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

Michael Denly†

November 25, 2022

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

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

Abstract

Does anti-corruption improve aid effectiveness? Although existing literature is less than sanguine about the potential of anti-corruption, I argue that anti-corruption improves aid effectiveness for two main reasons. 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 change how they approached corruption. In turn, aid agencies created large anti-corruption infrastructures and approached corruption risks on a more context-specific basis, enabling anti-corruption measures to have both the power and specificity to meaningfully contribute to aid effectiveness. To test the hypothesis, I individually coded all 3,663 World Bank investment projects approved from 2001-2016 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 better project outcomes than similar projects without GAAPs. The results suggest that principal-agent style monitoring remains useful, and weak institutions also do not automatically yield poor development outcomes.

*For excellent research assistance, I thank Briana Rocha, Gavin Loyd, Hanna Herbowy, Nehaarika Shridha, Trey Durio, and Angelina Braese. For feedback, I thank Mike Findley, Desha Girod, Matt Winters, Alex Wais, Kyosuke Kikuta, Xin Nong, Jiseon Chang, Olgahan Cat, Matt Carnes, Stephan Eggli, Cleo O'Brien-Udry, Maria Vannari, and Eric Langenbacher. For advice, I thank Rich Nielsen, Steve Burgess, Arthur Alik Lagrange, Luis Cano, Joel Turkewitz, Anna Pinto Hebert, Lina Ibarra, Pascale Schnitzer, Yang Li, Anna Wilczynski, Charlie Undeland, Naseer Rana, and Adriana Moreno.

†Postdoctoral Research Fellow, Institute for Advanced Study in Toulouse, michael.denly@iast.fr

From 1943-2013, multilateral aid donors alone spent an estimated $US 7.8 trillion on foreign aid,[1] and a large portion of that colossal amount of money went to countries with weak institutions and high levels of corruption (Alesina and Weder, 2002). Given that weak institutions and corruption imperil the effectiveness of foreign aid (World Bank, 2013*a*), it is essential to know: Does anti-corruption improve aid effectiveness?

Most scholarship follows Burnside and Dollar (2000) and suggests that aid is unlikely to be effective in the presence of weak institutions and corruption pressures. On that score, the conventional wisdom highlights that institutional problems are both paramount and intractable (e.g., Acemoglu and Robinson, 2012); aid can be a "fungible" nontax revenue that politicians and bureaucrats capture for corrupt and political purposes (e.g., Bueno de Mesquita and Smith, 2009; Andersen, Johannesen and Rijkers, 2022); and aid distribution is politically-motivated and thus inefficient (e.g., Booth, 2011; Vreeland, 2019).

Although corruption and weak institutions pose real impediments to aid effectiveness, I argue that anti-corruption measures can overcome these impediments and improve aid effectiveness for two main reasons. First, anti-corruption measures mostly target institutional constraints, which a large literature suggests are the primary cause of development outcomes (e.g., Burnside and Dollar, 2000; Rodrik, Subramanian and Trebbi, 2004). 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 (Clausen, Kraay and Nyiri, 2011; Johnson, 2011), aid providers have invested in large anti-corruption infrastructures (Rose-Ackerman and Carrington, 2013). In the process, aid agencies have received increased political backing, added numerous staff members dedicated to anti-corruption, and developed relevant expertise to better exercise their "fiduciary duty".[2] Notably, MDB bureaucrats have worked with aid-receiving countries

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

[2]The "fiduciary duty" refers to the clause in the founding Articles of Agreement of all major MDBs, specifying

to include context-specific anti-corruption action plans at the project level. Because these context-specific action plans focus bureaucratic effort, harness significant knowledge, and address relevant institutional risks through additional audits and other oversight measures, I argue that aid can succeed in countries with weak institutions when bureaucrats design for success. Essentially, aid agency bureaucrats have agency to overcome the structural constraints posed by weak institutions and corruption.

To operationalize aid projects operating in weak institutions with corruption concerns, I individually coded all 3,663 World Bank investment projects approved from 2001-2016 for their use of context-specific, project-level Governance and Anti-Corruption Action Plans (GAAPs). The latter capture aid operating in weak institutions with corruption concerns, because the World Bank only uses GAAPs as an *additional* layer of top-down monitoring controls that supplement the analyses of institutional risks that all projects undertake. In total, 352 of the projects during the time period of utilized GAAPs for the period under study, making the share of projects with GAAPs around 9% 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 conduct two separate tests for such concerns and find no evidence to support them.

I use King, Lucas and Nielsen's (2017) frontier matching to test the causal effects of the GAAPs on the standard measure of World Bank project success in the literature: Independent Evaluation Group (IEG) project rating scores, which fall on a 1-6 scale. 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, GAAPs, using the maximum balance for each observation in the sample. That re-testing of estimates at the entire range

that aid funds need to be spent for their intended purposes (e.g., World Bank, 1945).

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, 2022).

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. Especially given that GAAPs are context-specific, top-down monitoring tools, the present study's first contribution concerns the efficacy of principal-agent style monitoring. To that end, aid financiers supervise recipients to ensure that the latter implement projects according to agreed upon specifications, but aid financiers are not true principals (Nielson and Tierney, 2005, 786).[3] Making aid financiers'

[3]Aid financiers 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).

ability to supervise institutional risks even weaker is that they very rarely cancel projects due to corruption, especially because it is harmful for bureaucrats careers' to stop project disbursements (Booth, 2011; Buntaine, 2016). Similarly, highly-cited work from Persson, Rothstein and Teorell (2013), Mungiu-Pippidi (2013), and Fisman and Golden (2017) suggests that anti-corruption approaches rooted in the principal-agent model mostly fail for a simple reason: they are unable to change the relatively immutable norms that underpin corruption. Despite these challenges, and contra recent work stressing that extra top-down monitoring/reporting requirements and a focus on receipts instead of results hurts development outcomes (Kenny, 2017; Honig, 2019), the evidence suggests that GAAPs work. The present study's more positive results on the effectiveness of monitoring schemes thus challenges most recent literature, which favors collective action approaches to anti-corruption over ones rooted in principal-agent style monitoring (e.g., Mungiu-Pippidi, 2013; Fisman and Golden, 2017).

Second, the present paper helps clarify the conditions under which bureaucrats have agency to overcome 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 shows that institutions are the strongest drivers of development outcomes, thereby implying that GAAP success is very unlikely. Another challenge comes from Easterly's (2006, 2015) critiques of planning in the aid industry and the purported failure of what Acemoglu and Robinson (2012) call "the ignorance hypothesis"—i.e., developing countries need Western advice to engineer prosperity. However, consistent with how the literature on Western advice has recently reversed course (e.g., Estevadeordal and Taylor, 2013; Easterly, 2019; Grier and Grier, 2021), the present study shows that institutions are not fully deterministic. That is especially the case given that GAAPs target weaker institutional environments. Accordingly, the results of the present study align with Denizer, Kaufmann and Kraay (2013) and Bulman, Kolkma and Kraay (2017), who find that at least 75% of development outcomes at the World Bank and Asian Development Bank relate to project-level features, as opposed to macro-level country characteristics. More broadly, bureaucrats can overcome at least some of the risks of weak institutional environments when they design measures to address those risks.

1. Aid, Institutions, and Legitimacy Costs

To understand why context-specific, top-down monitoring of institutional challenges in aid projects can improve development outcomes, it is first necessary to understand how ideas, focusing events, and policy interacted to change what became feasible for foreign aid. As will become clear below, feasibility primarily refers to legitimacy costs and the steps that aid financiers undertook to avoid them. Ultimately, those legitimacy costs yielded a critical juncture, forcing aid agencies to take very significant action on institutional risks to development outcomes.

