Supplementary Information

Buntaine, Mark T., Bradley C. Parks, Benjamin P. Buch. "Aiming at the Wrong Targets: The Domestic Consequences of International Efforts to Build Institutions." *International Studies Quarterly*.

Table of Contents

A. The Difficulty of Form and Function Targets 2

B. Additional Tests and Model Checks 8

C. Coding Procedures for Targets 10

D. Robustness of Bandwidth Choices Near Eligibility Threshold 11

Appendix A. The Difficulty of Form and Function Targets

A key assumption in our theory is that targets that measure whether an institution or policy exists (e.g., a new law, task force, or training program) are easier to achieve and maintain than targets that require improvement in how well environmental institutions solve public problems (e.g. quality of forest management, reduction in pollution, or reduction in water waste). If this is correct, then the ability of governments that receive development financing and World Bank staff to act strategically regarding target selection is strong, since for countries that receive concessional financing the achievement and persistence of institutional targets determines future aid amounts.

To test this assumption, we collected additional data on whether the targets in our sample were achieved and if these achievements persisted after the project was completed. Baseline and completion values of the targets were often included in the post-project evaluations that were used to initially code the institutional targets. In cases where both baseline and completion data were available (around 85% of indicators), we computed an ordinal measure of the progress made towards the target during the project.

To monitor the durability of institutional development gains beyond the date of project closure, we compiled new measures of the targets from a comprehensive search of official documents, government websites, NGO reports, and news sources at the beginning of 2013. We used a variety of public news sources, such as LexisNexis and Google News. In addition, many government agencies issue annual reports on staffing and appropriations, allowing us to re-measure some targets. Government websites and NGO reports also offered valuable data. In all cases, we recorded the source and publication date of the re-measured target. In total we re-measured post-completion data for 347 of our in-sample targets, a success rate of around 42%.

An initial examination of the descriptive data suggests that form targets are easier to achieve and maintain than function targets (Figure A1). For targets with baseline data in the evaluation documents, 70% of function targets were achieved during project implementation, while 77% of action targets and 76% of form targets were achieved during project implementation. Differences in the difficulty of maintaining performance on targets following project completion are even clearer. Whereas only 59% of function targets were maintained following project completion, 95% of action targets and 97% of form targets were maintained following project completion.

