

Aid, Institutions, and the Potential of Anti-Corruption*

Michael Denly[†]

January 15, 2024

For the most recent version of this paper, please visit:

<https://mikedenly.com/research/aid-corruption-gaaps>

Abstract

Does anti-corruption improve aid effectiveness? I use insights from principal-agent theory, incomplete contracting, and historical institutionalism to argue that context-specific anti-corruption measures improve aid effectiveness. First, anti-corruption measures generally target institutional constraints, which a large literature suggests are the primary cause of development outcomes. Second, during the critical juncture period in the late 1990s, legitimacy challenges from civil society spurred aid agencies to create large anti-corruption infrastructures. Context-specific anti-corruption design measures thus have both the power and specificity to meaningfully contribute to aid effectiveness. To test the hypothesis, I individually coded all 4,303 World Bank investment projects approved from 2001-2018 for their use of context-specific, project-level Governance and Anti-Corruption Action Plans (GAAPs). Using frontier matching for causal inference, I find that projects with GAAPs have circa 3% better outcomes than similar projects without GAAPs. The results suggest that top-down monitoring remains a useful foil against corruption, and bureaucrats have agency to mitigate the institutional and structural constraints to development.

*For excellent research assistance, I thank Briana Rocha, Gavin Loyd, Hanna Herbowy, Nehaarika Shridha, Trey Durio, and Angelina Braese. For feedback or advice, I thank Arthur Alik-Lagrange, Mark Buntaine, Steve Burgess, Luis Cano, Matt Carnes, Olgahan Cat, Lorenzo Crippa, Valerie de Koeijer, Stephan Egli, Mike Findley, Desh Girod, Mirko Heinzl, Kyosuke Kikuta, Eric Langenbacher, Rich Nielsen, Dan Nielson, Xin Nong, Cleo O'Brien-Udry, Anna Pinto Hebert, Naseer Rana, Pascale Schnitzer, Joel Turzewitz, Charlie Undeland, Alex Wais, Matt Winters, Maria Vannari, and audiences at the Junior IO Scholars Workshop, the Institute for Advanced Study in Toulouse, the Swiss State Secretariat for Economic Affairs (SECO), the International Studies Association (ISA) Annual Meeting, and the Midwest Political Science Association (MPSA) Annual Meeting. All interviews referenced in this document were approved by Georgetown University's Institutional Review Board (2014-1055).

[†]Assistant Professor, Texas A&M University ✉ mdenly@tamu.edu

The paradox of foreign aid, once proclaimed by Peter Bauer (1972), is that the countries that need aid most are the same ones where it is least likely to work.¹ In particular, foreign aid is least likely to work in countries with weak institutions (Burnside and Dollar, 2000). Especially given that multilateral aid donors alone spent an estimated \$US 7.8 trillion on foreign aid from 1943 to 2013,² and a large portion of that colossal amount went to countries with weak institutions,³ it is essential to know: When do bureaucrats have agency to make aid more effective by addressing the structural constraints to development imposed by weak institutions and corruption?

Although corruption and weak institutions pose real impediments to aid effectiveness, I use insights from principal-agent theory, incomplete contracting, and historical institutionalism to argue that context-specific anti-corruption measures can mitigate these impediments. First, anti-corruption measures mostly target institutional constraints, which a large literature suggests are the primary cause of development outcomes (e.g., Acemoglu and Robinson, 2012). Second, during the critical juncture period surrounding former World Bank president James Wolfensohn’s (1996) “Cancer of Corruption” speech, legitimacy challenges from civil society spurred aid providers to change how they approached corruption. To curtail corruption-related legitimacy costs that affect both multilateral development banks (MDBs) and the donors overseeing them,⁴ aid providers have invested in large anti-corruption infrastructures (Rose-Ackerman and Carrington, 2013). Simultaneously, aid agencies received increased political backing, added numerous staff members dedicated to anti-corruption, and developed relevant expertise to better exercise their “fiduciary duty”.⁵ Notably, in some cases, MDB bureaucrats have worked with aid-receiving countries to include context-specific anti-corruption action plans at the project level. In practice, designing such action plans are costly in terms of bureaucrats’ time and effort. Nonetheless, they contribute to better

¹I borrow this idea from a 2023 talk at the London School of Economics by William Easterly. For a tweet expressing the same idea, see: https://twitter.com/bill_easterly/status/1275854422535475202?lang=en.

²These figures are presented in 2011 US dollars and come from AidData (Tierney et al., 2011).

³See, for example, Alesina and Weder (2002).

⁴See, for example, Clausen, Kraay and Nyiri (2011) and Johnson (2011).

⁵The “fiduciary duty” refers to the clause in the founding Articles of Agreement of all major MDBs, specifying that aid funds need to be spent for their intended purposes (e.g., World Bank, 1945).

development outcomes, because they help make the incomplete contract between the aid provider and recipient country more complete. By extension, aid agency bureaucrats have agency to mitigate the structural constraints posed by weak institutions and corruption.

To operationalize aid projects operating in weak institutions, I individually coded all 4,303 World Bank investment projects approved from 2001-2018 for their use of context-specific, project-level Governance and Anti-Corruption Action Plans (GAAPs). The latter capture aid projects operating in weak institutions, because the World Bank only uses GAAPs as an *additional* layer of top-down monitoring controls. That is, they supplement the institutional risks analyses that all projects undertake. In total, 362 of the projects during the time period of study utilized GAAPs for the period under study, making the share of projects with GAAPs around 8% of the sample. Given potential selection effects and post-treatment bias, my sample only counts GAAPs in projects issued before World Bank Executive Board approval. Thus, “problem projects” that the World Bank assigns a GAAP post-approval fall outside my sample, thereby mitigating endogeneity concerns. I also separately test for such concerns and find no evidence to support them.

To capture World Bank project success, I follow a large literature and use Independent Evaluation Group (IEG) project outcome ratings, which fall on a 1-6 scale. IEG outcome ratings represent an excellent measure of project success. In particular, they generally downgrade outcome ratings in Implementation Completion Reports (ICRs) completed by project Task Teams and their consultants, thereby guarding against potential biases.

To test the causal effects of GAAPs on IEG outcome ratings, I use King, Lucas and Nielsen’s (2017) frontier matching. Like any matching method, frontier matching finds projects with GAAPs and compares them to very similar projects without GAAPs. What distinguishes frontier matching from other matching methods is that it (re-)tests the causal effect of the treatment using the maximum balance for each observation in the sample. That re-testing of estimates at the entire range of possible sample sizes is crucial: for matching estimates to be credible, they must not be sensitive to researcher design choices that

may introduce bias-variance trade-offs or statistical power challenges. In any case, given that GAAP decisions are mostly project-specific, they do not have a clear panel structure or staggered adoption pattern. Accordingly, cross-sectional frontier matching is a more suitable method than potential alternatives, including the augmented synthetic control method and panel matching (see Ben-Michael, Feller and Rothstein, 2021; Imai, Kim and Wang, 2023).

Consistent with past literature stressing that matching is most suitable as a pre-processing method to reduce model dependence (Ho et al., 2007), I run two sets of parametric tests after matching. The first uses linear regression, and the second uses ordered multilevel logistic regression with country random effects. For both sets of models, I find that GAAPs indeed have a positive effect on World Bank outcomes across all possible balance and sample size configurations. Regarding the size of these estimates, the coefficients on the more directly interpretable linear regression estimates suggest that GAAPs improves project outcomes by an average of 0.14 points, which corresponds to a 2.3% percentage point increase given the six-point scale of the IEG scores. The ordered multilevel logistic estimates are also similar: they suggest that projects with a GAAP are, on average, 1.3 times more likely to receive the highest possible outcome rating than projects without a GAAP. Additionally, Athey and Imbens (2015) intervals on potential model misspecification are narrow, suggesting that the results are robust.

The paper makes two larger contributions. First, it shows that recent scholarly attacks on the utility of top-down anti-corruption monitoring from Persson, Rothstein and Teorell (2013), Kenny (2017), and others need further qualification. More broadly, the effectiveness of GAAPs shows that anti-corruption monitoring works in the presence of incomplete contracts, which are even weaker than the much-maligned principal-agent relationship. I reference incomplete contracts,⁶ because aid financiers are agents, not principals (Nielson and Tierney, 2005, 786). As such, recipient countries can “hold-up” agents due their power advantage. The latter derives from agents’ disbursement imperative, bureaucrats’ need to

⁶For more on incomplete contracts, see, for example, Hart (2017).

complete projects for career purposes, and the fact that all aid project preparation and supervision involves sunk costs.

Second, the present paper helps clarify the conditions under which bureaucrats have agency to mitigate structural constraints to achieving development outcomes. On that score, the literature is very pessimistic. For example, Acemoglu and Robinson (2012), Rodrik, Subramanian and Trebbi (2004), and others show that institutions are the strongest drivers of development outcomes, thereby implying that GAAP success is very unlikely. Another source of pessimism comes from the literature arguing that aid is a fungible resource that leaders and bureaucrats capture aid for political and personal ends (e.g., Bueno de Mesquita and Smith, 2009; Andersen, Johannesen and Rijkers, 2022). Still another source of pessimism derives from Easterly's (2006, 2015) critiques of the aid industry and the purported failure of what Acemoglu and Robinson (2012) call "the ignorance hypothesis".⁷ However, aid is mostly not fungible,⁸ the Western advice literature has recently reversed course with better data and methods,⁹ and the present study shows that institutions are not fully deterministic.¹⁰ A key reason why is that GAAPs are context-specific, so they avoid the "isomorphic mimicry" trap of assuming that best practices work across all contexts (see Andrews, Pritchett and Woolcock, 2017).

The results of the present study thus align with recent work stressing the essential nature of good project design (e.g., Ashton et al., 2023). Similarly, the results align with Denizer, Kaufmann and Kraay (2013) and Bulman, Kolkma and Kraay (2017), who find that most development outcomes at the World Bank and Asian Development Bank relate to project-level features, not macro-level country characteristics. More broadly, bureaucrats can mitigate at least some of the risks of weak institutional environments when they design

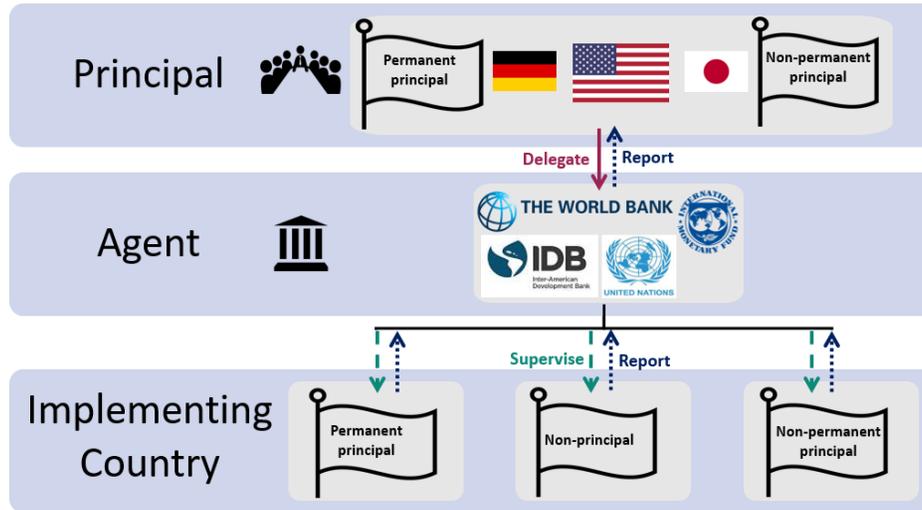
⁷The ignorance hypothesis refers to the idea that developing countries need Western advice to obtain better development outcomes (Acemoglu and Robinson, 2012).

⁸See, for example, Feyzioglu, Swaroop and Zhu (1998), van de Walle and Mu (2007), Altincekic and Bearce (2014), Bermeo (2016), and Jones and Tarp (2016).

⁹See, for example, Estevadeordal and Taylor (2013), Easterly (2019), and Grier and Grier (2021).

¹⁰This conclusion also aligns with Glaeser et al. (2004) and the latest work on the aid-growth nexus (e.g., Arndt, Jones and Tarp, 2015, 2016).

Figure 1: An Augmented Principal-Agent Framework for Foreign Aid



anti-corruption measures to address those risks.

1. Theoretical Framework

The present paper’s starting point for theorizing about the extent to which anti-corruption can improve aid effectiveness is the most commonly-used theoretical framework in the literature: the principal-agent model. It stresses that powerful donor countries (principals) delegate authority to their national development agencies or institutions like the World Bank (agents) to carry out their goals (Hawkins et al., 2006). Principals are mostly interested in pursuing development ends but sometimes use aid to advance their own strategic foreign policy goals (Stone, 2011), so principals are not always principled. For their part, agents have their own interests of financial “security, legitimacy, and policy advancement” (Johnson, 2013, 183), which agents pursue through the development of rules and organizational cultures (Barnett and Finnemore, 2004; Weaver and Nelson, 2016).