1.1. The Early Years and Structural Adjustment

From aid's inception in the 1940s through the early 1970s, major financiers such as the United States and the World Bank mainly used aid to fund infrastructure (Kapur, Lewis and Webb, 1997). The decision to focus on infrastructure, as opposed to institutions, social sectors, or other ends, was also consistent with the predominant intellectual paradigm of the time: modernization theory. It notably stressed that economic growth would lead to democratization (e.g., Lipset, 1959), and institutional problems such as corruption were a nuisance that may even be beneficial for growth and development (Leff, 1964; Huntington, 1965; Nye, 1967).

The first noteworthy change regarding aid and institutions came in the 1980s, when the World Bank and International Monetary Fund (IMF) began structural adjustment lending.[4] The latter represented a foray into institutionally-targeted lending that some practitioners viewed as potentially conflicting with the World Bank's political prohibition clause in its

[4] 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).

Articles of Agreement.[5] Even if the advent of adjustment lending represented one of the first formal recognitions that institutions are not merely a nuisance but critical for achieving development outcomes, the mostly neoliberal content of adjustment lending belied such a take-away. As numerous scholars underscore, structural adjustment lending prioritized the private sector over building a robust, honest state (e.g., Stiglitz, 2002; Rodrik, 2006).

1.2. The Critical Juncture

The actual critical juncture that forced aid financiers to dedicate significant efforts to improving institutional weaknesses in recipient states started during the 1990s. During that decade, major corruption scandals rocked Italy, Brazil, Japan, Mexico, India, Spain, Pakistan, Ecuador, Georgia, Germany, Peru, the United States, South Korea, and Switzerland, resulting in the resignation or impeachment of heads of state and high-ranking officials (Naím, 1995; Newell and Bull, 2003; Manzetti and Wilson, 2007; López Claros, 2015; Fisman and Golden, 2017). Meanwhile, international organizations like the World Bank and IMF continued their structural adjustment programs, which notably forced many countries to stop their social welfare programs for the poor. The confluence of structural adjustment, greater popular awareness of corruption, and other globalization pressures resulted in massive protests and legitimacy challenges for international organizations around the world.

The infamous 1999 "Battle in Seattle", in which massive protests turned violent over a World Trade Organization meeting, was perhaps the most visible manifestation of popular discontent with international organizations, but it was far from the only such occasion (Zürn, 2004). For example, in 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). The legitimacy challenges from corruption, in particular, became even more acute in 1997, when the political economist Jeffrey Winters held a press conference to detail his findings on World Bank lending in Indonesia. In response to Win-

[5]See Article IV, Section 10 of World Bank (1945). It specifically prohibits the World Bank from engaging in political affairs of its aid-receiving countries.

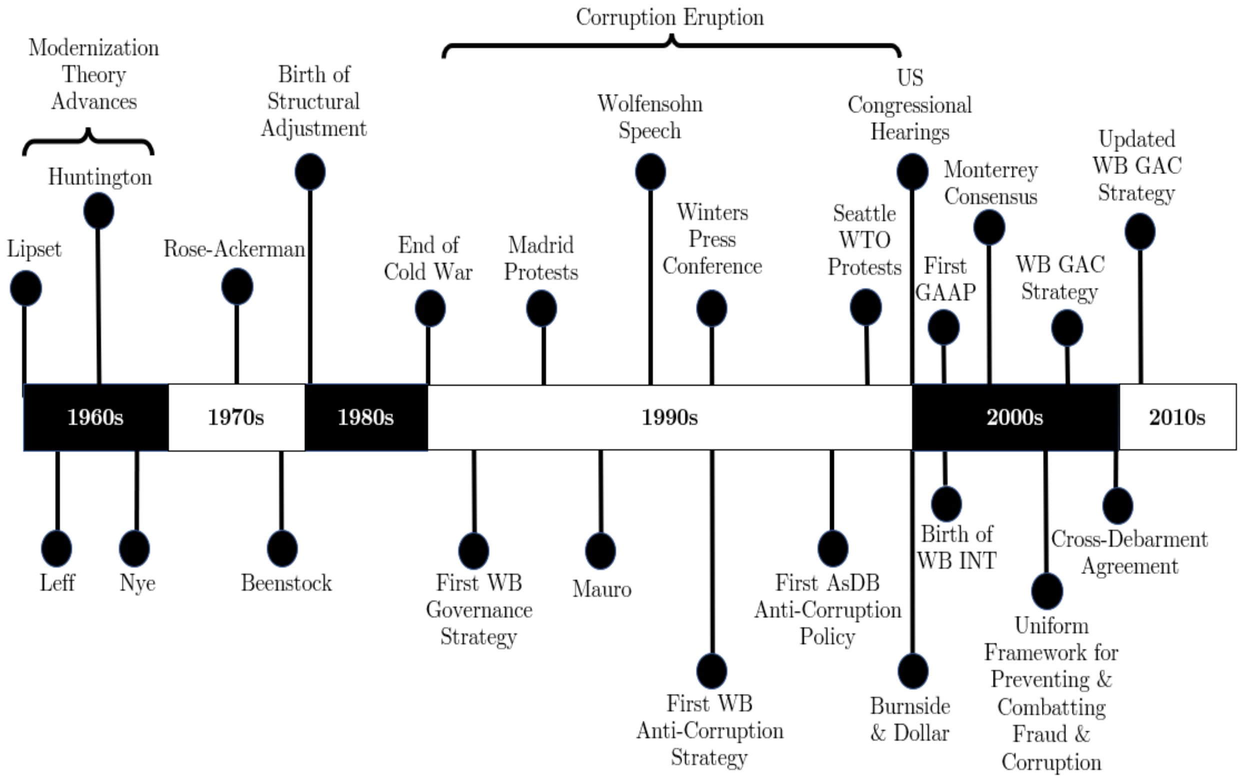
ters's (1997) allegation that more than $USD 10 billion of World Bank funds disappeared due to corruption, at least 126 non-governmental organizations from 35 countries wrote the institution in protest (Rich, 2002, 49-50). Subsequently, the United States invoked congressional hearings on the World Bank's ability to control corruption (US GAO, 2000). The legitimacy challenges for both donors and the agencies were well-founded, too, according to former World Bank president James Wolfensohn's (2010) autobiography:

> "I was perplexed that corruption seemed to be nowhere on the Bank's official agenda. . . I checked past speeches from annual meetings; I checked the minutes of the Development Committee...The term corruption simply did not appear. The only time I ever heard the subject discussed was in private gatherings in people's homes. There was a wall of silence surrounding this critical issue.
>
> 'What is going on here' I kept asking my staff. Finally, Ibrahim Shihata, the Bank's experienced and intelligent general counsel. . . took me into the hall outside his office. Looking over his shoulder, as if someone might hear, he warned that in the Bank, there was no room to discuss the 'C-word'. 'It would be offensive to our shareholders and risk political repercussions.' Attacking corruption, he made me understand, would insult some of the executive directors who represented countries where corruption had reached the highest levels. It would also insult some of the rich countries that were well-aware of the problem but used it to their advantage" (Wolfensohn, 2010, 294-295).

To change the institution's trajectory on corruption and attempt to overcome the legitimacy challenges, in 1996 Wolfensohn delivered a famous speech in which he decried the "cancer of corruption" and framed it as a development issue (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 (World Bank, 2003); 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),

Figure 1: Key Events and Scholarship Underpinning Aid Financiers' Reversal on Institutions



the Asian Development Bank (1998), African Development Bank (2006), and Inter-American Development Bank (2009), among other multilateral aid agencies,[6] 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*).