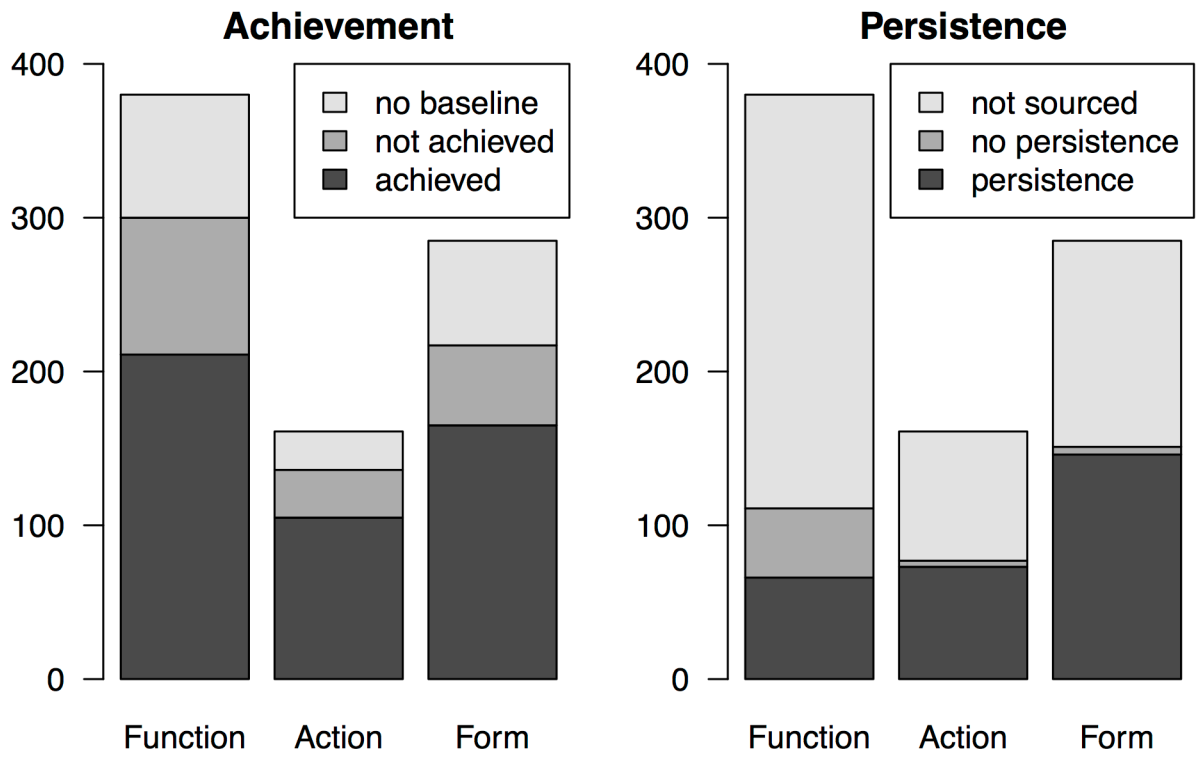

Figure A1. *Descriptive data on achievement and persistence of different types of targets*

If our results about the strategic choice of targets are correct, however, these descriptive data are likely to suffer from country-level selection effects. We find that countries that need to signal donor audiences are more likely to select form targets, while countries that have less of a need to signal donor audiences are more likely to select function targets (Table 2). These selection effects could work in two directions. Countries that receive concessional financing might have lower capacity and thus be less likely to meet and maintain performance on targets of any kind. However, these same countries may have stronger incentives to signal success to donors, making them more likely to strive for the achievement and maintenance of targets.

Another potential problem is that we were not able to collect data about achievement and persistence for all of the targets in our sample. To interpret regression results about the relative difficulty of targets in light of missing outcome data, we must make the strong assumption that the outcome variable is missing at random (MAR) — that missingness does not depend on the value of the outcome variable conditional on the predictor variables. It may be the case that even conditional on the predictor variables, targets that are not achieved or do not persist are less likely to be observed.

To address both of these potential problems, we developed a set of models to estimate *within-project* differences in achievement and persistence between form and function targets. If targets offer a strategic choice, then we should observe different rates of achievement and persistence even after we hold countries and projects constant. Fortunately, our sample includes many projects where countries select a mix of form and function targets. By estimating the effect of target type on the rate of achievement and persistence from *within-project* variation, we rule out the possibility that country-level selection effects drive our results.

Controlling for project-level variation in rates of achievement and persistence has the added advantage that we are more likely to satisfy the MAR assumption. To satisfy this assumption, we want to control for all factors that affect missingness. Controlling for project-level variation ensures that our results are not biased by missingness, to the extent that missingness is fully predicted at the project-level and with other included control variables. Even more reassuringly, when only the outcome variable is missing in logistic regression, the parameter estimates (not the intercept) will still be unbiased if the proportional odds assumption holds, even when the MAR assumption does not (Vach 1994). Note that multiple imputation would not help, since multiple imputation analysis and complete case analysis converge when missingness is only in the outcome variable. By controlling for variation through project-level fixed and random effects, rather than other independent variables that are subject to missingness, we remove bias caused by casewise deletion and decrease uncertainty associated with imputation of missing data.

We specify random-effects and fixed-effects models for both the achievement and persistence of targets. Our random-effects model is a hierarchical logistic regression model with project, country, and implementation level (municipal to national) random effects. The modeling strategy is a conservative approach to omitted variable bias and case selection bias, since we seek to account for all of the time-invariant country effects (e.g., statistical capacity), project-specific effects (e.g., difficulty of context for project implementation), and implementation-level effects (e.g., greater difficulty of supervising implementation by municipal agencies) that are independent from target type.

For both models related to the achievement and persistence of institutional targets, we include one minimally specified model that includes only the random-effects and the type of target (i.e., “form”, “action”, “function”) as the main predictor variable. In our second model for

both target achievement and persistence, we include an indicator for whether borrower performance was satisfactory during project implementation, as rated by the World Bank Independent Evaluation Group. In our first model of target persistence, we add a variable for the number of years between completion and our re-measurement data, since targets achieved long ago might be less likely to persist and the gap to re-measurement is not necessarily a project-specific characteristic. In our second model of target persistence, we also include a measure of whether the target was achieved during implementation, since this might correspond to borrower commitment or capacity, but this is not necessarily a project-specific effect.

Our fixed-effects model controls for project-level variation through project indicators. Because fixed-effects models must include variation on other predictor variables within the fixed-effects indicators, these models are restricted to a sample of projects where (1) both form and function targets were chosen; and (2) outcome data on either achievement or persistence exists for at least two targets. Table 3 summarizes the results, with the function target as the baseline category. The results justify our main assumption that form and action targets are easier to achieve and maintain than function targets, as estimated from within-project variation.