Given that the effectiveness of anti-corruption in aid depends on implementation risks, and most aid recipients to carry out implementation in line with the 2005 Paris Declaration on Aid Effectiveness,¹¹ I augment the typical principal-agent model. Figure 1 summarizes

¹¹See OECD (2005).

my framework, which includes an equal focus on the implementation country/aid recipient. The augmented principal-agent framework mostly corresponds to multilateral aid but also applies to some bilateral aid donors, such as France and Germany, which ensure that recipient countries implement their own projects (see Dietrich, 2021).

Regardless of the extent to which the augmented principal-agent framework applies to bilateral aid, it is necessary to characterize the nature of the agent's supervisory role over the implementer. On that score, Gutner (2005) suggests that scholars can analyze the agent's supervisory role over the implementer as a separate principal-agent relationship. However, Nielson and Tierney (2005) conclusively show that doing so entails stretching the principal-agent framework beyond its purview.¹²

Given that Nielson and Tierney (2005) did not provide an alternative theoretical framework for understanding the agent-implementer relationship, I propose that scholars can understand the agent's supervisory role over the country implementors as an incomplete contracting problem (see Hart, 2017). Fundamentally, just like the two firms in Grossman and Hart's (1986) seminal model of incomplete contracting, aid projects involve significant sunk costs in terms of preparation and supervision. Accordingly, once an agent begins a project, it has an incentive to finish it.

Recall agents' incentives of financial security, legitimacy, and policy advancement; the incomplete contract between agents and implementer relates to two of the three criteria. Failing to complete projects negatively affects financial security, given that agents earn interest and service fees on loans, credits, and grants.¹³ That is why agents have a so-called "disbursement imperative" (see Buntaine, 2016; Weaver, 2007, 84). With respect to legitimacy, failing to complete projects reflects poorly on agents' legitimacy and invites forum shopping, which has a negative downstream impact on financial security. From the per-

¹²Aid financiers like the World Bank are not true principals, because aid-receiving countries do not receive a conditional delegation of authority from aid financiers. In simpler terms, because aid-receiving countries are sovereign entities, they do not need permission from an aid financiers to operate on their own territory (Nielson and Tierney, 2005, 786).

¹³Although development credits to poorer recipient countries are technically interest-free, agents still earn money on credits through service fees, which are typically around 0.75% of credit amounts.

spective of bureaucrats, failing to complete projects also negatively affects career prospects, because such achievements are easy to monitor (Martens et al., 2002).

Aid-receiving countries are aware of the above issues, which provides with them a power advantage and a natural hold-up problem, and principal-agents challenges make these dynamics more severe. As Figure 1 demonstrates, implementing countries can simultaneously serve as a principal through their membership on executive boards, so the agent often supervises the principal in a compromised position of power. The most complete contract specifying zero tolerance for corruption cannot solve these power dynamic challenges. After all, monitoring corruption is often highly costly for the agent (Bourguignon and Gunning, 2020).

By the same token, lax enforcement of corruption problems risks loss of agent resources as well as legitimacy challenges from corruption scandals, which both principals and agents seek to avoid. Although agents suffer direct costs from legitimacy challenges, Johnson (2011) shows that principals suffer these costs as well through “guilt-by-association”. Essentially, because powerful states control aid agencies by means of their executive board positions (e.g., Hawkins et al., 2006), citizens across the world view the powerful states as responsible for the negative outcomes fostered by multilateral aid agencies that they steward. Corruption is also a particularly significant determinant of popular distrust in public institutions (Clausen, Kraay and Nyiri, 2011), so principals have a strong incentive to quell such concerns.

1.1. The Critical Juncture: How The Legitimacy Costs of Corruption Eclipsed Its Monitoring Costs

Until the start of the critical juncture period in the 1990s, the costs of monitoring corruption and foregoing disbursements outweighed the legitimacy costs from corruption scandals. The lack of focus on institutions and corruption stemmed from aid’s focus on the technical challenges of providing infrastructure (Kapur, Lewis and Webb, 1997), prohibitions

on engaging with political issues of aid recipients,¹⁴ the fact that institutional lending only began in the 1980s,¹⁵ and the predominant intellectual paradigm guiding aid: modernization theory. It notably stressed that economic growth would lead to democratization (e.g., Lipset, 1959), and institutional problems such as corruption were a relatively minor nuisance that may even be beneficial for growth and development (Leff, 1964; Huntington, 1965; Nye, 1967).

The actual critical juncture that made the legitimacy costs of corruption higher than the costs of monitoring and enforcing it for both principals and agents started during the 1990s. It marked the first decade since the end of the Cold War, which fostered a decline in the extent to which principals leaned on agents to provide favorable treatment to strategically important countries (Dunning, 2004; Bearce and Tirone, 2010; Bermeo, 2016). Even more crucially from the perspective of corruption were the major corruption scandals that rocked Italy, Brazil, Japan, Mexico, India, Spain, Pakistan, Ecuador, Georgia, Germany, Peru, the United States, South Korea, and Switzerland (Newell and Bull, 2003; Manzetti and Wilson, 2007; López Claros, 2015; Fisman and Golden, 2017). Famously dubbed the “corruption eruption” by Naím (1995), numerous high-ranking government officials and heads of state resigned or were impeached during this period. At around the same time, the confluence of unpopular structural adjustment programs, greater popular awareness of corruption, and other globalization pressures resulted in massive protests and legitimacy challenges for international organizations around the world. Some of the events that spurred the greatest legitimacy challenges were the 1994 Madrid protests;¹⁶ Jeffrey Winters’s (1997) press conference alleging that \$10 billion in World Bank money disappeared due to corruption in Indonesia; and the 1999 “Battle in Seattle” (see also Figure 2).¹⁷ Subsequently, the United

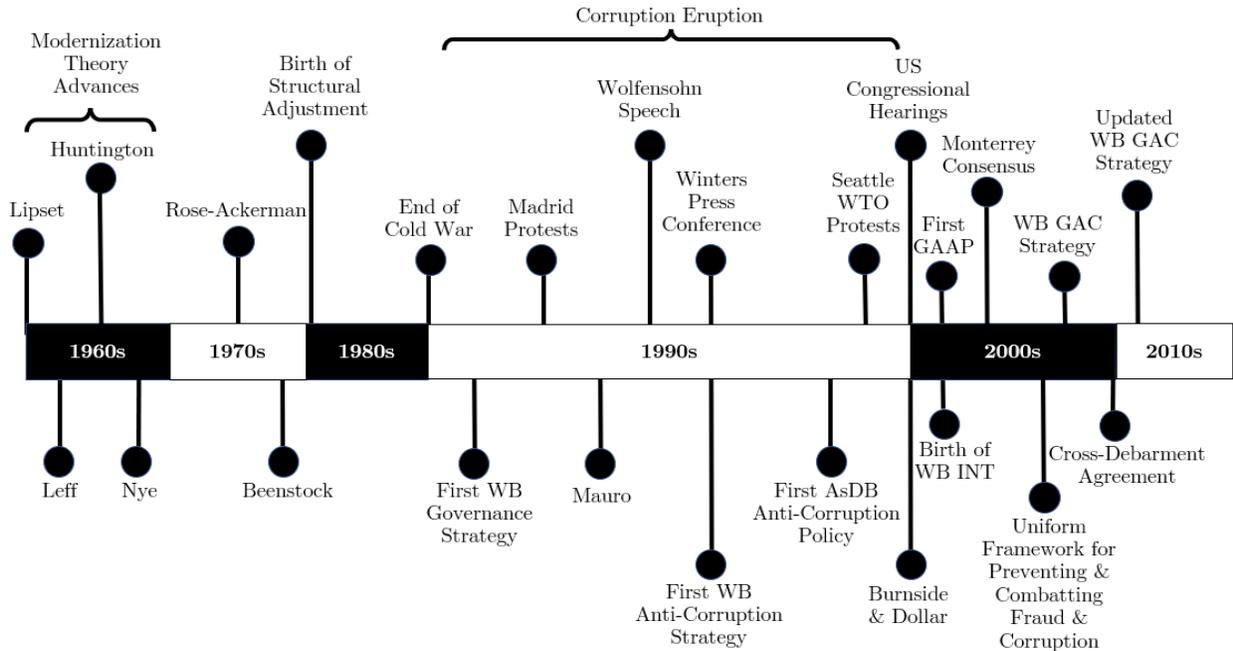
¹⁴See Article IV, Section 10 of World Bank (1945) and Article IV of IMF (1944).

¹⁵Here, I am referring to structural adjustment, which the World Bank and International Monetary Fund (IMF) began in 1980. Providing a full review of structural adjustment lending is beyond the scope of the present study. For excellent reviews of structural adjustment, see, for example, Easterly (2005) and Sharma (2013).

¹⁶In 1994, crowds in Madrid greeted former World Bank President James Wolfensohn with chants of “fifty years is enough”, referring to the institution’s then 50-year mandate (Levy, 2014, 203).

¹⁷In what became known as the “Battle in Seattle”, massive protests turned violent over a World Trade Organization meeting. For more, see Zürn (2004).

Figure 2: Key Events and Scholarship Underpinning Aid Financiers' Reversal on Corruption



States invoked congressional hearings on the World Bank's ability to control corruption (US GAO, 2000).

To change the institution's trajectory on corruption and attempt to overcome the legitimacy challenges, in 1996 World Bank President James Wolfensohn delivered a famous speech in which he decried the "cancer of corruption" (Wolfensohn, 1996). Thereafter, among other things, the World Bank developed its first anti-corruption strategy and significantly increased staffing in financial management and procurement (World Bank, 1997); drafted its first project-level anti-corruption strategy for the financing of Indonesia's Second Kecamatan Development Program loan in 2001;¹⁸ established the Integrity Vice Presidency (INT) to investigate "allegations of fraud and corruption in World Bank-financed projects" (World Bank, 2002); and stipulated precise anti-corruption guidelines for aid recipients' use of World Bank financing and sanctions in the case of fund misuse (World Bank, 2006).

Consistent with diffusion processes on other policies (e.g., Heldt and Schmidtke, 2019), the Asian Development Bank (1998), African Development Bank (2006), and Inter-American

¹⁸For more, see World Bank (2003).

Development Bank (2009), among other multilateral aid agencies,¹⁹ have largely followed the World Bank's lead on anti-corruption. For example, in 2002 these agencies and others signed the UN-backed Monterrey Consensus, in 2006 they established a Uniform Framework for Preventing and Combating Fraud and Corruption, and in 2010 they agreed to cross-debarment on sanctions policy (International Financial Institutions Anti-Corruption Task Force, 2006; World Bank, 2010*b*).

1.2. Controlling Local-Level Institutional Risks

Having a larger infrastructure and remit to deal with corruption and other institutional risks helps with curtailing relevant legitimacy costs and some risks, but they are not sufficient to consistently produce better development outcomes. The reason is that corruption, aid effectiveness, and institutional challenges are context-specific (Grindle, 2004; Gingerich, 2013). In particular, isomorphic mimicry of actions across projects is insufficient due to the differing implementation capacities of states and institutions within them (Andrews, Pritchett and Woolcock, 2017). That is especially the case for corruption: without context-sensitivity, “theories of change are meaningless” (Trapnell and Recanatini, 2017, 490).

Context-specific approaches to corruption and institutional risks also work, because there is now a significant knowledge base from which researchers and practitioners can draw. For example, aid agencies and governments around the world have effectively used financial, technical, and social audits to expose corrupt politicians, measure the quality and probity of road construction, and track the consequences of social fund diversion (e.g., Reinikka and Svensson, 2004, 2011; Olken, 2007; Ferraz and Finan, 2008; Bobonis, Cámara Fuertes and Schwabe, 2016; Gans-Morse et al., 2018). Similarly, procurement controls and e-procurement systems have constituted crucial means to prevent aid fungibility and elites from capturing the trajectory of bureaucrats' careers (e.g., Charron et al., 2017; Lewis-Faupel et al., 2016). Although the evidence on social accountability measures is more mixed, community-monitoring, citizen scorecards, and participatory budgeting, among other measures, have

¹⁹It is also possible list agencies such as the Islamic Development Bank (2012).

yielded positive public goods outcomes in various countries (Björkman and Svensson, 2009; Banerjee et al., 2010; Olken, 2010; Casey, Glennerster and Miguel, 2012; Joshi, 2013; Touchton and Wampler, 2013; Fox, 2015; Khemani et al., 2016; Björkman, de Walque and Svensson, 2017; Casey, 2018).

Better, institutionally-focused project design can thus help mitigate literature’s overall conclusion is that the wealth of anti-corruption information and tools available to practitioners has not led to clear progress in terms of overcoming the challenges of weak institutions (see Fukuyama and Recanatini, 2021, 472-473). In theoretical terms, better project design that addresses institutional and corruption risks helps make agent’s incomplete contracts with aid recipients/implementers more complete.

H1: Aid projects in weak institutional environments will be most likely to succeed when they contain a context-specific anti-corruption action plan to address those risks.