1.3. The Role of Scholarship and "Guilt by Association"

Although powerful powerful donor countries benefited from corruption in recipient countries (Wolfensohn, 2010, 294-295), politics underpins multilateral aid flows (Vreeland, 2019), aid agencies need to disburse money to survive (Booth, 2011), and "international organi-

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

zations are frequently pawns of developed states" (Lake and McCubbins, 2006, 341), two factors enabled the above pivot on corruption and institutions more broadly. The first factor relates back to ideas and, more specifically, those of Burnside and Dollar (2000). Building off insights from a second generation of scholars who used stronger methods and empirical data to significantly challenge the claims of the aforementioned early modernization theorists (Rose-Ackerman, 1978; Beenstock, 1979; Mauro, 1995; Wei, 2000; Acemoglu, Johnson and Robinson, 2001), Burnside and Dollar (2000) helped change the development landscape. Even though the study received numerous challenges (e.g., Easterly, Levine and Roodman, 2004), Burnside and Dollar's (2000, 2004) conclusion that aid only spurred growth in countries with strong institutional environments reverberated around policy and lay audiences. The large diffusion of the results—as well as related ones from Collier and Dollar (2002) and Kosack (2003)—supported efforts by the World Bank (1992, 1997), Asian Development Bank (1998), and other development financiers as they implemented revised mandates on governance and corruption. After all, if aid only works in countries with strong institutions, and institutions are a primary cause of development outcomes (Rodrik, Subramanian and Trebbi, 2004; Acemoglu and Robinson, 2012), then it is necessary to address those institutional weaknesses.

The second factor underpinning aid financiers' wholescale change regarding their willingness to directly address institutional weaknesses in recipient countries regards the significant evidence of what Johnson (2011) calls "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 donors have an incentive to qwell such concerns. That is especially true since the end of the Cold War and subsequent fracturing of the liberal international order (Bearce and Jolliff Scott, 2019), particularly given the competition from China and other "new" donors (Hernandez, 2017).

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 which, as detailed above, matter to both powerful donor countries and individual aid financiers. However, greater anti-corruption infrastructures, remits, and even anti-corruption knowledge are insufficient to prevent legitimacy costs on their own. The reason is that corruption, aid effectiveness, and institutional issues are context-specific (Grindle, 2004; Gingerich, 2013). To achieve better aid outcomes at the project-level and avoid more legitimacy costs, it is thus necessary to mitigate the risks of aid fungibility, elite capture, and their resulting legitimacy costs with context-specific measures. 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).

A robust literature on the effectiveness of anti-corruption tools suggests that it is possible to control corruption in weak institutional environments by enacting specific measures. 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 important 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 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). By the same token, the 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 (Fukuyama and Recanatini, 2021, 472-473).

One challenge that previous literature has not sufficiently considered, however, is the competing demands on bureaucrats' time. As the "mission creep" of aid agencies has increased over time to cover gender mainstreaming, climate change co-benefits, and many other aspects initially outside of aid agencies' purviews (see Einhorn, 2001; Weaver, 2008), there are more competing demands on aid project leaders' time. That is particularly significant given that individual project leaders have significant impacts on aid projects outcomes (Denizer, Kaufmann and Kraay, 2013; Bulman, Kolkma and Kraay, 2017). Accordingly, the greater infrastructures and knowledge to control corruption are likely insufficient to control legitimacy costs and the challenges of weak institutional environments unless project leaders have a commitment device to focus their attention. For these reasons, I hypothesize:

> **H1**: Aid projects in weak institutional environments will be most likely to succeed when they contain an action plan to address those institutional risks.

3. Research Design

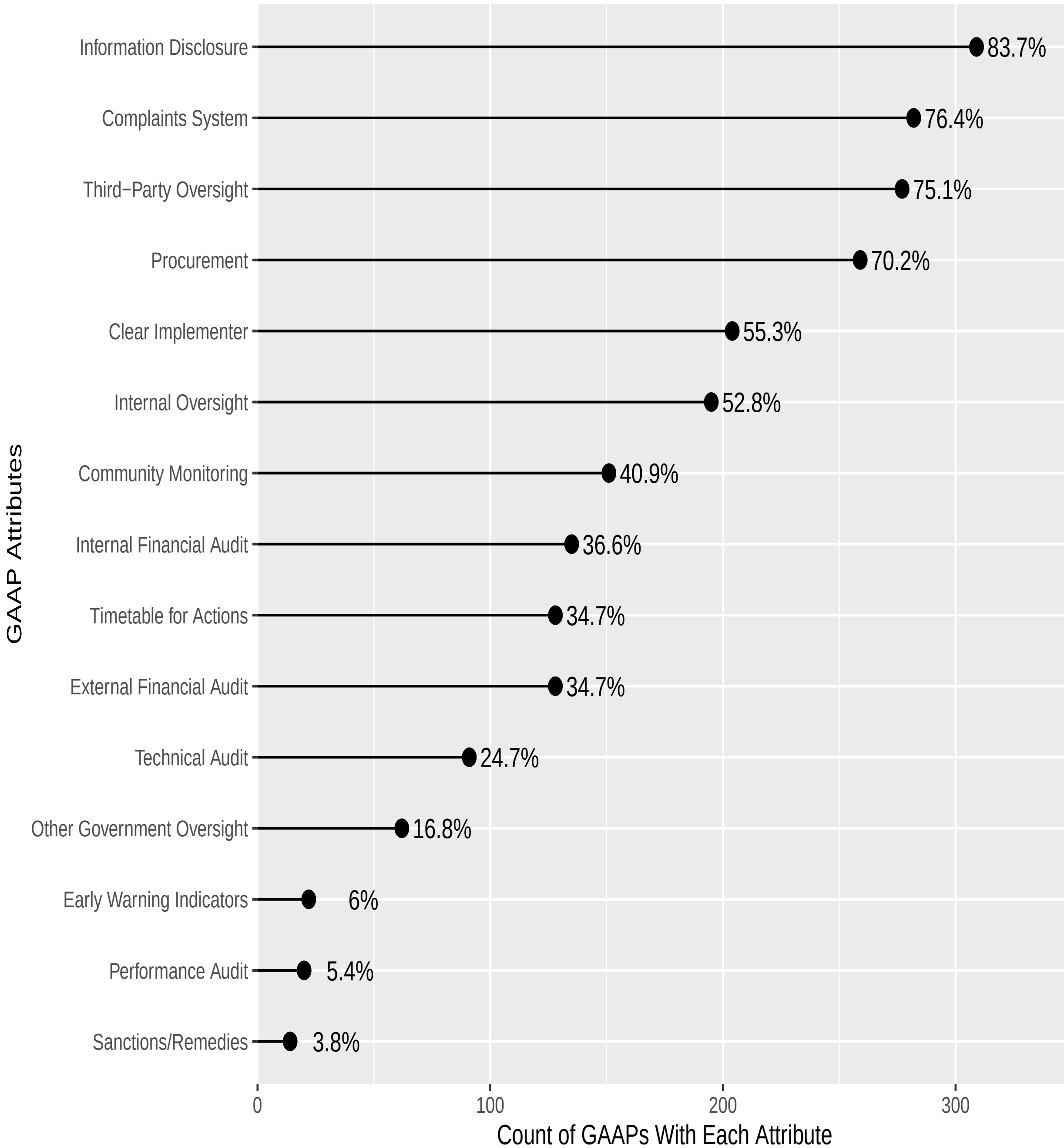
3.1. Governance and Anti-Corruption Action Plans (GAAPs)

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,[7] GAAPs comprise a project-specific mix of anti-corruption tools, such as audits, procurement controls, and social accountability measures that *supplement* existing project-level requirements.[8] Figure 2 provides a detailed breakdown of the attributes of GAAPs, which

[7] See Appendix A for more details.

