcomes

factors. Thus, the present analysis affirms past work showing that most of the variation in project outcomes is due to factors that vary across projects within recipient countries (Denizer et al. 2013; Bulman et al. 2017).

While Fig. 5 suggests that all donors are part of the same trend, it should be noted that I cannot offer a strong test of the presence of this trend because few donors ever have more than 10 active projects per recipient per year. Thus, the results from this section rely on some visual extrapolation and should probably be viewed more skeptically than the results from the other three sections. Nonetheless, the most natural interpretation of Fig. 5 is that there is a general trend influencing all donors, and that recipient-level characteristics are never very important in practically explaining project outcomes.[20]

2.4 Project-level factors

In response to the finding that recipient-level factors do not actually explain that much of the variation in project success, past research moved to examining how project-level factors influence project success (Denizer et al. 2013; Bulman et al.

[20]It is important not to read this result in causal terms, as sample selection is a major issue. Given that a donor is active in a country, local conditions appear to not explain much of the variation in outcomes across its projects. Of course, donors may choose not to operate in places with particularly bad conditions because of the causal effect of those conditions on outcomes.

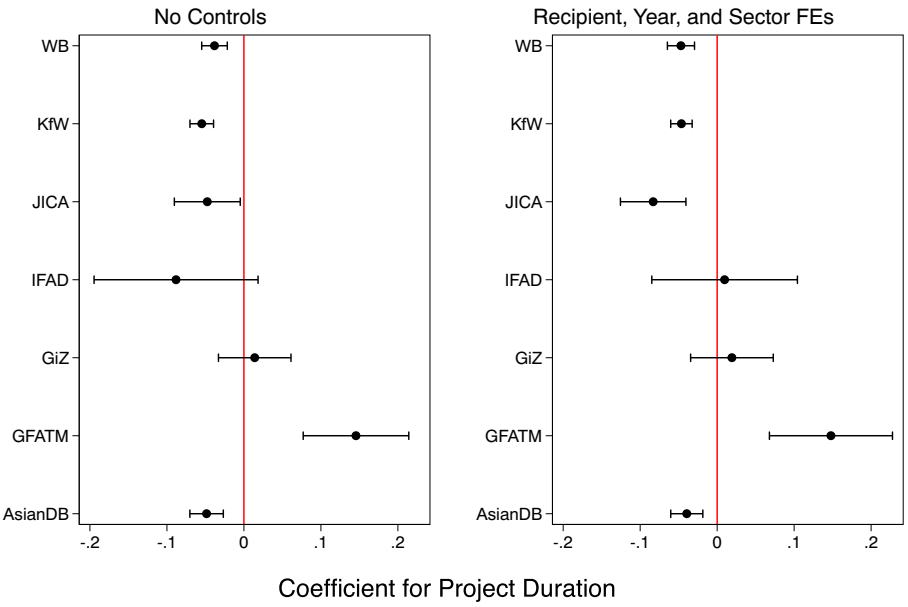

Fig. 6 Relationship between project duration and project success

2017). However, the present dataset lacks harmonized versions of most of the project-level variables tested in past work. One exception is project duration, which is coded for all donors. Thus, I am able to test the relationship between the duration of projects and project outcomes.

Following the analysis strategy in Section 2.2, I explain the outcome score first simply as a function of the duration of the project (in years), a set of donor dummies, and an interaction between the two. I then add to this recipient country, year, and sector fixed effects. Standard errors are clustered on recipient countries. The results are presented graphically in Fig. 6.

Longer projects by the AsianDB, JICA, KfW, and WB were less likely to be successful, a finding which is in line with past research (Denizer et al. 2013; Bulman et al. 2017). IFAD and GiZ had noisier and/or insignificant results, but as noted previously we should put less stock in results from these donors as the database under study has relatively few of their projects. However, the GFATM had a robustly positive relationship between project duration and project success. Thus, the negative relationship between project duration and project outcomes seems common but not universal.[21] Further cross-donor analysis will require the creation of additional harmonized variables.