2. Research Design

2.1. Governance and Anti-Corruption Action Plans (GAAPs)

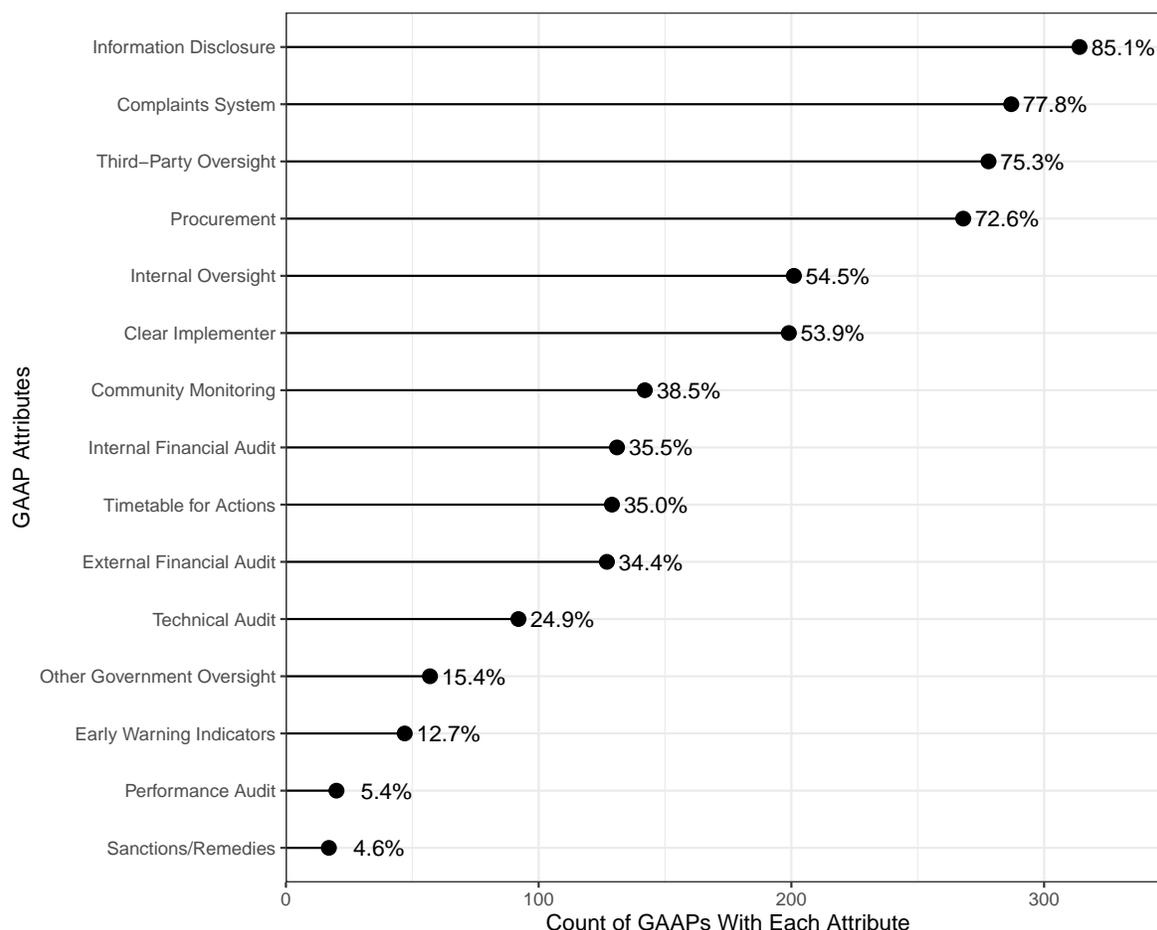
To demonstrate the empirical relevance of the argument, I examine the effectiveness of the World Bank’s use of Governance and Anti-Corruption Actions Plans (GAAPs) on improving project-level development outcomes. Although GAAPs often have slightly different names,²⁰ GAAPs comprise a project-specific mix of anti-corruption tools that *supplement* existing project-level requirements.²¹ Figure 3 provides a detailed breakdown of the attributes of GAAPs, which the World Bank only uses in investment lending, not more fungible budget support, including Program for Results and structural adjustment loans.²² As Figure 3 un-

²⁰See Appendix A for more details.

²¹In examining potential GAAPs, I only found 8 projects with governance or anti-corruption action plans that were not specific to the individual projects. Given this very small number relative to the overall number of GAAPs and projects, the present study does not focus on country-level action plans.

²²Structural Adjustment Lending (SAL), Development Policy Lending (DPL), and Development Policy Financing (DPF) are all equivalent. The World Bank currently refers these instruments as DPFs. In any

Figure 3: Governance and Anti-Corruption Action Plan Attributes (2001-2018)

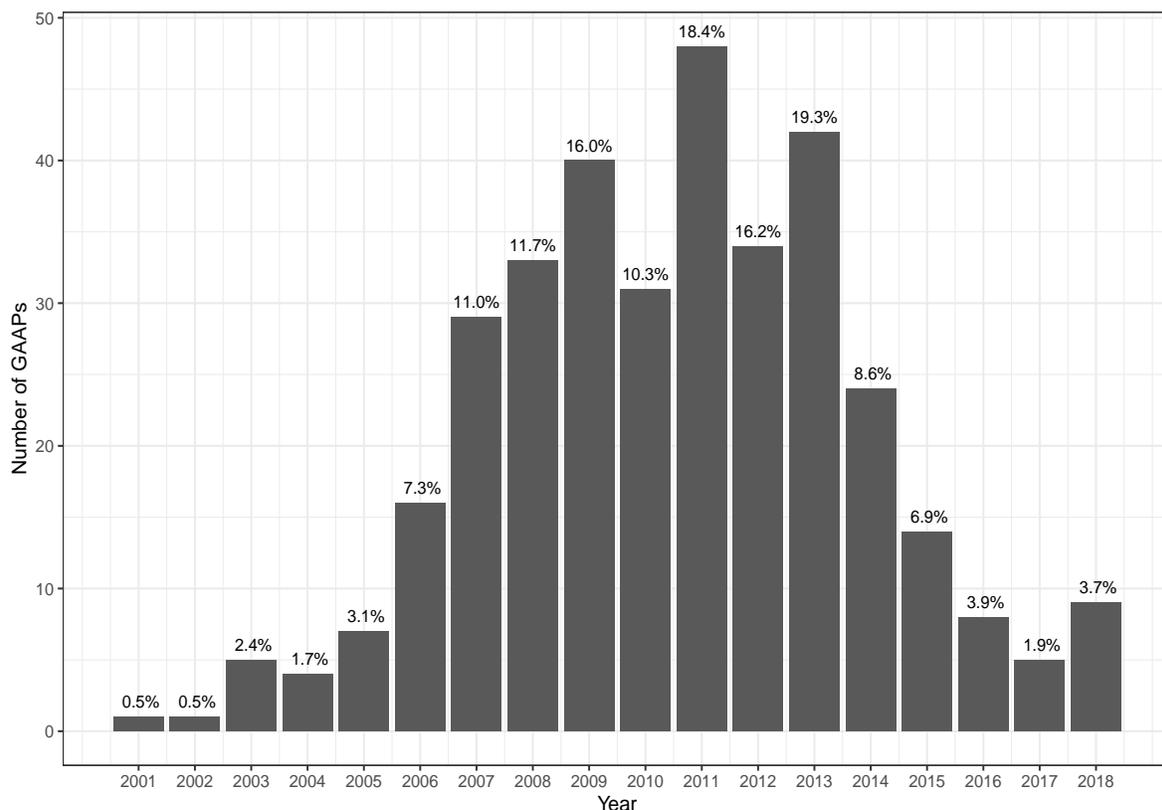


Note: Own coding based on project approval documents as well as project evaluation documents. The coding only includes projects with GAAPs present in the relevant approval document, because coding post-hoc GAAPs used for poorly-performing projects might introduce an endogeneity problem or post-treatment bias. Appendix A provides further details on the coding, and the percent numbers above refer to the share of projects with GAAPs that have each attribute.

derscores, some of the most prominent attributes include additional information disclosure, grievance redress systems, and third-party monitoring. GAAPs also use various types of audits, procurement controls, and sanctions measures slightly less frequently. Regardless, all of these measures contribute to better project outcomes by helping ensure that the aid

case, the World Bank does not use GAAPs for budget support. Accordingly, concerns relating to loan type targeting (Winters, 2010), the failings of conditionality involving prior actions (e.g., Svensson, 2003), and the potential fungibility of general budget support fall outside the scope of the present study. This distinction is critical because project-related investment aid is generally not fungible, especially as compared to budget support.

Figure 4: Yearly Usage of Governance and Anti-Corruption Action Plans (2001-2018)



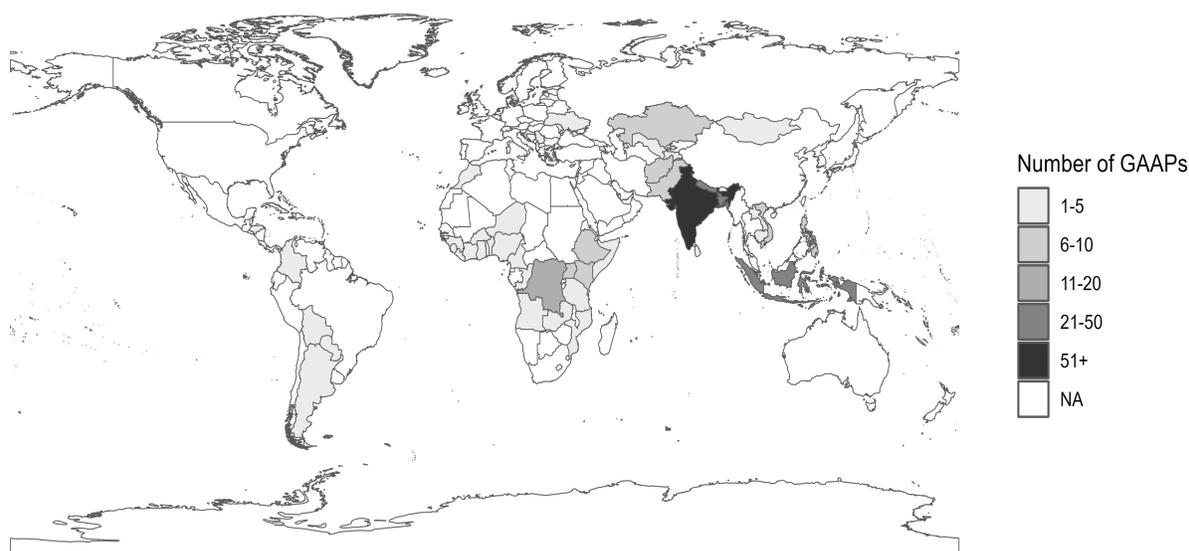
Note: The percentages above the bars above refer to the share of projects in each year with a GAAP.

is used for its intended purposes.

Following the aforementioned legitimacy scandals from corruption in Indonesian lending (Rich, 2002), the World Bank designed its first GAAP in 2001 for the financing of Indonesia's Second Kecamatan Development Program.²³ Notably because senior management and Board of Directors advocated for their use in the World Bank's (2007) combined Governance and Anti-Corruption (GAC) Strategy, GAAPs figured prominently in the proceeding years (see Figure 4). However, the World Bank (2012) updated that strategy a few years later, and GAAPs did not receive a single reference in the update, portending their subsequent decline in use.

²³Indonesia's Kecamatan Development Program's famous community-driven development project has previously received scholarly attention (e.g., Olken, 2007).

Figure 5: Country Usage of Governance and Anti-Corruption Action Plans (2001-2018)



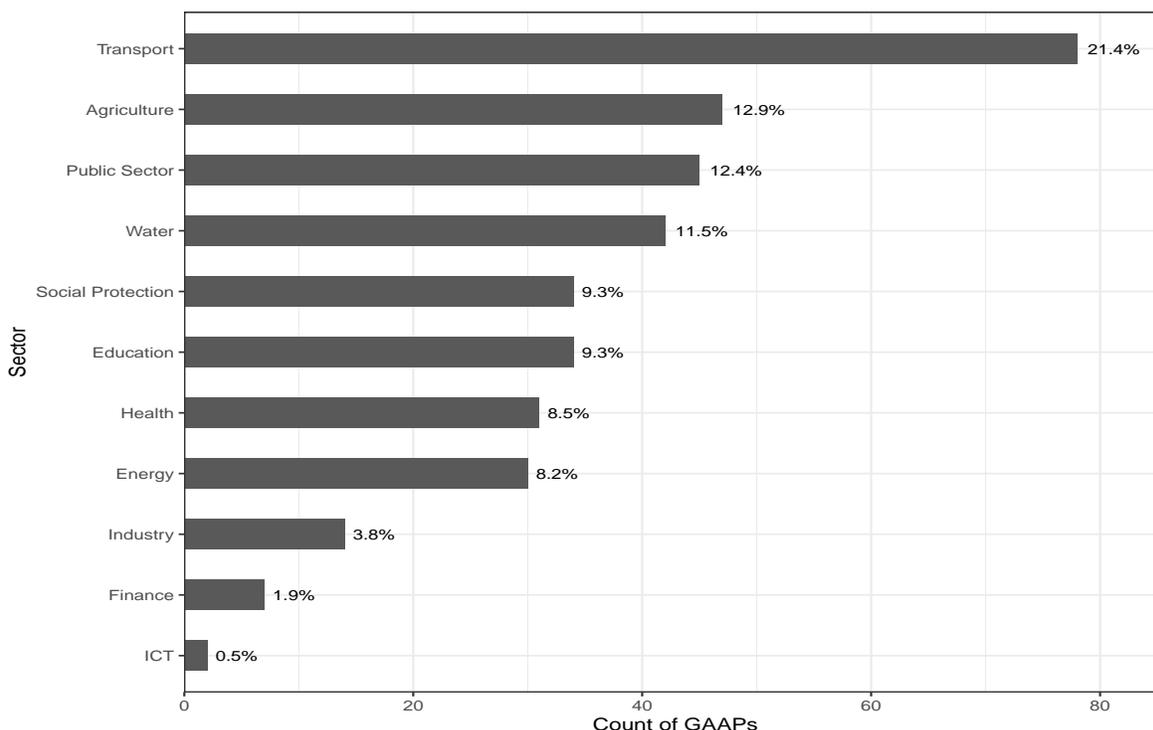
Indonesia, India, Bangladesh, and Nepal have received the most GAAPs (see Figure 5), which was largely a result of country-level decisions to include them in all of the countries' respective projects for at least some time. Outside of Asia, the Democratic Republic of the Congo, Uganda, and Kenya employed circa 5-12 GAAPs from 2001-2018. For its part, projects for Latin American countries used GAAPs more sparingly, with only Argentina and Honduras receiving a total of GAAPs 4 for the 2001-2018 period. In total, 59 countries used these tools for the study period, encompassing around 8% of approved projects.²⁴ Based on author interviews with project Team Leaders and other World Bank staff, GAAPs diffused based on Team Leaders taking up new positions as well as at the direction of regional World Bank anti-corruption advisors (see also World Bank, 2009, 2010*a*, 2013).