[8] In 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 (see Figure 3), the present study does not focus on country-level action plans.

Figure 2: Governance and Anti-Corruption Plans Attributes (2001-2016)



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. In any case, 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.

the World Bank only uses in investment lending, not more fungible budget support, including Program for Results and structural adjustment loans.[9] As Figure 2 underscores, 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 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.[10] 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 3b). 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.

Indonesia, India, Bangladesh, and Nepal have received the most GAAPs (see Figure 3a), 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 2001-2016. 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 2001-2016 period. In total, 59 countries used these tools for the study period, encompassing around 9% of approved projects.[11] Based on

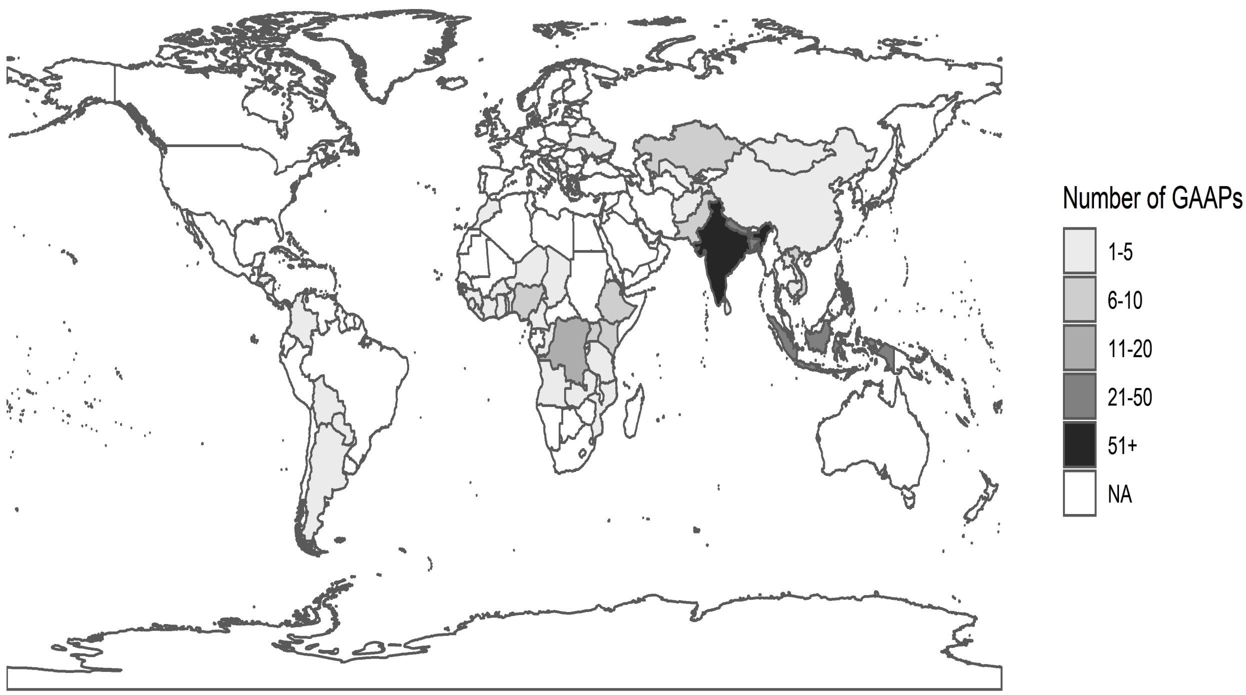
[9] Structural Adjustment Lending (SAL), Development Policy Lending (DPL), and Development Policy Financing (DPF) are all equivalent. The World Bank currently refers these instruments as DPFs. In any case, the World Bank does not use GAAPs for budget support. Accordingly, concerns relating to loan type 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.

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

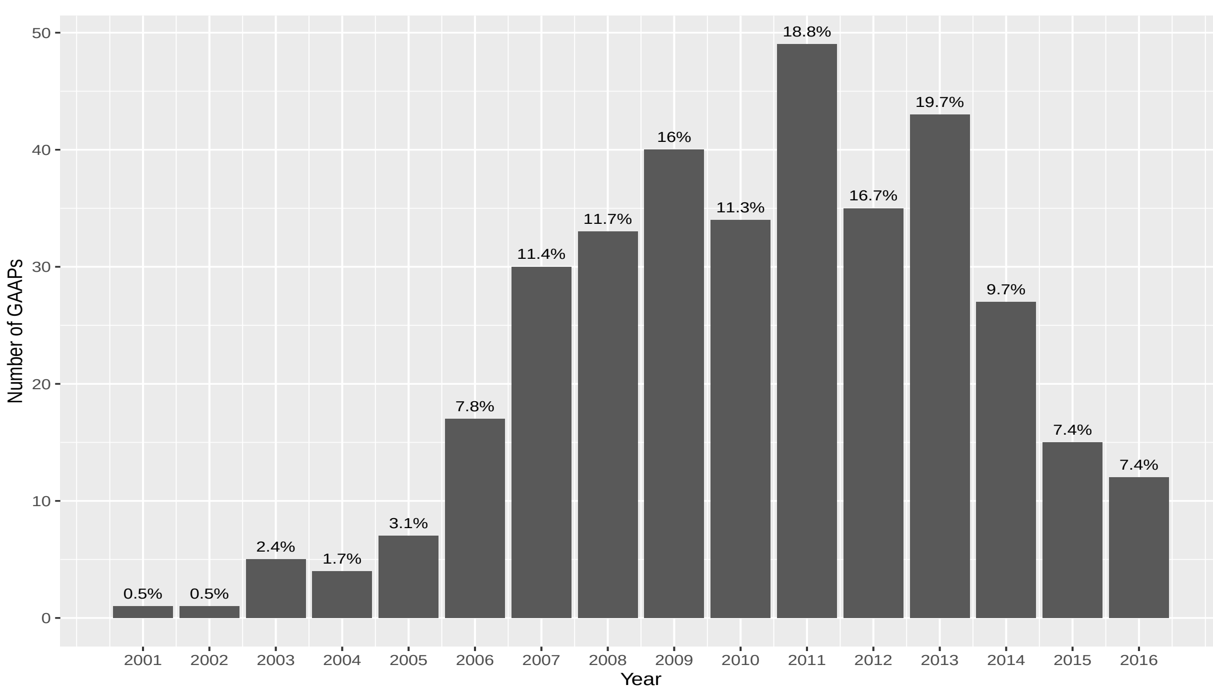
[11] See Appendix A for more details on the coding strategy.

Figure 3: World Bank's Usage of Governance and Anti-Corruption Plans (2001-2016)

(a) Country Usage



(b) Frequency by Year



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

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*b*).

The present study only considers GAAPs incorporated into the design of a project prior 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 the two measures correlate only at -0.03, it is difficult to argue that endogeneity or selection concerns characterize the use of GAAPs.

3.2. 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.