Table A1. *Achievement and persistence of institutional development targets*

DV:	**Achieve**	**Achieve**	**Achieve**	**Persist**	**Persist**	**Persist**
Form Indicator	**0.41** (0.28) [0.08]	**0.43** (0.29) [0.07]	**1.06** (0.55) [0.03]	**3.03** (0.51) [0.00]	**3.04** (0.59) [0.00]	**3.86** (1.42) [0.00]
Action Indicator	0.35 (0.32) [0.14]	0.40 (0.33) [0.11]	**0.89** (0.60) [0.07]	**2.55** (0.56) [0.00]	**2.60** (0.59) [0.00]	**4.18** (1.95) [0.02]
Satisfactory Borrower		0.44 (0.30)			-0.53 (0.50)	
Gap to Remeasure					0.11 (0.08)	
Achievement					-0.04 (0.49)	
Project	R.E. (208)	R.E. (204)	F.E. (81)	R.E. (147)	R.E. (126)	F.E. (39)
Country	R.E. (79)	R.E. (79)		R.E. (73)	R.E. (66)	
Level	R.E. (4)	R.E. (4)		R.E. (4)	R.E. (4)	
Observations	653	645	345	339	281	132
Deviance Reduction w/ Estimated Parameters	0.01	0.01	0.02	0.18	0.18	0.26

Model cells list: Parameter estimate; (Standard Error); [p-value of one-sided z-test]

To aid substantive interpretation of these models, we display first difference simulations for the increase in probability of achieving and then maintaining a form target versus a function target for the hypothetical average project and country based on our models (Figure A2). We take 1000 draws from the distributions of model coefficients, including average random effect levels for project and country, and then vary only whether the hypothetical target measured institutional function or form. These simulations show that on average choosing a form target for

a municipal project increases the probability of achievement from 73% to 79% and increases the probability of maintaining one's performance on the target from 59% to 96%. Thus, the choice of targets, even within the confines of specific projects that are more or less challenging to implement, offers an important strategic choice for both governments that receive development financing and World Bank staff.

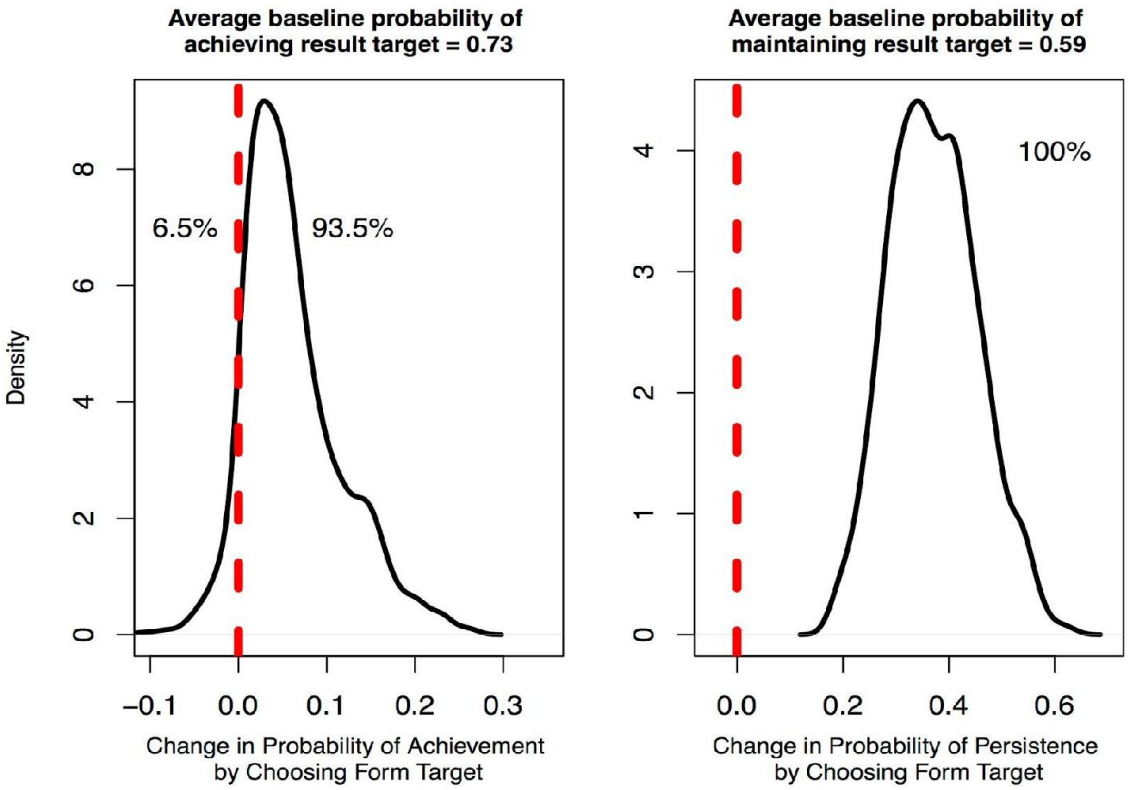

Figure A2. *The effect of target choice on achievement and persistence of institutional developments*