Sectorally, as Figure 6 demonstrates, the transport sector has the largest share of projects with GAAPs, totalling 21%. The infrastructurally-focused sectors of agriculture (13%) and water (12%) similarly use GAAPs quite frequently. Nevertheless, GAAPs are not merely for infrastructure, as public sector (12%) and social protection (9%) projects also use GAAPs.

The present study only considers GAAPs incorporated into the design of a project prior

²⁴See Appendix A for more details on the coding strategy.

Figure 6: Sectoral Usage of Governance and Anti-Corruption Action Plans (2001-2018)



Note: Given that World Bank projects have up to 5 sectors, I follow Heinzel (2022) and assign each project only one sector, corresponding to the largest share of financing for each project. The percent number above reflect the share of all projects with GAAPs from the respective sector.

to Board approval. Such a design feature mitigates potential endogeneity or post-treatment bias concerns associated with failing projects receiving a GAAP post-hoc. In any case, given the potential for endogeneity or selection problems, I further investigate them in two ways. First, I test whether GAAP incidence correlates with the Varieties of Democracy’s political corruption measure, which McMann et al. (2022) show is likely the best available country-level corruption measure. I find that the two measures only correlate at 0.06. The second endogeneity/selection test uses a more relevant measure for World Bank bureaucrats: the institution’s official yearly designation of countries that staff assess to be “fragile states”. Given that GAAPs and fragile state designations correlate only at -0.03, it is difficult to argue that endogeneity or selection concerns characterize the use of GAAPs.

2.2. Dependent Variable: Project-Level Outcome Ratings

To assess the effectiveness of GAAPs on development outcomes, I follow past literature and use IEG's project outcome rating. IEG rates each project after completion based on their respective outcomes (Independent Evaluation Group, 2016), using all project documents and, in many cases, interviews with project team members. To accompany and justify the ratings, IEG provides a Project Performance Assessment Report (PPAR). The latter largely mirrors project Implementation Completion Reports (ICRs) prepared by the respective project team members and their consultants. The key difference between the final IEG PPAR ratings and the ICR ratings is that those of PPAR tend to be lower, thereby providing an additional layer of protection against biased or inflated ratings (see also Dreher et al., 2013; Girod and Tobin, 2016).

The IEG project outcome ratings have six potential scores: highly unsatisfactory (1), unsatisfactory (2), moderately unsatisfactory (3), moderately satisfactory (4), satisfactory (5), and highly satisfactory (6). Higher-rated projects achieve their respective development objective efficiently and without little-to-no shortcomings, whereas lower-rated projects exhibit problems with efficiency and achievement of development objectives (Independent Evaluation Group, 2016, 1-2). With respect to the development objectives, all World Bank projects have a specific development objective, which project teams disaggregate into sub-indicators. Every six months in each project's life cycle, teams provide an update on the achievement of the development objective and the respective sub-indicators in a required Implementation Status Report (ISR). Accordingly, there is lots of detailed documentation that evaluators can assess when assigning the final project outcome ratings. Heinzl (2022) provides further details.

2.3. Identification Strategy

I use King, Lucas and Nielsen's (2017) frontier matching to identify the causal effect of GAAPs on IEG project outcomes scores. Frontier matching is preferable to other potential

methods because GAAP decisions are mostly project-specific and do not have a clear staggered adoption or panel structure. The only exceptions were Indonesia, India, and Nepal, all of which at one point decided to use GAAPs in all of their projects. However, at a later time, the countries reversed their decisions and no longer required their projects to have GAAPs. Consequently, alternative potential identification strategies, such as the generalized synthetic control method (Xu, 2017), the augmented synthetic control method (Ben-Michael, Feller and Rothstein, 2021), and panel matching (Imai, Kim and Wang, 2023), are not feasible for the present study.

Like any matching method, frontier matching finds projects with GAAPs and compares them to very similar projects without GAAPs. What distinguishes frontier matching from other matching methods is how it (re-)tests the causal effect of GAAPs. It does so by first pre-processing the data, finding the maximum balance for each observation in the sample. Next, it re-tests the causal effect of the GAAP treatment on project outcomes using the appropriate parametric techniques for each maximally balanced matched sample. Given that frontier matching allows researchers to see how the causal effects differ as the sample size increases and decreases, it enables analysts to discern how bias-variance tradeoffs and statistical power considerations may affect the results. By contrast, traditional matching methods, such as propensity score matching, which King and Nielsen (2019) argue have many issues, conduct one such analysis based on finding the one sample with the maximum balance and common support. Even other techniques that use matching as a pre-processing technique only conduct their analysis on one potential sample (e.g., Ho et al., 2007), so frontier matching represents a significant improvement over most matching methods. As King, Lucas and Nielsen (2017, 474) summarize, “no matching method can outperform the matching frontier, provided that both use the same imbalance metric.”

On the subject of the imbalance metric, I employ Iacus, King and Porro’s (2011) L_1 statistic primarily for reasons pertaining to the estimand (see Greifer and Stuart, 2022). King, Lucas and Nielsen (2017) also provide a method for obtaining the average Maha-

lanobis distance as an imbalance metric, but they only do so for the Feasible Sample Average Treatment Effect on the Treated (FSATT). The latter is not the estimand of interest in the present study: the FSATT would only examine the 210 projects with GAAPs and only compare them against the 210 best-matching projects without GAAPs, dropping control projects consecutively along the frontier. Limiting the sample as such would not only entail statistical power trade-offs but also external validity consequences given the inability of the smaller FSATT sample to correspond to a larger sample of World Bank development projects. That would be a problem, because the present study is attempting to estimate the effect on GAAPs in countries that are mostly like to need them. For these reasons, I focus on the Sample Average Treatment Effect on the Treated (SATT): that is, the effect of GAAPs on project outcomes across treated projects. Consistent with the frontier matching approach, I estimate the SATT at multiple points along the balance-sample size frontier. When doing so, I use fixed-ratio matching because Abadie and Spiess (2022) show that any type of post-matching inference involving replacement, including variable-ratio matching, produces problematic standard errors.

After matching, I first estimate the model dependence of the treatment effects using Athey and Imbens's (2015) robustness to model misspecification estimator. Then, I supplement those model dependence estimates using more traditional linear regression with robust standard errors as well as an ordinal multilevel model with country random effects. The ordinal multilevel model is likely superior not only due to the ordinal nature of the dependent variable data but also because Abadie and Spiess (2022) show that robust post-matching inference requires accounting for the clustering patterns in the data. On that score, the multilevel model generally outperforms standard "no pooling" or "complete pooling" models, including those with clustered standard errors. The reason is that the multilevel model does not treat the standard errors of similar observations from groups as a nuisance to correct but models that variation. More specifically, the multilevel estimates entail partial pooling by taking a weighted average of the mean of the unpooled model (i.e., for all projects) and the mean for the completely pooled model (i.e., for all countries) (Gelman and Hill, 2007, 253).

That partial pooling is particularly critical here given the different sample sizes of projects received for each country.

Mathematically, the multilevel ordered logit model takes the following form:

$$Pr(y_{IEG\ outcome\ rating(i,j)}^*) = \Lambda(\alpha_{country(j[i])} + \beta_{GAAP(i,j)} + \beta_{controls(i,j)}) \quad (1)$$

where subscripts i refer to the project and j to the implementing country; $\beta_{GAAP_{i,j}}$ corresponds to the primary effect of interest; $\beta_{controls_{i,j}}$ refer to the control variables; and $\alpha_{country(j[i])}$ is a random intercept that captures the (mostly) time-invariant country-level factors j for project i . Because the (latent) dependent variable, $y_{IEG\ outcome\ rating(i,j)}^*$, has six ordered categories, it is possible to classify $y_{IEG\ outcome\ rating(i,j)}^*$ in the following way, where τ_i are the cutpoints for each imposed category:

$$y_{IEG\ outcome\ rating(i,j)}^* = \begin{cases} 1 \text{ (highly unsatisfactory),} & \text{if } y_{outcome\ rating}^* \leq \tau_2 \\ 2 \text{ (unsatisfactory),} & \text{if } \tau_2 < y_{outcome\ rating}^* \leq \tau_3 \\ 3 \text{ (moderately unsatisfactory),} & \text{if } \tau_3 < y_{outcome\ rating}^* \leq \tau_4 \\ 4 \text{ (moderately satisfactory),} & \text{if } \tau_4 < y_{outcome\ rating}^* \leq \tau_5 \\ 5 \text{ (satisfactory),} & \text{if } \tau_5 < y_{outcome\ rating}^* \leq \tau_6 \\ 6 \text{ (highly satisfactory),} & \text{if } \tau_6 < y_{outcome\ rating}^* \end{cases} \quad (2)$$

2.4. Potential Control Variables

To discern the causal effects of GAAPs on project outcomes with any degree of confidence, it is necessary to take into account the potential determinants of higher and lower project outcome scores other than GAAPs. Following Denizer, Kaufmann and Kraay (2013), Kilby (2015), and Bulman, Kolkma and Kraay (2017), I control for both project- and country-level predictors. With respect to the project-level determinants, commitment amounts (project size) and preparation costs are particularly crucial: both the aid recipients and World Bank may be likely dedicate more supervision time and effort to projects that cost more to prepare and entail higher legitimacy in the event of poor outcomes. I also include

an indicator variable for additional or supplemental financing projects, because they tend to be for well-performing initial projects (Bulman, Kolkma and Kraay, 2017). Denizer, Kaufmann and Kraay (2013) further argue for the importance of mid-project risk flags, supervision costs, and effectiveness delays.²⁵ However, effectiveness delay information and mid-project risk flags are not publicly available, and in a follow-up study Bulman, Kolkma and Kraay (2017) find inconsistent effects of effectiveness delays. More significantly, though, I do not control for risk flags, effectiveness delays, and supervision costs, because doing so would clearly introduce post-treatment bias—or what Angrist and Pischke (2008) call “bad controls”.

At the country level, I follow Denizer, Kaufmann and Kraay (2013), Kilby (2015), and Bulman, Kolkma and Kraay (2017) by controlling for the (deflated) GDP growth rate, debt as a percent of GNI, and log GDP per capita from the World Bank’s (2017) World Development Indicators. To account for democracy, I include the Varieties of Democracy (V-Dem) polyarchy measure given its superiority over competing measures, such as Polity (see Vreeland, 2008). I also use V-Dem’s corruption measure given its superior performance relative to alternatives (McMann et al., 2022), such as those from the Worldwide Governance Indicators and Transparency International. Finally, following Honig (2019), I attempt to account for state fragility using Polity IV’s state fragility measure and the World Bank’s yearly classification of fragile states. Given the lack of available Country Policy and Institutional Assessment (CPIA) index data for the entire time period of study,²⁶ I use Henisz’s (2000) political constraints index to control for the effect of institutions on project outcomes.

2.5. Balance Assessments and Determining the Final Model

Determining the final set of covariates to include when using matching to pre-process the data and remove model dependence primarily relates to two factors. The first concerns the “conditional independence”, “unconfoundedness”, “selection on observables”, or “no

²⁵Effectiveness delays refer to starting project late after Board approval, which often happens when there are implementation challenges.

²⁶See Denly (2021) for more on available CPIA data.

omitted variable bias” assumption (e.g., Angrist and Pischke, 2008). For the case of the present study, treatment assignment to GAAPs must be independent of potential project outcomes given the final set of covariates. The second consideration concerns balance: that is, the similarity between treatment and control observations, which is what randomization attempts to achieve. In the present observational study, balance is a function of both the final set of chosen covariates and the number of observations included in the sample.