3.3. 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, 2022), 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 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 Mahalanobis 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 the 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:

$$Pr(y^*_{outcome\ rating(i,j)}) = \wedge\left(\alpha_{country(j[i])} + \beta_{GAAP_{(i,j)}} + \beta_{controls_{(i,j)}}\right) \tag{1}$$

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, $y^*_{outcome\ rating(i,j)}$, has six ordered categories, it is possible to classify $y^*_{outcome\ rating(i,j)}$ in the following way, where τ_i are the cutpoints for each imposed category:

$$y_{outcome\ rating(i,j)} = \begin{cases} 1, & if \quad y^*_{outcome\ rating} \leq \tau_2 \\ 2, & if \quad \tau_2 < y^*_{outcome\ rating} \leq \tau_3 \\ 3, & if \quad \tau_3 < y^*_{outcome\ rating} \leq \tau_4 \\ 4, & if \quad \tau_4 < y^*_{outcome\ rating} \leq \tau_5 \\ 5, & if \quad \tau_5 < y^*_{outcome\ rating} \leq \tau_6 \\ 6, & if \quad \tau_6 < y^*_{outcome\ rating} \end{cases} \tag{2}$$

3.4. 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.[12] 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) control for for the (deflated) GDP growth rate, debt as a

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

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,[13] I use Henisz's (2000) political constraints index to control for the effect of institutions on project outcomes.

3.5. 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 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

[13]See Denly (2021) for more on available CPIA data.

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 Mahalonobis 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.

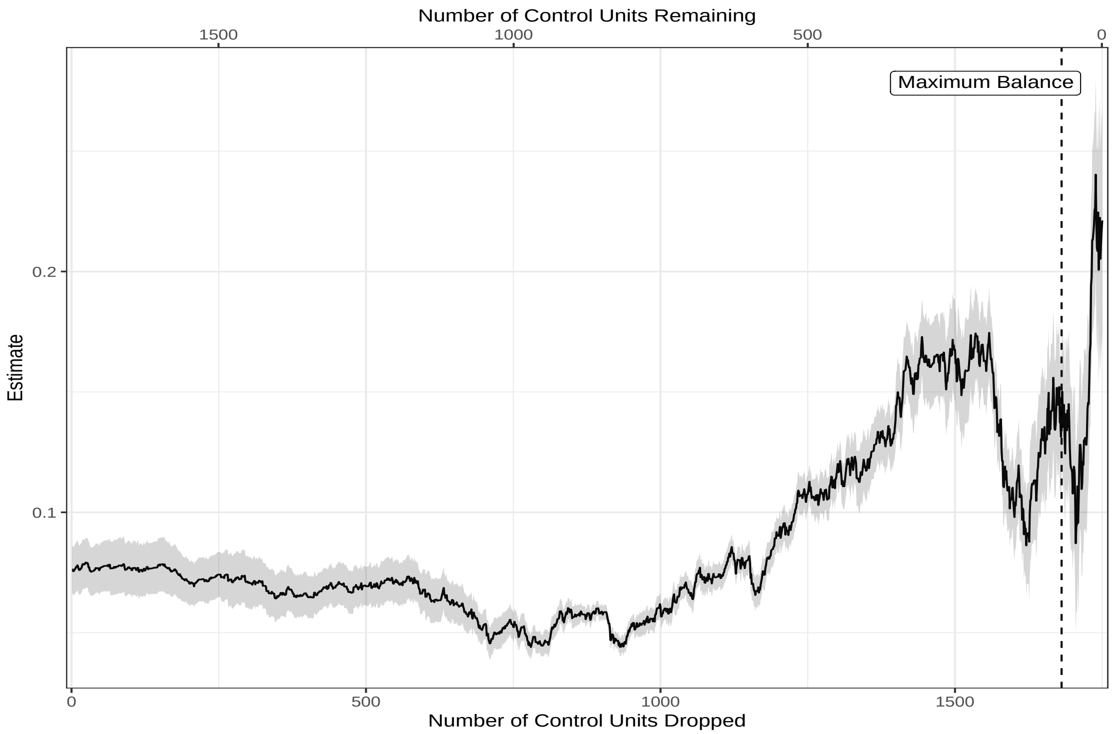
4. Post-Matching Results

Figures 4 and 5 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. As shown by Figure 4a, 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 4). These estimates suggest that, on average, GAAP incidence increases project outcomes by 0.09 points (see Figure 4). 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.

Figure 4: Model Dependence and Linear Regression Results

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



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

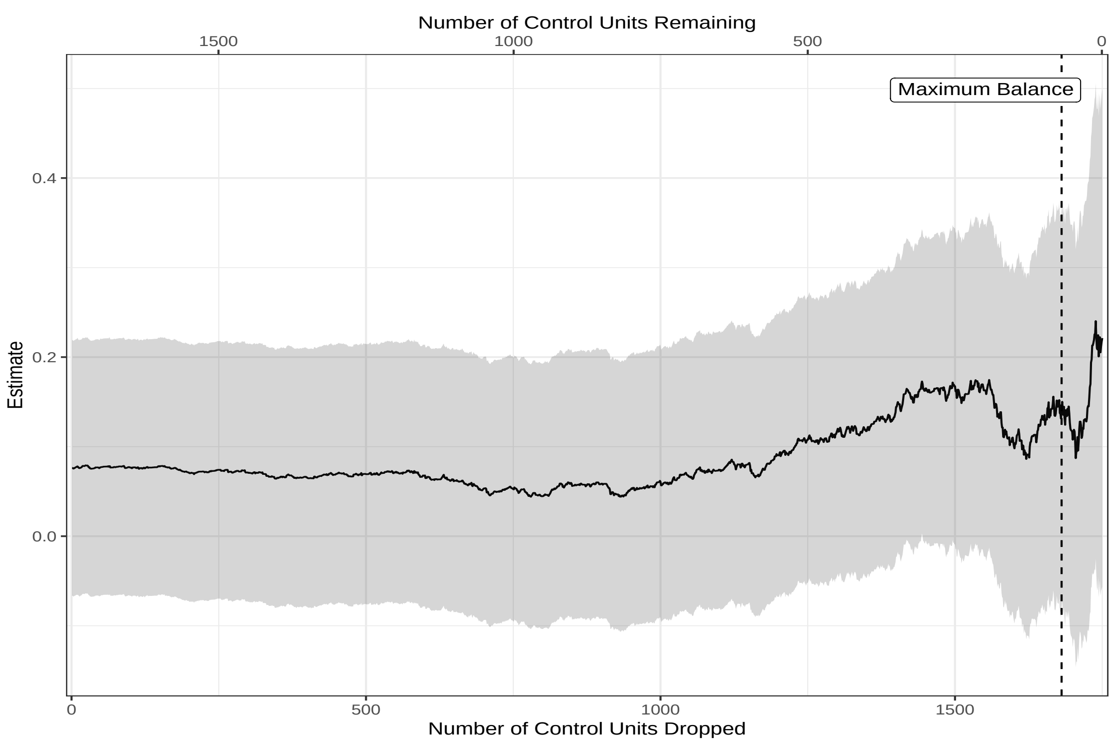
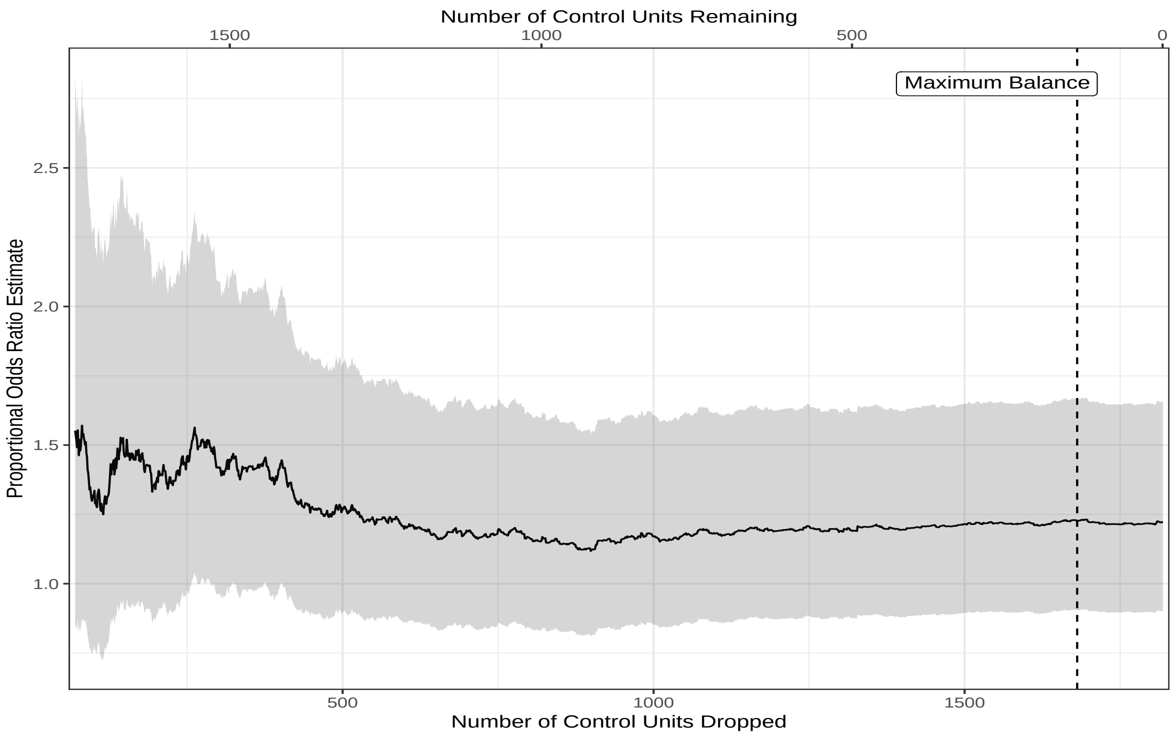