Appendix B. Additional Tests and Model Checks

As reported in the main text, the primary theoretical focus of this paper centers on the distinction between institutional form and function targets. From a theoretical perspective, we do not consider action targets to be substantially different from form targets since both involve measuring whether an institution or process exists, rather than whether the institution or process solves a public problem. For the main analysis reported in Table 2, we collapse form and action targets together as the baseline category. If we remove action targets from the sample entirely, we obtain the results reported in Table B1. In no case does the exclusion of action targets change the substantive results or diminish the level of statistical significance. This result justifies collapsing form and action targets into a single baseline category in the main models reported in Table 2.

Table B1. *Choice of function institutional development targets (at project approval) with action targets removed from the sample*

DV: Function Target	**1**	**2**	**3**
Project IDA proportion	**-0.95** (0.36) [0.00]		
Portfolio IDA proportion (AY)		**-0.73** (0.42) [0.04]	
IDA-only operational classification (AY)			**-0.83** (0.34) [0.01]
Resource Rents / GDP (% @ AY-1)	**0.03** (0.02) [0.09]	0.02 (0.02) [0.13]	**0.03** (0.02) [0.07]
New Environmental Ministry (≤ 5 yrs.)	-0.44 (0.39)	-0.53 (0.39)	-0.56 (0.39)
Established Environmental Ministry (≥ 6 yrs.)	-0.31 (0.47)	-0.39 (0.46)	-0.27 (0.47)
Approval Year (centered linear)	-0.02 (0.04)	-0.03 (0.04)	-0.05 (0.04)
Environment Sector Board	**-1.65** (0.38)	**-1.49** (0.36)	**-1.60** (0.37)
Level R.E.	Yes (4)	Yes (4)	Yes (4)
Country R.E.	Yes (84)	Yes (80)	Yes (84)
Observations	649	644	649

Model cells list: Parameter estimate; (Standard Error); [p-value of one-sided z-test]
All models are random-intercept logit fitted by Laplace approximation with levels as indicated

Appendix C. Coding Procedures for Targets

We found that projects with at least 10% financing for the strengthening of environmental institutions, calculated using World Bank sector and theme codes, reliably contained institutional performance targets that met our coding definitions. Due to the resources involved with this large coding effort, we chose to focus on the documentation where we could reliably identify targets. We did not restrict the sample of project based on the type of target or the sub-sector of the project, so we consider these targets to be broadly representative of institutional targets.

Trained coders then systematically extracted targets from these documents, checking the introduction, sections about results or outcomes, appendices, and the conclusion. All targets were linked to the documents from which they were collected and the page and/or section number for their locations were recorded. In addition to recording the raw target text, our team also compiled any available measurements of the target value at the project baseline and completion for the in-sample dataset.

After extracting the full set of indicators, three research assistants independently coded each target after completing a known training set with at least 90% reliability. For 72% of targets, all three coders agreed. For the disputed codes, one of the co-authors arbitrated and assigned the final code. Either all coders agreed or the arbitrator agreed with the majority of the coders 95% of the time. In the remaining 5% of cases, the arbitrator agreed with the minority coder. Of the 826 targets, 380 were coded as measuring function, 285 as measuring form, and 161 as measuring action. A large number of projects contained both form and function targets, which allows us to examine the average relative difficulty of targets *within* projects (see Appendix A).

We chose to use targets as our variable of interest rather than projects because often a single project had targets implemented at different levels of government (e.g., both national and municipal level targets). Focusing on targets rather than projects allows us to account for the impact that levels of implementation may have on target selection in our model.