Given that the previous subsection deals with the conditional independence assumption by considering other potential causes of project outcomes, I now turn to assessing balance. As I show in Figure B.1b, adding each additional covariate entails a bias-variance trade-off with respect to the base specification of only including the treatment, GAAPs, as well as the following essential covariates from the base specification: commitments, preparation costs, GDP growth rates, and indicator variables for supplemental/additional financing projects as well as ones taking place in fragile states. With respect to the World Bank’s fragile state measure, I chose it over the Polity IV state fragility index measure, because the latter adds significant imbalance compared to World Bank’s own fragile state measure (see Figure B.1a). Additionally, the World Bank’s fragile state classification is more relevant and known by its bureaucrats.

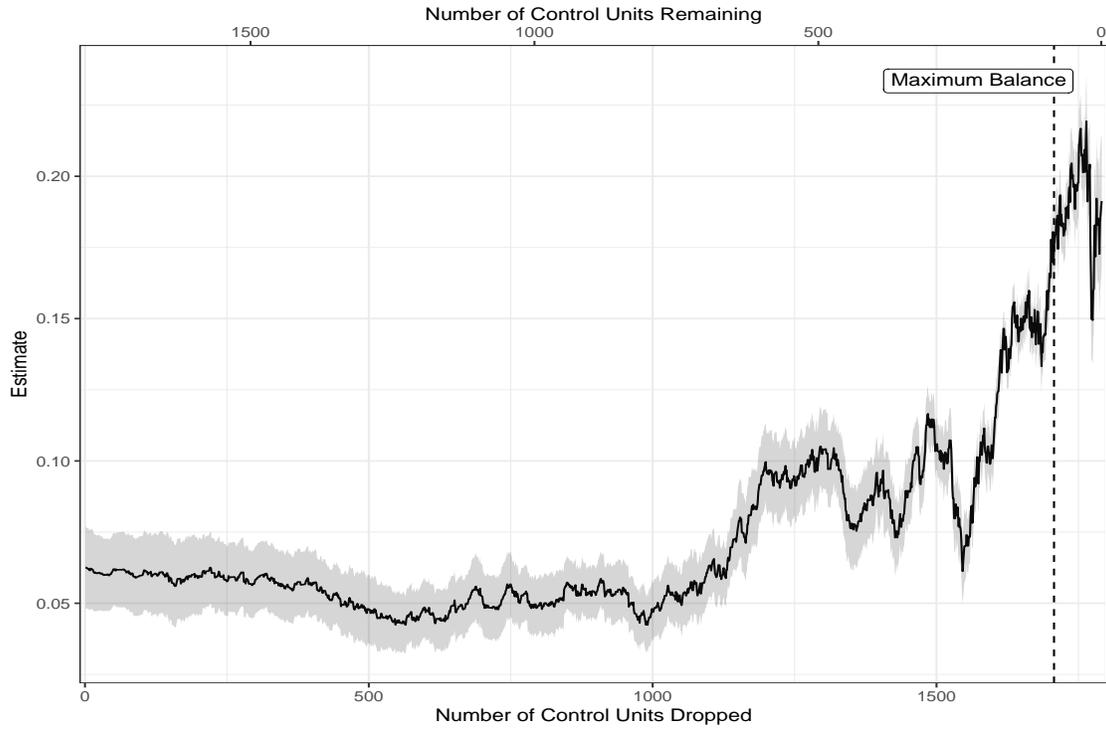
Regardless of the above balance trade-offs, Figure B.1b shows that the overall L_1 imbalance metric for the base specification is low, ranging from 0.45 (no control units removed) to 0.12 (most control units removed). Accordingly, there is no need to supplement analysis of the SATT with that of the FSATT using the average Mahalanobis distance. Such an analysis of the FSATT would only be necessary if it was impossible to obtain adequate matches using the larger dataset necessary to calculate the SATT. That is not the case here.

3. Post-Matching Results

Figures 7 and 8 presents the main results for the base specification after pre-processing the data via matching. Overall, GAAPs have a positive effect on IEG project outcomes.

Figure 7: Model Dependence and Linear Regression Results

(a) Athey and Imbens (2015) Model Dependence Intervals



(b) Linear Regression with Robust Standard Errors (95% Confidence Intervals)

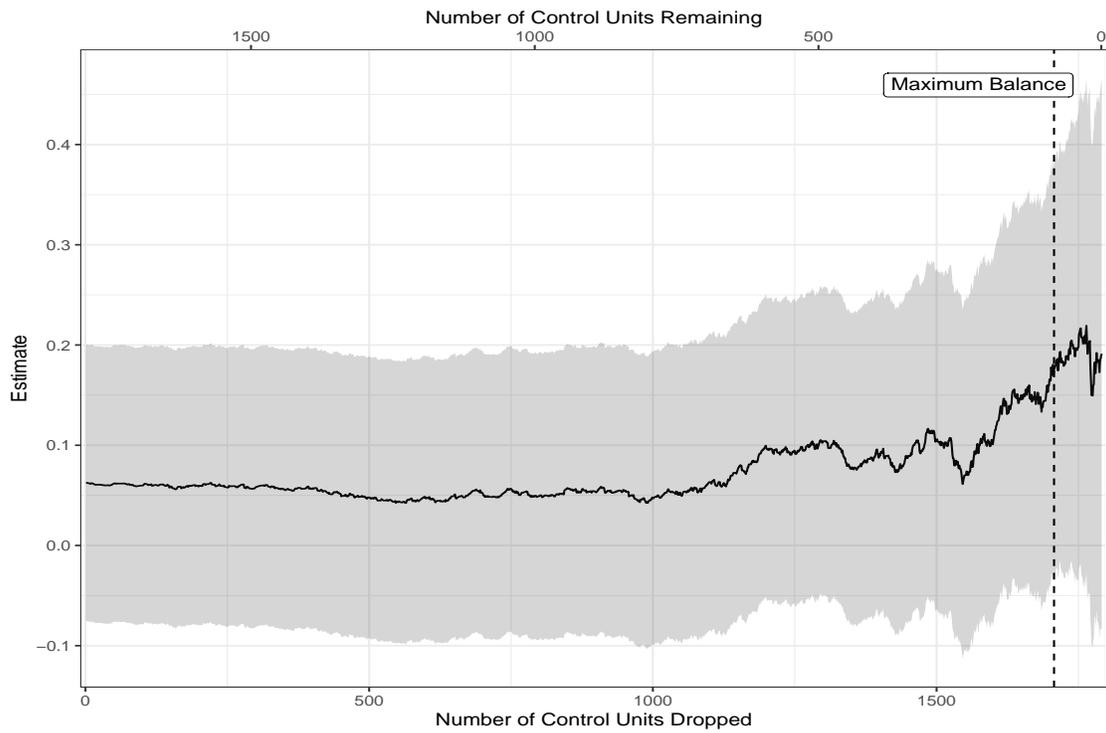
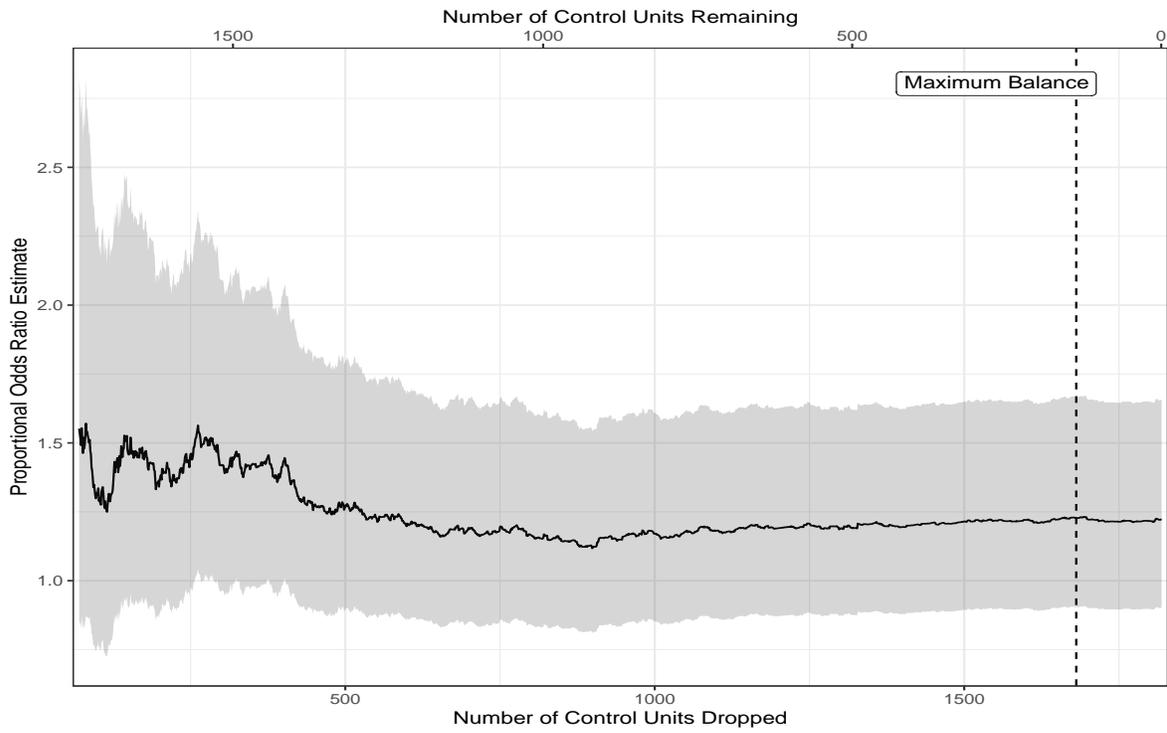


Figure 8: Base Specification of the Multilevel Ordered Logistic Regression

(a) Proportional Odds Ratios Estimates



As shown by Figure 7a, the Athey and Imbens (2015) intervals for robustness to model misspecification are quite narrow, suggesting that the treatment effect of GAAPs is precisely estimated. That is quite significant given that one of the primary challenges to any type of matching is that it is very difficult to satisfy the aforementioned unconfoundedness assumption.

The results from the linear regression model with 95% confidence intervals show positive treatment effects as well (see Figure 7). These estimates suggest that, on average, GAAP incidence increases project outcomes by 0.09 points (see Figure 7). When balance and internal validity are maximized, by dropping 1,681 control units, the effect of GAAPs on project outcomes increases slightly to 0.14. Considering that project outcomes are measured on a 1-6 scale, these treatment effect sizes from linear regression are quite high.

Using the likely more appropriate ordered multilevel logit model with a country-level random intercept does not alter the interpretation of the results. As Figure 8 showcases, the proportional odds ratio estimates range from 1.5 to 1.2. These estimates suggest that the use of a GAAP increases the odds of a project being in the highest outcome score category as compared to the lower five categories by 1.2-1.5 times.

4. External Validity

The present paper makes multiple efforts to improve the external validity of the estimates underpinning it. By design, frontier matching tests the sensitivity of the estimates due to balance and sample size considerations. Especially given the diversity of contexts arising from the larger number of projects that the World Bank finances, including different numbers of control units through the use of the frontier provides the estimates with a greater degree of generalizability. (More Coming Soon)

In terms of transportability, the author is unaware of any other aid financier that consistently carried anti-corruption actions plans across a large number of projects. By the

same token, given previous research on the difference between aid financiers (Isaksson and Kotsadam, 2018; Dreher et al., 2019), it is likely that the financier matters. Accordingly, the treatment effects of GAAPs are most likely to be applicable to financiers with strong procedures like those of the World Bank.

5. Conclusion

Numerous scholars cite Bauer’s (1972) aid paradox as a main reason for why foreign aid mostly cannot deliver the development outcomes that it promises (e.g., Easterly, 2007; Deaton, 2013). The fundamental idea behind the aid paradox is unequivocally correct: it is harder to engineer better development outcomes in poorer countries and ones with weaker institutions (e.g., Burnside and Dollar, 2000; Collier and Dollar, 2002). Note, however, that I use the word “harder”, not “impossible”, because the present study shows that context-specific project design efforts focused on institutional and corruption risks contribute to better development outcomes.

Primarily due to the endogenous relationships between development outcomes and anti-corruption measures, credible tests of my hypothesis previously proved elusive (see Gans-Morse et al., 2018). To overcome this challenge, I coded a new dataset of GAAPs used in World Bank projects and tested for potential endogenous relationships in the data. After I found no such endogenous relationships, I proceeded to test my hypothesis and found support for it in a number of models after pre-processing the data through frontier matching. Because GAAPs consist of extra top-down monitoring controls to address institutional weaknesses, the present study’s results challenge the recent wave of studies questioning the utility of top-down monitoring (e.g., Persson, Rothstein and Teorell, 2013). More broadly, the results suggest that aid bureaucrats have agency to mitigate the structural constraints imposed by weak institutions and corruption.