Figure 5: Base Specification of the Multilevel Ordered Logistic Regression

(a) Proportional Odds Ratios Estimates



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 5 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.

5. Robustness

A potential concern with the above results concerns the violation of the Stable Unit Treatment Value Assumption (SUTVA) (see Gerber and Green, 2012, Chapter 8). While the context-specific nature of World Bank projects mitigates SUTVA violations concerning spillovers, the SUTVA challenge with GAAPs pertains to the fact that they are never uniform due to the sundry institutional and corruption challenges in each project. By extension, there are multiple versions of the GAAP treatments, meaning that potential project outcomes under each version of the treatment are not precisely defined. Although having multiple versions of the treatment violates the base case of SUTVA (see Neyman, 1935), VanderWeele and Hernán (2013) show that such an issue is not an inferential problem under two scenarios: (1) when there is only one version of the control; and (2) when the treatment version does not precede the treatment itself. In the present study, there is only one version of the control—i.e., projects without GAAPs. Similarly, treatment assignment precedes the version of it, because World Bank bureaucrats decide to assign a GAAP to a project before they choose on the specifics of the GAAP.

In any case, I use the data referenced in Figure 2 to assess whether the treatment effects differ when I restrict the analysis to "strong" versions of the GAAP treatment. From the perspective of SUTVA, the advantage of strong GAAPs *vis-à-vis* regular ones is that strong GAAPs have more well-defined potential outcomes. Consequently, strong GAAPs are less likely to yield violations of the base case of SUTVA.

Even though it is difficult to classify any area of focus of GAAPs, such as audits, procurement, or community monitoring, as inherently stronger than another, some GAAP attributes facilitate classification into stronger GAAPs. Notably, some GAAPs have attributes that assist with their implementation, thereby making them arguably stronger treatments. Along those lines, it is possible to argue that GAAPs clearly specifying implementation responsibilities and timetables for task completion are the strongest.

[Coming soon: analysis of strong GAAPs]

6. Analysis of the Mechanism

[Coming soon: analysis of whether GAAPs predict better Task Team Leader (TTL) Quality (when I finish cleaning the data). The reasoning here is that GAAPs provide some mechanism to keep project leaders' attention on the relevant measures. Thus, GAAPs should be highly correlated with better TTL quality.]

7. 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. Beyond these unit- and settings-based generalizability improvements, the study also takes into account treatment effect heterogeneity by comparing the estimates of all GAAPs versus strong GAAPs.

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 difference 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 transfer to financiers with strong procedures like those of the World Bank.

8. Conclusion

Previous literature has indicated that it is mostly impossible to obtain good aid outcomes in weak institutional environments due to the politics of aid allocation, the inadequacy of Western advice, the "fungibility" of aid, and the primary importance of institutions to economic development. In contrast, I argued that multilateral aid can succeed in weak institutional environments because corruption scandals create legitimacy costs that nowadays force aid agencies and their donors to take countermeasures. These countermeasures notably include the development of large anti-corruption infrastructures in aid agencies and the use of GAAPs that are context-specific to each aid project.

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 investment projects and tested for potential endogenous relationships in the data. Because I found no such endogenous relationships, I proceeded to causally test my argument and found support for my hypothesis in a number of models after pre-processing the data through frontier matching. Each of the models suggest in GAAPs cause better development outcomes in contexts with weak institutions.

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 principal-agent style monitoring (e.g., Persson, Rothstein and Teorell, 2013; Kenny, 2017; Rasul and Rogger, 2018; Honig, 2019). More broadly, the results suggest that bureaucrats have agency to overcome the structural constraints imposed by weak institutions. Given that countries with weak institutions mostly need aid more than countries with strong institutions (Acemoglu and Robinson, 2012; Deaton, 2013), the policy implications of the results suggest that there is less of a trade-off between need and efficiency than most scholars and practioners believe. When there are institutional risks, bureaucrats can monitor them and design for success with relevant action plans.

Appendices

A. Coding Strategy

A.1. Coding of GAAPs

As specified in Section 3.1, 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.[14] 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

[14] The 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 GAAPsto 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 3.1).

A.2. Coding of GAAP Attributes

As Figure 2 showcases, I code the following components:

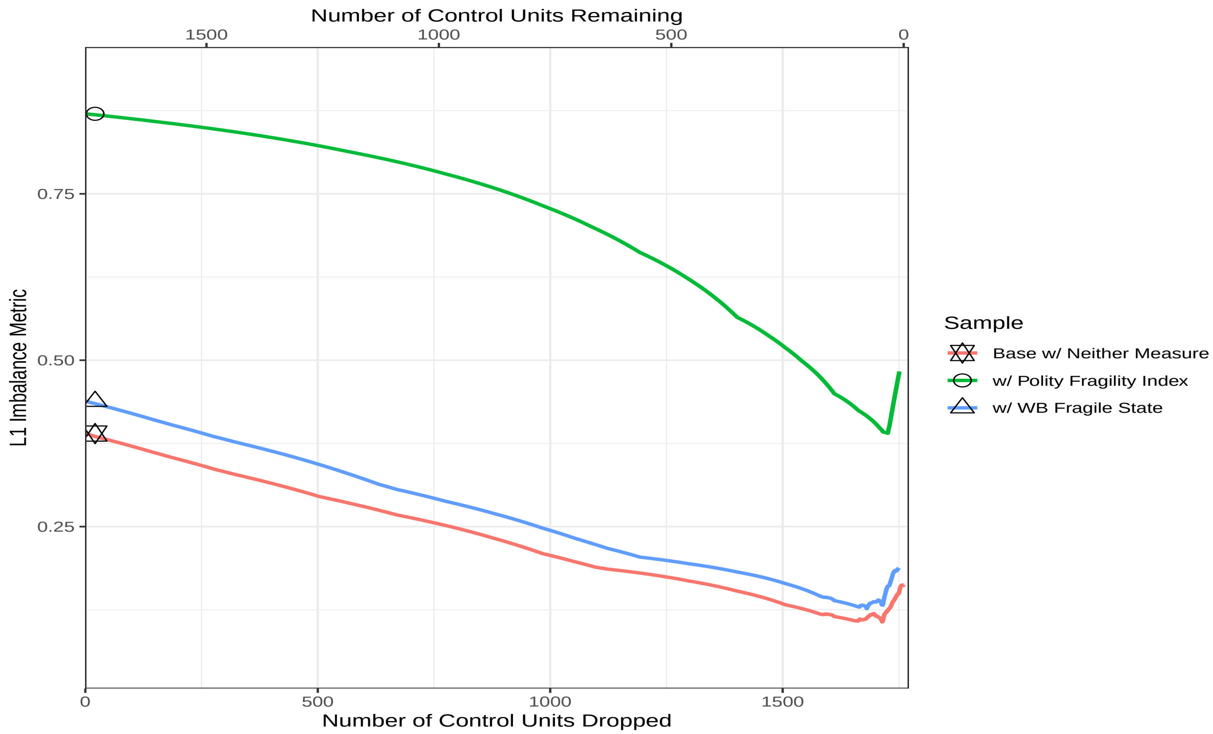
- internal audits: financial audit conducted by auditors within the same government implementing agency.
- external audit: financial audit conducted by either a different government agency or external firm.
- performance audit: audit designed to improve performance, as opposed to monitor it.
- technical 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

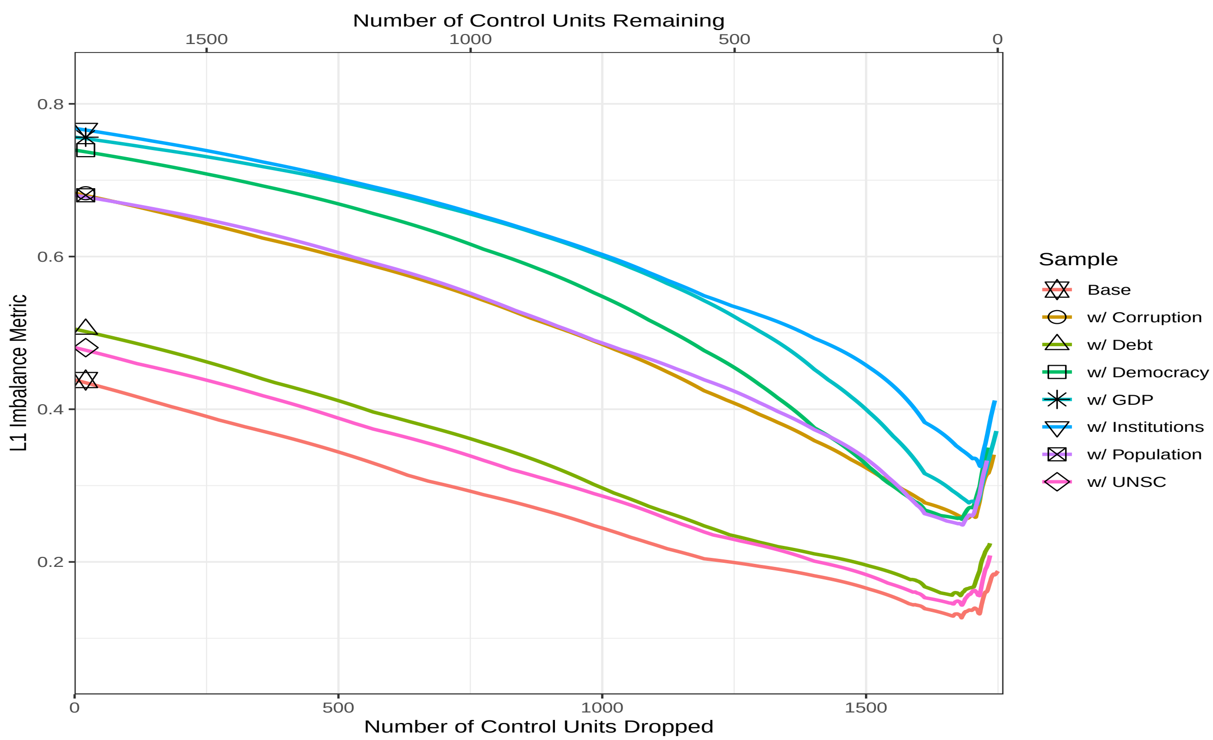
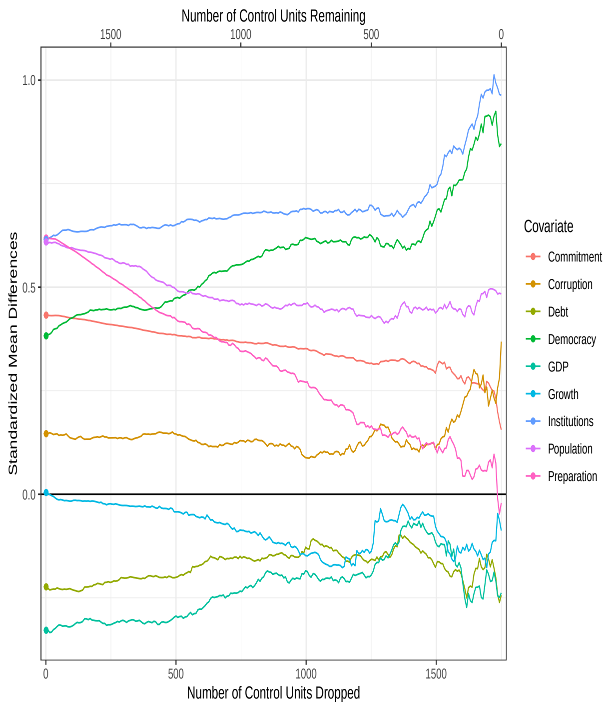
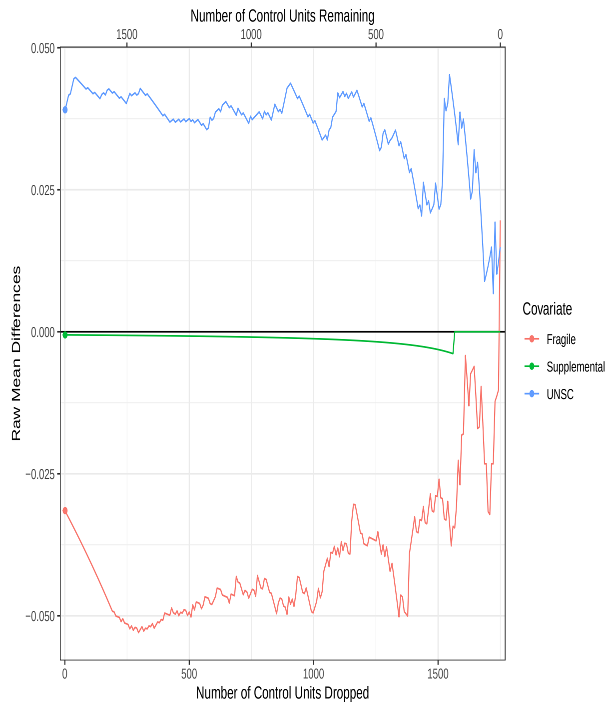