Appendix D. Robustness of Bandwidth Choices Near Eligibility Threshold

The two bandwidth choices that we report in Table 3 have substantive meanings, but they are not the only choices available. To further examine the robustness of our results at different income bandwidths, we iteratively subset the dataset based on closeness to the eligibility threshold for non-concessional financing. We then estimate our primary models from Table 2 on each of these subsets and record the coefficient estimate and standard error. In Figure D1, we report the results of this process, plotting coefficient estimates and 90% confidence intervals (equivalent to one-tailed test at $\alpha = 0.05$ when the standard error line crosses the zero horizontal). For the majority of specifications, we find a significant, negative effect of various measures of concessional-only status, even after the very substantial observational penalty involved with losing more than half of our sample. The results are especially strong for our primary and cleanest measure — concessional-only classification. Taken together this collection of results rules out an income effect; the need to signal foreign donors is the most plausible explanation for the choice of targets in institution building.

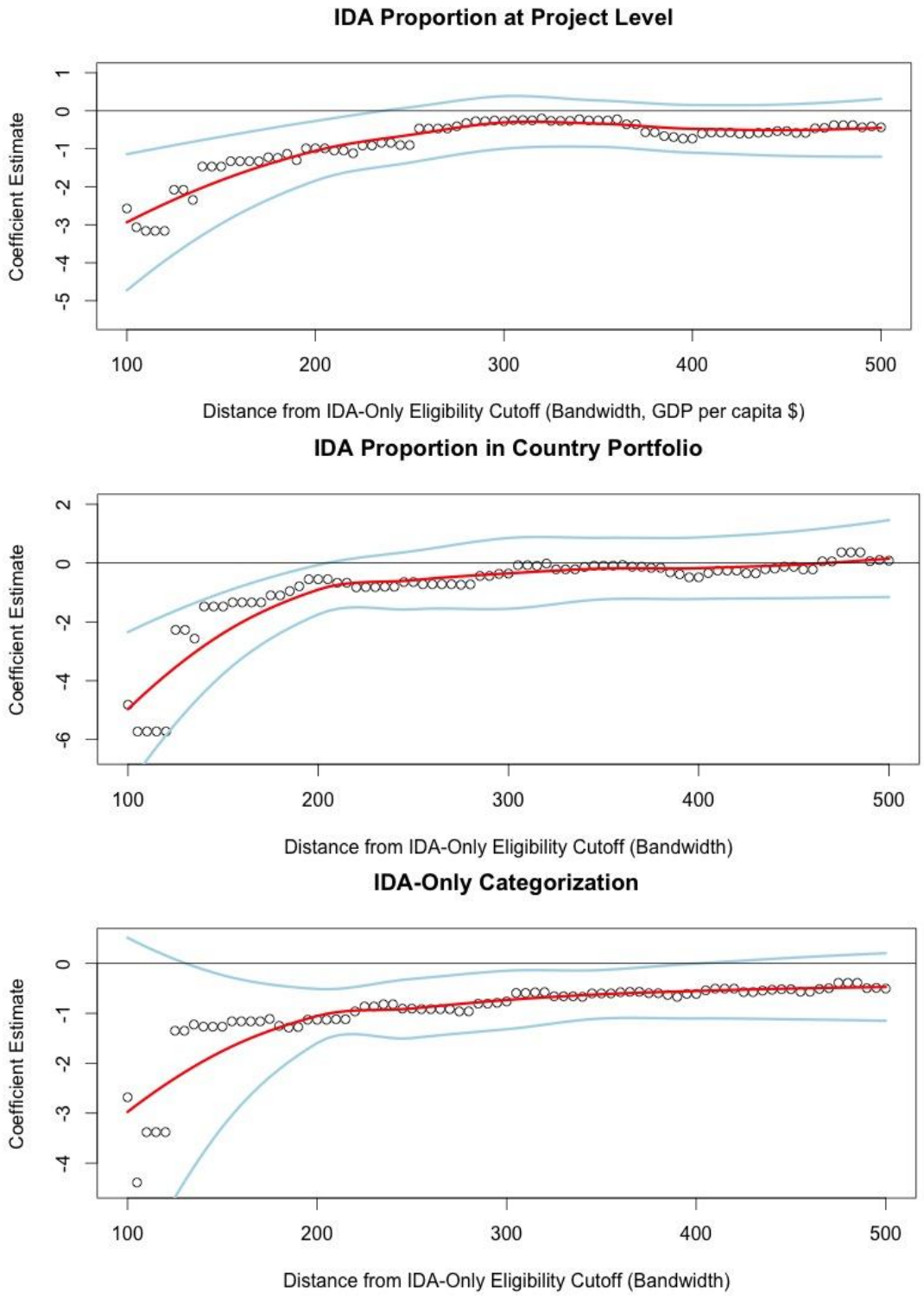

Figure D1. *Estimates of the effect of concessional lending across different income bandwidths*