The results also have at least main four policy implications. First, there is value in rigorous quantitative assessment and waiting for the evaluation reports to be ready before

making policy decisions. I say this because the World Bank all but abandoned the use of GAAPs after 2013 (see Figure 4), which was well before a significant share of projects with GAAPS were completed and evaluated. Second, there may be less need for aid providers to continue lending selectivity exercise based on rankings like the World Bank's Country Policy and Institutional Assessments (CPIA) and its equivalents at the regional development banks.²⁷ Indeed, aid could reduce poverty to a much greater extent if it targeted countries that need it most (Collier and Dollar, 2002). Third, because the results only apply to GAAPs, which are context-specific, template-style monitoring efforts may be helpful but are likely insufficient to achieve better development outcomes. Essentially, bureaucrats need autonomy to best design their projects (Honig, 2018), but they also need top-down directives to ensure that they pay more attention to institutional and corruption risks. Fourth, and finally, the utility of GAAPs questions the development community's recent embrace of results-based financing, which in many cases obviates the need to rigorously monitor corruption (see Kenny, 2017).

²⁷For more the World Bank CPIA and its equivalent at regional development banks, see Denly (2021).

Appendices

A Coding Strategy	28
A.1 Coding of GAAPs	28
A.2 Coding of GAAP Attributes	29
B Matching Balance	30
Bibliography	33

A. Coding Strategy

A.1. Coding of GAAPs

As specified in Section 2.1, GAAPs only take place in investment projects, not structural adjustment/development policy or Program for Results loans. The differing structures of the appraisal documents containing the GAAPs presents one relevant challenge for the coding.²⁸ Most of the appraisal documents detailed the GAAPs in an annex. These cases tended to involve easy decisions to count the respective GAAP. It was more difficult to make an accurate assessment when the Task Teams decided to include the GAAPs outside of a dedicated annex. Generally, I tended not to count these instances, because they did not provide anti-corruption measures outside of the required measures that all projects must include.

I counted GAAPs that contained: (1) measures outside the scope of standard, required financial management, procurement, and demand-side governance controls; (2) governance and oversight arrangements that exceeded regular smart project design regarding internal controls and overlapping accountability structures; and/or (3) specified responsibilities for undertaking the relevant governance and anti-corruption measures. Some of the documents with sections labeled “anti-corruption action plan” or “governance strategy” did not meet the above criteria, so I excluded these projects from my count of GAAPs and governance/anti-corruption strategies.

The final types of governance/anti-corruption strategies that I excluded from my count are those included in the project risk frameworks, such as the ORAF or SORT. Such strategies are not formal governance/anti-corruption strategies and correspond more with overall risk management and the “GAC is everyone’s business” approach (see Kunicová, 2013) than additional anti-corruption measures consistent with GAAPs. Although I examined guarantee projects from the sample of potential projects with GAAPs, none of them employed either

²⁸The relevant appraisal document include: Project Appraisal Documents (PADs) (for normal projects), Project Papers (for additional financing loans), Technical Annexes (in case of Interim Strategy Note or part of a larger program), or Program Documents (in case of supplemental Investment Loans).

tool. Consequently, despite their presence in the IBRD/IDA (PE) product line, I removed guarantees from my sample before conducting the analysis.

The World Bank has occasionally added GAAPs to investment projects after Board approval. I did not code for such instances in my data set of GAAPs and governance/anti-corruption strategies. The lack of consistent data on such instances would have complicated the relevant coding, but endogeneity issues accounted for the primary reason behind my approach (see Section 2.1).

Finally, all projects in the sample

A.2. Coding of GAAP Attributes

As Figure 3 showcases, I code the following components:

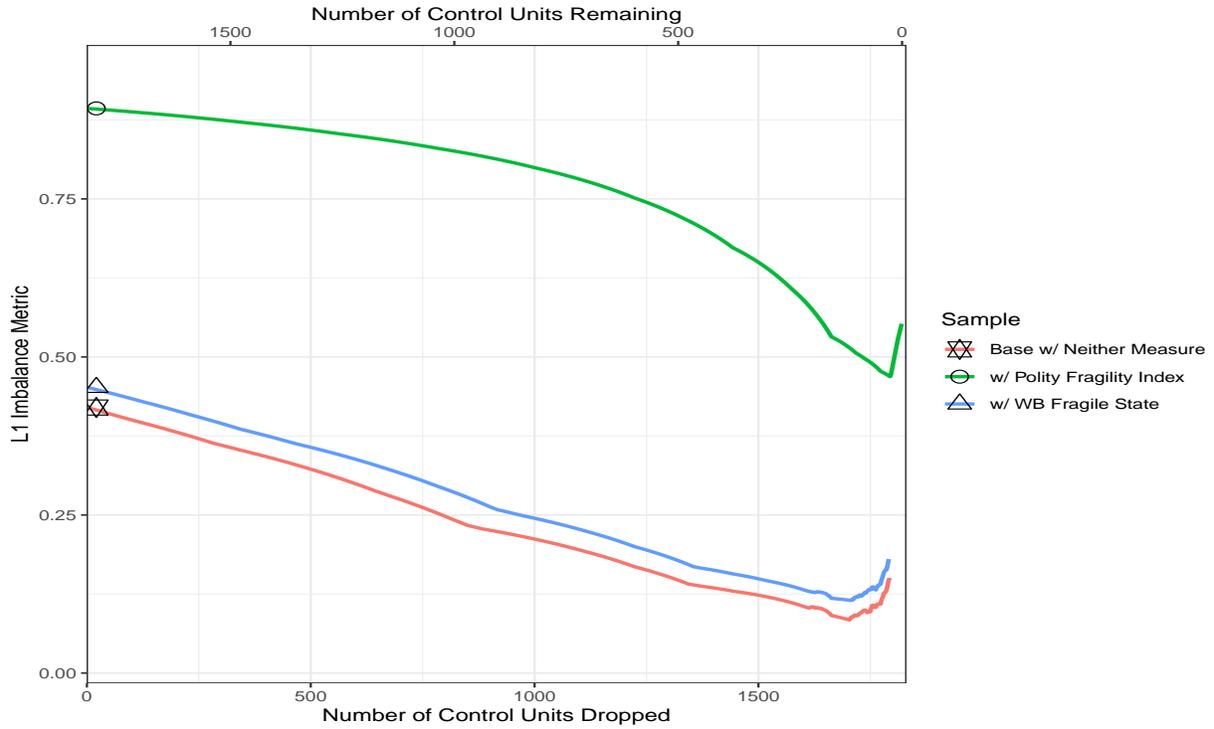
- internal audits: financial audit conducted by auditors within the same government implementing agency.
- external audit: financial audit conducted by either a different government agency or external firm.
- performance audit: audit designed to improve performance, as opposed to monitor it.
- technical audit: audits designed to measure the quality of infrastructure, such as in Olken (2007).
- procurement controls: measures to control corruption in procurement beyond the required ones that all projects use.
- internal oversight: additional oversight measures by employees of the relevant implementing agency of the government.
- other government oversight: additional oversight measures by a different government agency, such as an ombudsman.

- third-party oversight: additional oversight measures by a private company or different aid agency.
- community monitoring: monitoring by citizens/beneficiaries living near the implementation of the project, including score cards, report cards, social audits.
- clear implementor responsibility: the GAAP specifies the responsible actors necessary for completing the required actions.
- timetable: timetable for completing the attributes/actions.
- sanctions and remedies: extra project-level sanctions and remedies beyond those captured by the World Bank's sanctions and debarment framework.
- early warning indicators: indicators for further action in case of certain negative outcomes.

B. Matching Balance

Figure B.1: Balance-Sample Size Frontier Plots

(a) Frontier Plot for World Bank Fragile State/Polity State Fragility Index



(b) Frontier Plot for Other Variables

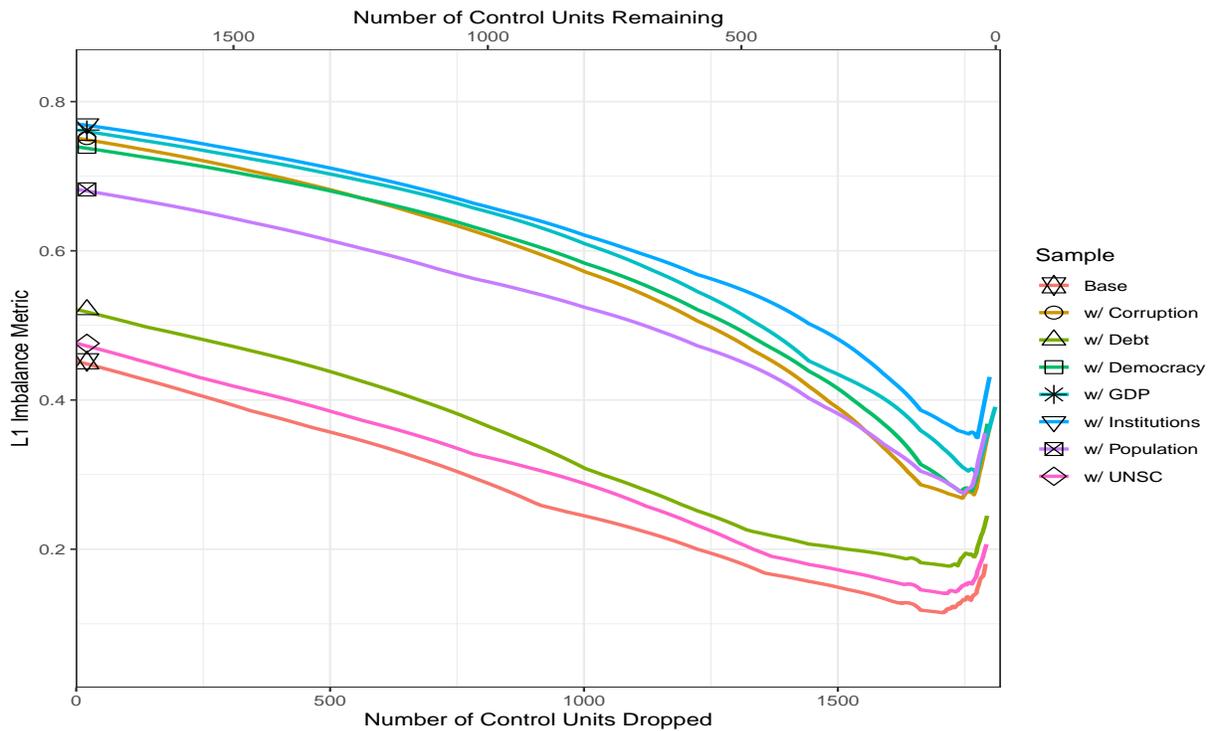
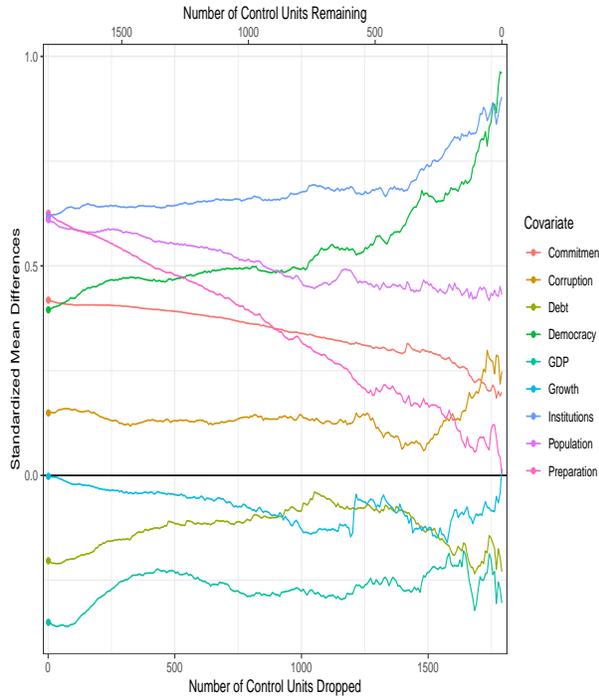
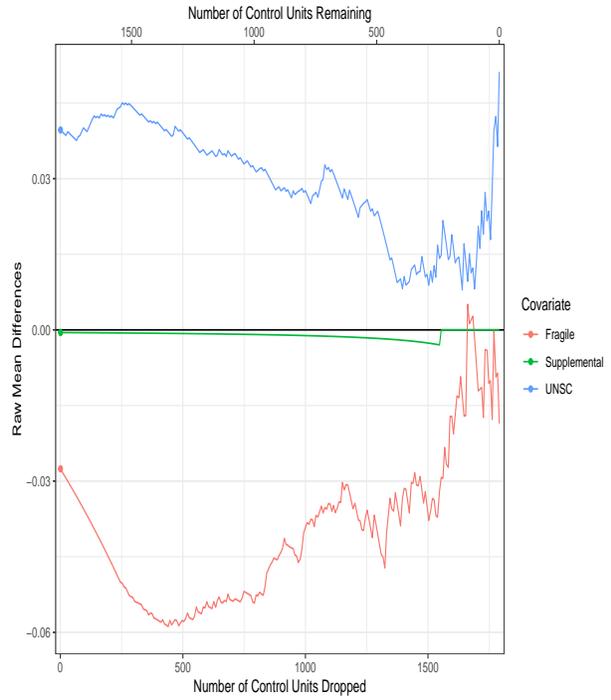