Figure B.2: Balance Diagnostics

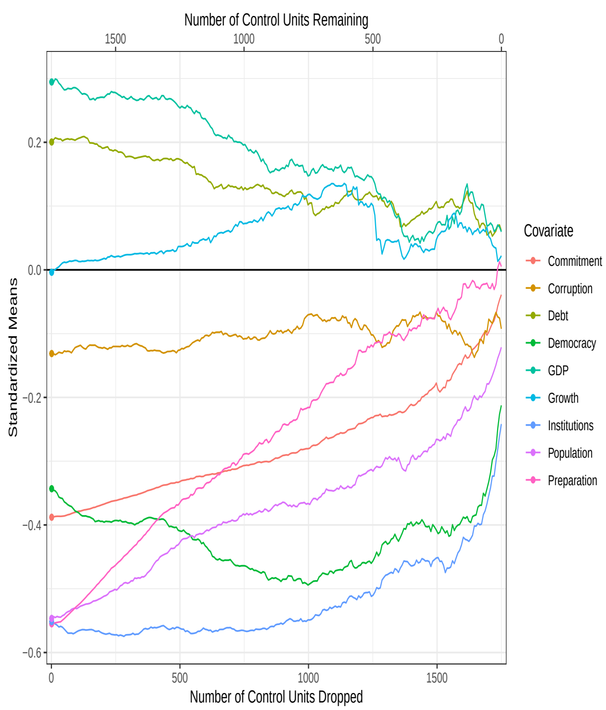
(a) Standardized Mean Differences



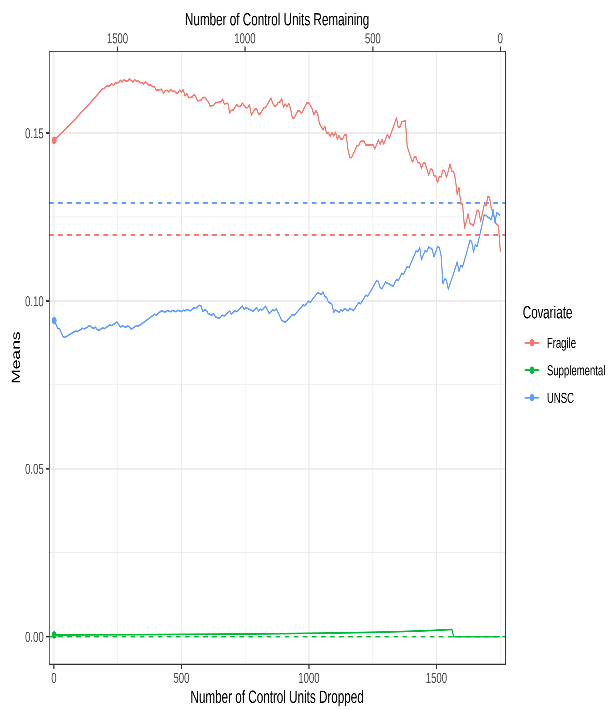
(b) Raw Mean Differences



(c) Standardized Means



(d) Raw Means



Bibliography

Abadie, Alberto and Jann Spiess. 2022. "Robust Post-Matching Inference." *Journal of the American Statistical Association* .

Acemoglu, Daron and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty.* New York: Crown Business.

Acemoglu, Daron, Simon Johnson and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91(5):1369–1401.

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.

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.

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.

Bearce, David H. and Brandy J. Jolliff Scott. 2019. "Popular Non-Support for International Organizations: How Extensive and What Does This Represent?" *Review of International Organizations* 14(2):187–216.

Beenstock, Michael. 1979. "Corruption and Development." *World Development* 7(1):15–24.

Ben-Michael, Eli, Avi Feller and Jesse Rothstein. 2021. "The Augmented Synthetic Control Method." *Journal of the American Statistical Association* 116(536):1789–1803.

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.

Booth, David. 2011. "Aid, Institutions and Governance: What Have We Learned?" *Development Policy Review* 29:s5–s26.

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.

Burnside, Craig and David Dollar. 2004. "Aid, Policies, and Growth: Reply." *American Economic Review* 94(3).

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*.

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.

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. 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.".

Easterly, William, Ross Levine and David Roodman. 2004. "Aid, Policies, and Growth: Comment." *American Economic Review* 94(3):774–780.

Einhorn, Jessica. 2001. "The World Bank's Mission Creep." *Foreign Affairs* 80(5):22–35.

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.

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 Multilevel/Hierarchical Modelling*. Cambridge: Cambridge University Press.

Gerber, Alan S. and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton & Company.

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.

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.

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.

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.

Hernandez, Diego. 2017. "Are "New" Donors Challenging World Bank Conditionality?" *World Development* 96(2007):529–549.

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. 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. 2022. "Matching Methods for Causal Inference with Time-Series Cross-Sectional Data." *American Journal of Political Science* .

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.

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.

Kosack, Stephen. 2003. "Effective Aid: How Democracy Allows Development Aid to Improve the Quality of Life." *World Development* 31(1):1–22.

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.

Lake, David A. and Mathew D. McCubbins. 2006. The Logic of Delegation to International Organizations. In *Delegation and Agency in International Organizations*, ed. Darren G. Hawkins, David A. Lake, Daniel L. Nielson and Michael J. Tierney. New York: Cambridge University Press chapter 12, pp. 341–368.

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.

Mauro, Paolo. 1995. "Corruption and Growth." *Quarterly Journal of Economics* 110(3):681–712.

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.

Mungiu-Pippidi, Alina. 2013. "Controlling Corruption through Collective Action." *Journal of Democracy* 24(1):102–115.

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.

Neyman, Jersey. 1935. "Statistical Problems in Agricultural Experimentation." *Journal of the Royal Statistical Society* II(2):108–180.

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.

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.

Rasul, Imran and Daniel Rogger. 2018. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." *Economic Journal* 128(608):413–446.

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. 2006. "Goodbye Washington Consensus, Hello Washington Confusion? A Review of the World Bank's Economic Growth in the 1990s: Learning from a Decade of Reform." *Journal of Economic Literature* 44(4):973–987.

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. 1978. *Corruption: A Study in Political Economy.* New York: Academic Press.

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.

Stiglitz, Joseph E. 2002. *Globalization and Its Discontents.* New York: W.W. Norton & Company.

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.

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: .

VanderWeele, Tyler J. and Miguel A. Hernán. 2013. "Causal Inference under Multiple Versions of Treatment." *Journal of Causal Inference* 1(1):1–20.

Vreeland, James Raymond. 2008. "The Effect of Political Regime on Civil War: Unpacking Anocracy." *Journal of Conflict Resolution* 52(3):401–425.

Vreeland, James Raymond. 2019. "Corrupting International Organizations." *Annual Review of Political Science* 23:205–224.

Weaver, Catherine E. 2008. *Hypocrisy Trap: The World Bank and the Poverty of Reform.* Princeton, New Jersey: Princeton University Press.

Wei, Shang-Jin. 2000. "How Taxing Is Corruption on International Investors?" *Review of Economics and Statistics* LXXXII(1):1–11.

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.".

Wolfensohn, James D. 2010. *A Global Life: My Journey Among Rich and Poor, from Sydney to Wall Street to the World Bank.* PublicAffairs.

World Bank. 1945. "International Bank for Reconstruction and Development Articles of Agreement.".

World Bank. 1992. Governance and Development. Technical report World Bank Washington, DC: .

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*a*. "Corruption is "Public Enemy Number One" in Developing Countries, says World Bank Group President Kim.".

World Bank. 2013*b*. 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.