[21] There does not appear to be much sector-level heterogeneity in the relationship between project duration and outcome scores. For evidence supporting this claim, see the appendix.

3 Discussion

The above results largely support the idea that the factors that shape project success are similar across donors. All donors found the same country-time periods relatively easy or hard, though the degree of agreement varies across pairs of donors. All donors had more successfully rated projects in recipient countries with faster growth and no donor's outcomes were correlated with Freedom House scores once recipient country fixed-effects were included in the analysis. All donors seem to be part of a general trend where recipient-level factors end up explaining only about 15% of the variation in project outcomes, though data limitations make this finding tentative. Finally, all but one of the donors with a large sample of projects in the database show a negative relationship between project duration and outcome ratings.

In sum, similar factors seem to drive project success across donors. This mirrors the finding that similar forces seem to shape cross-donor spatial aid targeting within countries (Öhler and Nunnenkamp 2014; Öhler et al. 2019; Custer et al. 2017; Briggs 2017, 2018a, b). However, these findings of similarity across donors contrast with much research on cross-national aid targeting, which has typically found sharp differences between bilateral aid and multilateral aid (Maizels and Nissanke 1984; Dollar and Levin 2006; Nunnenkamp and Thiele 2006). They also contrast with work finding different effects of Chinese and WB aid (Brazys et al. 2017; Isaksson and Kotsadam 2018).

Here I propose a tentative model for understanding the domains in which we might expect to see aid from different donors behave more or less similarly. In the most-similar cases, we should be able to learn a lot about aid 'in general' even if we base our research on data from a single donor. I differentiate between factors that influence aid before implementation and factors that affect the implementation process. If we are trying to understand aid targeting then we are asking a question about aid before implementation. If we are trying to understand factors that shape project outcomes or the degree to which the funding for an aid project is appropriated by the regime in power or implementing partner, then we are asking about the factors that affect the implementation process.[22]

I base my discussion of factors that influence aid before implementation on the idea that aid is a negotiated bargain between donors and recipients.[23] I expect that for each domain, each donor and each recipient have a set of possible outcomes that it finds acceptable. We should then see donors act more similarly the larger the set of choices that they find acceptable (and the more those choice sets overlap) and the smaller the set of choices that recipients find acceptable. To take a trivial example, imagine donors are willing to give loans or grants but a recipient will accept only grants. In this case, donors will be similar in the domain of 'aid concessionality.' To take a somewhat more realistic example, one could also imagine that recipients would

[22] This discussion brackets research on the effects of aid and instead focuses on aid as a dependent variable. Informally, there seems to be more heterogeneity in the effects of aid on various outcomes than there is on the effects of various factors on aid.

[23] For an extended discussion of how these bargains work themselves out in practice, see Swedlund (2017b).

prefer for aid to flow to the birthplace of their president but they will accept aid that does not. Some donors, perhaps China, are willing to allow their aid to be directed to the President's place of birth while others, such as the WB, are not. We would then expect to see divergence in this domain, which is in fact what we find (Dreher et al. 2019).

I expect that single donor studies of the factors influencing aid implementation will generalize better across donors the more that the domain of analysis is technical and the more that it is handled by low-level staff. In both cases, I expect similarity because in these areas there is likely to be more deference to standardized routines and "best practices". Many of the factors that influence aid implementation are at least treated as if they were technical in nature and are handled by low-level staff or implementing partners. I expect studies of these domains to generalize broadly across donors. For example, Kilby (2015) finds that longer preparatory periods for WB projects improves outcome ratings. I expect this result would generalize across most donors. On the other hand, the factors that lead to aid suspensions should be less likely to generalize across donors because the process for making such a decision is less dependent on standard operating procedures and is taken by higher-level staff with more discretion. Indeed, it appears that decisions about aid suspensions depend on strategic considerations of donor governments and institutional incentives within donor agencies, both of which can generate heterogeneity across donors (Swedlund 2017a).