Figure B.2: Balance Diagnostics

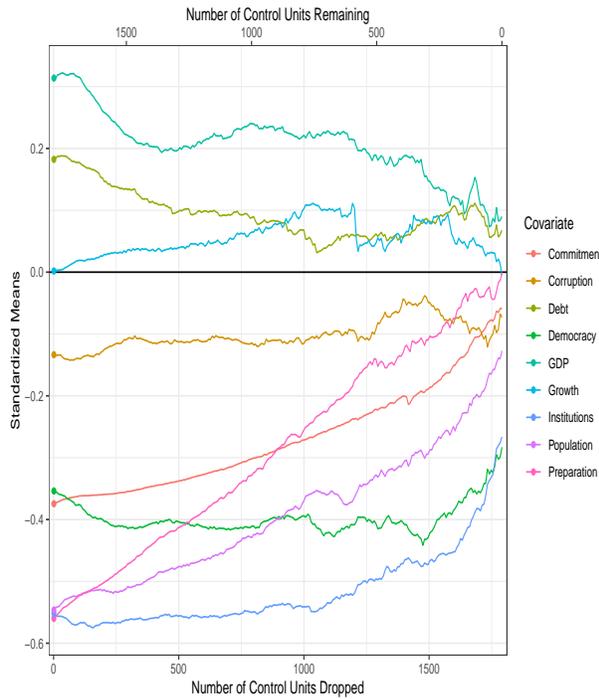
(a) Standardized Mean Differences



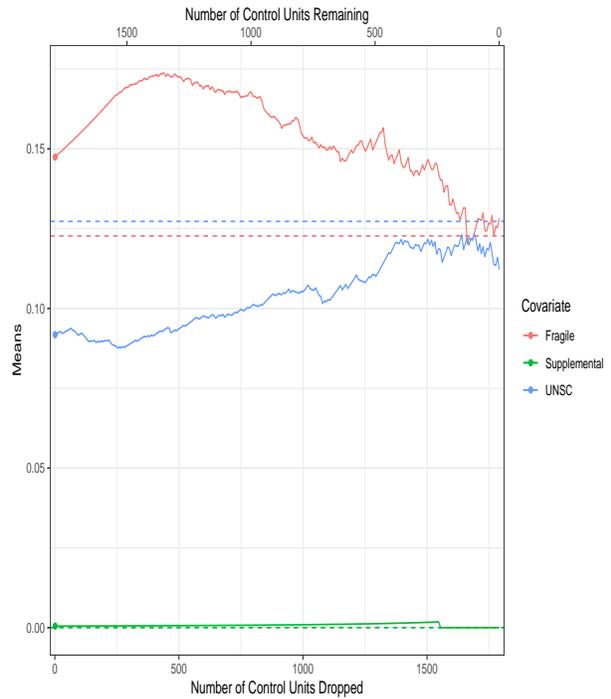
(b) Raw Mean Differences



(c) Standardized Means



(d) Raw Means



Bibliography

- Abadie, Alberto and Jann Spiess. 2022. “Robust Post-Matching Inference.” *Journal of the American Statistical Association* 117(538):983–995.
- Acemoglu, Daron and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. New York: Crown Business.
- African Development Bank. 2006. “Integrity and Anti-Corruption Department of the African Development Bank Group: Standard Operating Procedures.”
- Alesina, Alberto and Beatrice Weder. 2002. “Do Corrupt Governments Receive Less Foreign Aid?” *American Economic Review* 92(4):1126–1137.
- Altincekic, Ceren and David H. Bearce. 2014. “Why There Should Be No Political Foreign Aid Curse.” *World Development* 64(October 2012):18–32.
- Andersen, Jørgen Juel, Niels Johannesen and Bob Rijkers. 2022. “Elite Capture of Foreign Aid: Evidence from Offshore Bank Accounts.” *Journal of Political Economy* 130(2):388–425.
- Andrews, Matt, Lant Pritchett and Michael Woolcock. 2017. *Building State Capability: Evidence, Analysis, Action*. New Haven, Connecticut: Oxford University Press.
- Angrist, Joshua D. and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, New Jersey: Princeton University Press.
- Arndt, Channing, Sam Jones and Finn Tarp. 2015. “Assessing Foreign Aid’s Long-run Contribution to Growth and Development.” *World Development* 69:6–18.
- Arndt, Channing, Sam Jones and Finn Tarp. 2016. “What is the Aggregate Economic Rate of Return to Foreign Aid?” *World Bank Economic Review* 30(3):446–474.

- Ashton, Louise, Jed Friedman, Diana Goldemberg, Mustafa Zakir Hussain, Thomas Kenyon, Akib Khan and Mo Zhou. 2023. “A Puzzle with Missing Pieces: Explaining the Effectiveness of World Bank Development Projects.” *World Bank Research Observer* 38:115–146.
- Asian Development Bank. 1998. *Anticorruption Policy*. Manila: Asian Development Bank.
- Athey, Susan and Guido W. Imbens. 2015. “A Measure of Robustness to Misspecification.” *American Economic Review: Papers and Proceedings* 105(5):476–480.
- Banerjee, Abhijit V., Rukmini Banerji, Esther Duflo, Rachel Glennerster and Stuti Khemani. 2010. “Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India.” *American Economic Journal: Economic Policy* 2(1):1–30.
- Barnett, Michael N. and Martha Finnemore. 2004. *Rules for the World: International Organizations in Global Politics*. Ithaca, New York: Cornell University Press.
- Bauer, Peter T. 1972. *Dissent on Development*. Cambridge, Massachusetts: Harvard University Press.
- Bearce, David H. and Daniel C. Tirone. 2010. “Foreign Aid Effectiveness and the Strategic Goals of Donor Governments.” *Journal of Politics* 72(03):837–851.
- Ben-Michael, Eli, Avi Feller and Jesse Rothstein. 2021. “The Augmented Synthetic Control Method.” *Journal of the American Statistical Association* 116(536):1789–1803.
- Bermeo, Sarah Blodgett. 2016. “Aid Is Not Oil: Donor Preferences, Heterogeneous Aid, and the Aid-Democratization Relationship.” *International Organization* 70(1):1–32.
- Björkman, Martina, Damien de Walque and Jakob Svensson. 2017. “Experimental Evidence on the Long-Run Impact of Community-Based Monitoring.” *American Economic Journal: Applied Economics* 9(1):33–69.
- Björkman, Martina and Jakob Svensson. 2009. “Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda.” *Quarterly Journal of Economics* 124(2):735–769.

- Bobonis, Gustavo J., Luis R. Cámara Fuertes and Rainer Schwabe. 2016. “Monitoring Corruptible Politicians.” *American Economic Review* 106(8):2371–2405.
- Bourguignon, François and Jan Willem Gunning. 2020. Foreign Aid and Governance: A Survey. In *Handbook of Economic Development and Institutions*, ed. Jean-Marie Baland, François Bourguignon, Jean-Philippe Platteau and Thierry Verdier. Princeton, New Jersey: Princeton University Press chapter 9, pp. 308–355.
- Bueno de Mesquita, Bruce and Alastair Smith. 2009. “A Political Economy of Aid.” *International Organization* 63(2):309–340.
- Bulman, David, Walter Kolkma and Aart Kraay. 2017. “Good Countries or Good Projects? Macro and Micro Correlates of World Bank and Asian Development Bank Project Performance.” *Review of International Organizations* 12:335–363.
- Buntaine, Mark T. 2016. *Giving Aid Effectively: The Politics of Environmental Performance and Selectivity at Multilateral Development Banks*. Oxford: Oxford University Press.
- Burnside, Craig and David Dollar. 2000. “Aid, Policies, and Growth.” *American Economic Review* 90(4):847–868.
- Casey, Katherine. 2018. “Radical Decentralization: Does Community Driven Development Work?” *Annual Review of Economics* 10:139–163.
- Casey, Katherine, Rachel Glennerster and Edward Miguel. 2012. “Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan.” *Quarterly Journal of Economics* 127(4):1755–1812.
- Charron, Nicholas, Carl Dahlström, Mihály Fazekas and Victor Lapuente. 2017. “Careers, Connections, and Corruption Risks: Investigating the Impact of Bureaucratic Meritocracy on Public Procurement Processes.” *Journal of Politics* 79(1):89–104.

- Clausen, Bianca, Aart Kraay and Zsolt Nyiri. 2011. "Corruption and Confidence in Public Institutions: Evidence from a Global Survey." *World Bank Economic Review* 25(2):212–249.
- Collier, Paul and David Dollar. 2002. "Aid Allocation and Poverty Reduction." *European Economic Review* 46(8):1475–1500.
- Deaton, Angus. 2013. *The Great Escape: Health, Wealth, and the Origins of Inequality*. Princeton, New Jersey: Princeton University Press.
- Denizer, Cevdet, Daniel Kaufmann and Aart Kraay. 2013. "Good Countries or Good Projects? Macro and Micro Correlates of World Bank Project Performance." *Journal of Development Economics* 105(2013):288–302.
- Denly, Michael. 2021. Institutional Autonomy and Donor Strategic Interest in Multilateral Foreign Aid: Rules vs. Influence. In *Political Economy of International Organizations*.
- Dietrich, Simone. 2021. *States, Markets, and Foreign Aid*. Cambridge: Cambridge University Press.
- Dreher, Axel, Andreas Fuchs, Roland Hodler, Bradley C. Parks, Paul A. Raschky and Michael J. Tierney. 2019. "African Leaders and the Geography of China's Foreign Assistance." *Journal of Development Economics* 140(August 2018):44–71.
- Dreher, Axel, Stephan Klasen, James Raymond Vreeland and Eric Werker. 2013. "The Costs of Favoritism: Is Politically Driven Aid Less Effective?" *Economic Development and Cultural Change* 62(1):157–191.
- Dunning, Thad. 2004. "Conditioning the Effects of Aid: Cold War Politics, Donor Credibility, and Democracy in Africa." *International Organization* 58(2):409–423.
- Easterly, William. 2005. "What Did Structural Adjustment Adjust? The Association of Policies and Growth with Repeated IMF and World Bank Adjustment Loans." *Journal of Development Economics* 76(2005):1–22.

- Easterly, William. 2006. *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*. New York: Penguin Books.
- Easterly, William. 2007. "Was Development Assistance a Mistake?" *American Economic Review* 97(2):328–332.
- Easterly, William. 2015. *The Tyranny of Experts: Economists, Dictators, and the Forgotten Rights of the Poor*. New York: Basic Books.
- Easterly, William. 2019. "In Search of Reforms for Growth: New Stylized Facts on Policy and Growth Outcomes."
- Estevadeordal, Antoni and Alan M. Taylor. 2013. "Is the Washington Consensus Dead? Growth, Openness, and the Great Liberalization, 1970s-2000s." *Review of Economics and Statistics* 95(5):1669–1690.
- Ferraz, Claudio and Frederico Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics* 123(2):703–745.
- Feyzioglu, Tarhan, Vinaya Swaroop and Min Zhu. 1998. "A Panel Data Analysis of the Fungibility of Foreign Aid." *World Bank Economic Review* 12(1):29–58.
- Fisman, Raymond and Miriam A. Golden. 2017. *Corruption: What Everyone Needs to Know*. Oxford: Oxford University Press.
- Fox, Jonathan A. 2015. "Social Accountability: What Does the Evidence Really Say?" *World Development* 72:346–361.
- Fukuyama, Francis and Francesca Recanatini. 2021. Corruption, Elites, and Power: An Overview of International Policy Efforts to Improve the Quality of Government. In *Oxford Handbook of the Quality of Government*, ed. Andreas Bågenholm, Monika Bauhr, Marcia Grimes and Bo Rothstein. Oxford, UK: Oxford University Press chapter 22, pp. 471–494.

- Gans-Morse, Jordan, Mariana Borges, Alexey Makarin, Theresa Mannah-Blankson, Andre Nickow and Dong Zhang. 2018. “Reducing Bureaucratic Corruption: Interdisciplinary Perspectives on What Works.” *World Development* 105:171–188.
- Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multi-level/Hierarchical Modelling*. Cambridge: Cambridge University Press.
- Gingerich, Daniel W. 2013. “Yesterday’s Heroes, Today’s Villains: Ideology, Corruption, and Democratic Performance.” *Journal of Theoretical Politics* 26(2).
- Girod, Desha M. and Jennifer L. Tobin. 2016. “Take the Money and Run: The Determinants of Compliance with Aid Agreements.” *International Organization* 70(1):209–239.
- Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de Silanes and Andrei Shleifer. 2004. “Do Institutions Cause Growth?” *Journal of Economic Growth* 9(3):271–303.
- Greifer, Noah and Elizabeth A. Stuart. 2022. “Choosing the Estimand When Matching or Weighting in Observational Studies.”
- Grier, Kevin and Robin M. Grier. 2021. “The Washington Consensus Works: Causal Effects of Reform, 1970-2015.” *Journal of Comparative Economics* 49(1):59–72.
- Grindle, Merilee S. 2004. “Good Enough Governance: Poverty Reduction and Reform in Developing Countries.” *Governance* 17(4):525–548.
- Grossman, Sanford J. and Oliver Hart. 1986. “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration.” *Journal of Political Economy* 94(4):691–719.
- Gutner, Tamar. 2005. “World Bank Environmental Reform: Revisiting Lessons from Agency Theory.” *International Organization* 5(3):773–783.
- Hart, Oliver. 2017. “Incomplete Contracts and Control.” *American Economic Review* 107(7):1731–1752.

- Hawkins, Darren, David A. Lake, Daniel L. Nielson and Michael J. Tierney. 2006. Delegation under Anarchy: States, International Organizations, and Principal-Agent Theory. In *Delegation and Agency in International Organizations*. New York: Cambridge University Press chapter 1, pp. 3–38.
- Heinzel, Mirko. 2022. “International Bureaucrats and Organizational Performance. Country-Specific Knowledge and Sectoral Knowledge in World Bank Projects.” *International Studies Quarterly* 66(sqac013):1–13.
- Heldt, Eugénia C. and Henning Schmidtke. 2019. “Explaining Coherence in International Regime Complexes: How the World Bank Shapes the Field of Multilateral Development Finance.” *Review of International Political Economy* 26(6):1160–1186.
- Henisz, Witold J. 2000. “The Institutional Environment for Multinational Investment.” *Journal of Law, Economics, and Organization* 16(2):334–364.
- Ho, Daniel E., Kosuke Imai, Gary King and Elizabeth A. Stuart. 2007. “Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference.” *Political Analysis* 15(3):199–236.
- Honig, Dan. 2018. *Navigation by Judgment: Why and When Top Down Foreign Aid Management Doesn't Work*. Oxford: Oxford University Press.
- Honig, Dan. 2019. “When Reporting Undermines Performance: The Costs of Politically Constrained Organizational Autonomy in Foreign Aid Implementation.” *International Organization* 73(1):171–201.
- Huntington, Samuel. 1965. “Political Development and Political Decay.” *World Politics* 17(3):386–430.
- Iacus, Stefano M., Gary King and Giuseppe Porro. 2011. “Multivariate Matching Methods that Are Monotonic Imbalance Bounding.” *Journal of the American Statistical Association* 106(493):345–361.

- Imai, Kosuke, In Song Kim and Erik H. Wang. 2023. “Matching Methods for Causal Inference with Time-Series Cross-Sectional Data.” *American Journal of Political Science* .
- IMF. 1944. “Articles of Agreement of the International Monetary Fund.” .
- Independent Evaluation Group. 2016. Harmonized Criteria for ICR and OED Evaluations. Technical report Washington, DC: .
- Inter-American Development Bank. 2009. “IDB Moves to Strengthen Anti-Corruption Framework.” .
URL: <http://www.iadb.org/en/news/news-releases/2009-12-08/idb-moves-to-strengthen-anti-corruption-framework,6079.html>
- International Financial Institutions Anti-Corruption Task Force. 2006. Uniform Framework for Preventing and Combating Fraud and Corruption. Technical Report September.
- Isaksson, Ann Sofie and Andreas Kotsadam. 2018. “Chinese Aid and Local Corruption.” *Journal of Public Economics* 159:146–159.
- Islamic Development Bank. 2012. Islamic Development Bank Anti-Corruption Guidelines. Technical report Islamic Development Bank Jeddah, Saudi Arabia: .
- Johnson, Tana. 2011. “Guilt by Association: The Link between States’ Influence and the Legitimacy of Intergovernmental Organizations.” *Review of International Organizations* 6(1):57–84.
- Johnson, Tana. 2013. “Institutional Design and Bureaucrats’ Impact on Political Control.” *Journal of Politics* 75(01):183–197.
- Jones, Sam and Finn Tarp. 2016. “Does Foreign Aid Harm Political Institutions?” *Journal of Development Economics* 118:266–281.
- Joshi, Anuradha. 2013. “Do They Work? Assessing the Impact of Transparency and Accountability Initiatives in Service Delivery.” *Development Policy Review* 31(S1):29–48.

- Kapur, Devesh, John P Lewis and Richard Webb. 1997. *The World Bank: Its First Half Century (Volume 1: History)*. Vol. 1 Washington, DC: Brookings University Press.
- Kenny, Charles. 2017. *Results Not Receipts: Counting the Right Things in Aid and Corruption*. Washington, DC: Center for Global Development.
- Khemani, Stuti, Ernesto Dal Bó, Claudio Ferraz, Frederico Finan, Corinne Stephenson, Adesinaolo Odugbemi, Dikshya Thapa and Scott Abrahams. 2016. *Making Politics Work for Development: Harnessing Transparency and Citizen Engagement*. Washington, DC: World Bank.
- Kilby, Christopher. 2015. “Assessing the Impact of World Bank Preparation on Project Outcomes.” *Journal of Development Economics* 115:111–123.
- King, Gary, Christopher Lucas and Richard A. Nielsen. 2017. “The Balance-Sample Size Frontier in Matching Methods for Causal Inference.” *American Journal of Political Science* 61(2):473–489.
- King, Gary and Richard A. Nielsen. 2019. “Why Propensity Scores Should Not Be Used for Matching.” *Political Analysis* 27(4):435–454.
- Kunicová, Jana. 2013. The Role of the World Bank in Promoting Good Governance and Anti-Corruption Reforms. In *Anti-Corruption Policy: Can International Actors Play a Constructive Role?*, ed. Susan Rose-Ackerman and Paul D Carrington. Durham, NC: Duke University Press.
- Leff, Nathaniel. 1964. “Economic Development through Bureaucratic Corruption.” *American Behavioral Scientist* 8(3):8–14.
- Levy, Brian. 2014. *Working with the Grain: Integrating Governance and Growth in Development Strategies*. New York: Oxford University Press.

- Lewis-Faupel, Sean, Yusuf Neggers, Benjamin A. Olken and Rohini Pande. 2016. "Can Electronic Procurement Improve Infrastructure Provision? Evidence from Public Works in India and Indonesia." *American Economic Journal: Economic Policy* 8(3):258–283.
- Lipset, Seymour Martin. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *American Political Science Review* 53(1):69–105.
- López Claros, Augusto. 2015. "Removing Impediments to Sustainable Economic Development: The Case of Corruption." *Journal of International Commerce, Economics and Policy* 6(1):35.
- Manzetti, Luigi and Carole J. Wilson. 2007. "Why Do Corrupt Governments Maintain Public Support?" *Comparative Political Studies* 40(8):949–970.
- Martens, Bertin, Uwe Mummert, Peter Murrell and Paul Seabright. 2002. *The Institutional Economics of Foreign Aid*. Cambridge: Cambridge University Press.
- McMann, Kelly, Daniel Pemstein, Brigitte Seim, Jan Teorell and Staffan I. Lindberg. 2022. "Assessing Data Quality: An Approach and An Application." *Political Analysis* 30(3):426–449.
- Naím, Moisés. 1995. "The Corruption Eruption." *Brown Journal of World Affairs* 2(2):245–262.
- Newell, James L and Martin J Bull. 2003. Introduction. In *Corruption in Contemporary Politics*, ed. Martin J Bull and James L Newell. New York: Palgrave Macmillan chapter 1, pp. 1–11.
- Nielson, Daniel L. and Michael J. Tierney. 2005. "Theory, Data, and Hypothesis Testing: World Bank Environmental Reform Redux." *International Organization* 59(3):785–800.
- Nye, Joseph. 1967. "Corruption and Political Development: A Cost-Benefit Analysis." *American Political Science Review* 61(2):417–427.

- OECD. 2005. "Paris Declaration on Aid Effectiveness."
- Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115(21):200–249.
- Olken, Benjamin A. 2010. "Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia." *American Political Science Review* 104(2):243–267.
- Persson, Anna, Bo Rothstein and Jan Teorell. 2013. "Why Anticorruption Reforms Fail—Systemic Corruption as a Collective Action Problem." *Governance* 26(3):449–471.
- Reinikka, Ritva and Jakob Svensson. 2004. "Local Capture: Evidence from a Central Government Transfer Program in Uganda." *Quarterly Journal of Economics* 119(2):679–705.
- Reinikka, Ritva and Jakob Svensson. 2011. "The Power of Information in Public Services: Evidence from Education in Uganda." *Journal of Public Economics* 95(7-8):956–966.
- Rich, Bruce. 2002. The World Bank under James Wolfensohn. In *Reinventing the World Bank*, ed. Jonathan Pincus and Jeffrey A Winters. Ithaca and London: Cornell University Press pp. 26–53.
- Rodrik, Dani, Arvind Subramanian and Francesco Trebbi. 2004. "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development." *Journal of Economic Growth* 9(2):131–165.
- Rose-Ackerman, Susan and Paul D. Carrington. 2013. *Anti-Corruption Policy: Can International Actors Play a Constructive Role?* Durham, N.C.: Carolina Academic Press.
- Sharma, Patrick. 2013. "Bureaucratic Imperatives and Policy Outcomes: The Origins of World Bank Structural Adjustment Lending." *Review of International Political Economy* 20(4):667–686.
- Stone, Randall W. 2011. *Controlling Institutions: International Organizations and the Global Economy*. New York: Cambridge University Press.

- Svensson, Jakob. 2003. "Why Conditional Aid Does Not Work and What Can Be Done About It?" *Journal of Development Economics* 70(2003):381–402.
- Tierney, Michael J., Daniel L. Nielson, Darren Hawkins, J. Timmons Roberts, Michael G. Findley, Ryan M. Powers, Bradley C. Parks and Sven E. Wilson. 2011. "More Dollars than Sense: Refining Our Knowledge of Development Finance Using AidData." *World Development* 39(11):1891–1906.
- Touchton, Michael and Brian Wampler. 2013. "Improving Social Well-Being Through New Democratic Institutions." *Comparative Political Studies* 47(10):1442–1469.
- Trapnell, Stephanie and Francesca Recanatini. 2017. "Evidence, Corruption, and Reform: the Importance of Context-Sensitivity." *Crime, Law and Social Change* 68(4):477–491.
- US GAO. 2000. Report to Congressional Committees: World Bank, Management Controls Stronger, but Challenges in Fighting Corruption Remain. Technical Report April United States General Accounting Office Washington, DC: .
- van de Walle, Dominique and Ren Mu. 2007. "Fungibility and the Flypaper Effect of Project Aid: Micro-Evidence for Vietnam." *Journal of Development Economics* 84(2):667–685.
- Vreeland, James Raymond. 2008. "The Effect of Political Regime on Civil War: Unpacking Anocracy." *Journal of Conflict Resolution* 52(3):401–425.
- Weaver, Catherine E. 2007. "The World's Bank and the Bank's World." *Global Governance* 13(4):493–512.
- Weaver, Catherine E. and Stephen C. Nelson. 2016. Organizational Culture. In *Oxford Handbook of International Organizations*, ed. Jacob Katz Cogan, Ian Hurd and Ian Johnstone. Oxford: Oxford University Press chapter 43, pp. 920–942.
- Winters, Jeffrey. 1997. "Down with the World Bank." *Far Eastern Economic Review* p. 13.

- Winters, Matthew S. 2010. "Choosing to Target: What Types of Countries Get Different Types of World Bank Projects." *World Politics* 62(3):422–458.
- Wolfensohn, James D. 1996. "People and Development."
- World Bank. 1945. "International Bank for Reconstruction and Development Articles of Agreement."
- World Bank. 1997. *Helping Countries Combat Corruption: The Role of the World Bank*. Technical report World Bank Washington, DC: .
- World Bank. 2002. *The World Bank Annual Report*. Technical report World Bank Washington, DC: .
- World Bank. 2003. *Anti-Corruption Guide: Developing an Anti-Corruption Guide for Reducing Fiduciary Risk in New Projects - Lessons from Indonesia*. Technical report Jakarta, Indonesia: .
- World Bank. 2006. *Guidelines on Preventing and Combatting Fraud and Corruption in Projects Financed by IBRD Loans and IDA Credits and Grants*. Technical report World Bank Washington, DC: .
- World Bank. 2007. *Strengthening World Bank Group Engagement on Governance and Anticorruption*. Technical report World Bank Washington, DC: .
- World Bank. 2009. *Governance and Anticorruption in Lending Operations: A Benchmarking and Learning Review*. Technical report Washington, DC: .
- World Bank. 2010*a*. *Preparing Your Project Governance and Accountability Action Plan: Reducing Fiduciary Risk through Increased Transparency and Accountability - A Guidance Note for New Projects in India*. Technical report Washington, DC: .
- World Bank. 2010*b*. "World Bank Report Highlights Strong Year of Progress in Managing Corruption Risks."

URL: <http://www.worldbank.org/en/news/press-release/2010/09/08/world-bank-report-highlights-strong-year-of-progress-in-managing-corruption-risks>

World Bank. 2012. Strengthening Governance: Tackling Corruption – The World Bank Group’s Updated Strategy and Implementation Plan. Technical report World Bank Washington, DC: .

World Bank. 2013. Governance and Accountability Action Plans in the South Asia Region (SAR): Taking Stock and Moving Forward. Technical report Washington, DC: .

World Bank. 2017. “World Development Indicators.”.

URL: <http://data.worldbank.org/data-catalog/world-development-indicators>

Xu, Yiqing. 2017. “Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models.” *Political Analysis* 25(2009):57–76.

Zürn, Michael. 2004. “Global Governance and Legitimacy Problems.” *Government and Opposition* 18(1):99–109.