It should be emphasized that this is a basic, initial framework and that it should be revised or redone as more data become available. The point of presenting such a tentative framework is to generate further discussion around when and where we might expect to be able to learn about aid in general from studies of single donors. Given the paucity of detailed cross-donor data, understanding when we should be more or less confident about the generality of single donor studies seems a useful endeavor. Thanks to Honig (2018) we now have harmonized outcome rating data. Further tests of generalization will likely be most limited by a lack of relevant project-level independent variables such as GPS locations, length of project preparation periods, or names of project leaders.[24]

4 Conclusion

This comment showed that much past research on aid project success generalizes across donors. In the case of the Freedom House variable, where I could not replicate past work, I still found a similar null result across donors. I also presented a tentative framework for understanding when we should expect results from single donor studies to generalize to other donors. It is not obvious that single-donor studies in new domains will generalize as well as research on spatial poverty targeting or project outcome ratings. However, the present results should probably shift people

[24]The names of people in TTL-equivalent roles could be use to test the degree to which variation in outcomes is associated with fixed features of team leaders (Denizer et al. 2013).

with skeptical priors a little in the direction of believing that we can learn something about aid in general from research on single donors.

References

Brazys, S., Elkink, J.A., Kelly, G. (2017). Bad neighbors? How co-located Chinese and World Bank development projects impact local corruption in Tanzania. *The Review of International Organizations*, *12*(2), 227–253.

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.

Bulman, D., Kolkma, W., Kraay, A. (2017). Good countries or good projects? Comparing macro and micro correlates of World Bank and Asian Development Bank project performance. *The Review of International Organizations*, *12*(3), 335–363.

Canavire, G., Nunnenkamp, P., Thiele, R., Triveno, L. (2006). Assessing the allocation of aid: Developmental concerns and the self-interest of donors. *The Indian Economic Journal*, *54*(1), 26–51.

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

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

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

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

Dollar, D., & Svensson, J. (2000). What explains the success or failure of structural adjustment programmes? *The Economic Journal*, *110*(466), 894–917.

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

Graham, B.A., & Tucker, J.R. (2019). The international political economy data resource. *The Review of International Organizations*, *14*(1), 149–161.

Honig, D. (2018). Navigation by judgment: Why and when top down management of foreign aid doesn't work. Oxford University Press.

Honig, D. (2019). When reporting undermines performance: The costs of politically constrained organizational autonomy in foreign aid implementation. *International Organization*, *73*(1), 171–201.

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

Isham, J., Kaufmann, D., Pritchett, L.H. (1997). Civil liberties, democracy, and the performance of government projects. *The World Bank Economic Review*, *11*(2), 219–242.

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

Kilby, C. (2013). The political economy of project preparation: An empirical analysis of World Bank projects. *Journal of Development Economics*, *105*, 211–225.

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

Kilby, C., & Michaelowa, K. (2016). What influences World Bank project evaluations? Working Paper.

Maizels, A., & Nissanke, M.K. (1984). Motivations for aid to developing countries. *World Development*, *12*(9), 879–900.

Neumayer, E. (2003). *The Pattern of Aid Giving: The Impact of Good Governance on Development Assistance*. London: Routledge.

Nielson, D.L., Parks, B., Tierney, M.J. (2017). International organizations and development finance: Introduction to the special issue. *The Review of International Organizations*, *12*(2), 157–169.

Nunnenkamp, P., & Thiele, R. (2006). Targeting aid to the needy and deserving: Nothing but promises? *The World Economy*, *29*(9), 1177–1201.

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

Rodrik, D. (1995). Why is there multilateral lending?. In Bruno, M., & Pleskovic, B. (Eds.) *Annual World Bank Conference on Development Economics* (pp. 167–193). Washington: International Monetary Fund.

Swedlund, H.J. (2017a). Can foreign aid donors credibly threaten to suspend aid? Evidence from a cross-national survey of donor officials. *Review of International Political Economy*, *24*(3), 454–496.

Swedlund, H.J. (2017b). The development dance: How donors and recipients negotiate the delivery of foreign aid. Cornell University Press.

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