Measuring Corruption Using Governmental Audits: A New Approach and Dataset*

Michael Denly†

September 30, 2020

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

www.mikedenly.com/files/audit-measurement.pdf

Abstract

For about 25 years, empirical scholarship on corruption has primarily relied on perceptions data, but the drawbacks of these measures are ample and well-known. More recently, analyses centered on Brazil have showcased the utility of randomly assigned audits as a more objective alternative to perception-based measures. However, Brazil is the only country with randomized audits and has many unique institutional features that limit the external validity of the numerous studies using the Brazil data. In this paper, I provide a new framework to assess the quality of audit data even when they are not randomly assigned. Specifically, I show that it is acceptable to use experimental or observational audit data to measure corruption when: 1) the auditing institution is legally independent from the executive branch; 2) the distribution of audits is not biased against opposition party politicians, especially following close elections; and 3) the intensity/dosage is consistent across similar types of audits. I demonstrate the utility of the framework by analyzing a massive new dataset of subnational audits from India, Mexico, Honduras, and Guatemala. The new data and framework proposed in this paper will help researchers undertake more objective analyses of governmental corruption around the world.

*For extraordinary research assistance, I thank the following current and former Research Affiliates and Graduate Research Fellows at UT Austin's Innovations for Peace and Development: Adityamohan Tantravahi, Akshat Gautam, Alejandra Zuniga, Amila Lulo, Ana Suarez, Ana Ramirez, Awab Ahmed, Blaine Finstein, Brayan Salgado, Caleb Rudow, Carmen Magana, Caroline Hastings, Caroline Woodman, Connor Hanna, David Tibaquira, Dorna Abdi, Elena Solimano, Emily Ibarra, Erin Eggleston, Ethan Tenison, Evelin Caro Gutierrez, Fabiola Ramirez, Felipa Mendez, Francesca D'Annunzio, Isabel Campos, Iulia Tothezan, Ishani Pandya, Ivana Jelensky, Jenny Rodriguez, Joelean Hall, Jonah Isaac, Josh Meuth Alldredge, Jude Rios, Judy Lane, Kais Bhandari, Karan Kanatala, Kate Hopp, Katherine Alonzo, Lissette Almanza, Luca Venegoni, Luisa Venegoni, Magdalena Ibarra, Marifer Guerrero, Mary White, Mishan Kara, Mohamed Abulfalgha, Mobin Piracha, Nicole Domingo, Nicky Pownall, Paige Johnson, Rachel Boles, Rachel Rosenburg, Ryan Jabra, Saamia Imtiaz, Samuel Gilette, Sara Lowe, Sterling Mosley, Tara Jackson, Temi Ajibola, Vanessa Gonzales, and Vanessa Lizcano. For advice or feedback, I thank Akshat Gautam, Arun Manuja, Ilona Wysmulek, Jan Pierskalla, John Gerring, John Ogallo, Lorenzo Crippa, Mike Findley, Ken Greene, Serge Nkuindja, and Xiaobo Lu. All errors are those of the author.

†PhD Candidate, Department of Government, University of Texas at Austin. mdenly@utexas.edu

Since the publication of Mauro's (1995) seminal study just over 25 years ago, empirical scholarship on governmental corruption has primarily relied on perceptions data.[1] As numerous studies have demonstrated, that is a grave problem: perceptions data suffer from myriad methodological challenges that make them unsuitable for advancing new theory and knowledge (e.g., Kurtz and Schrank, 2007*a*,*b*; Andersson and Heywood, 2009; Hollyer, 2018). Nevertheless, the empirical findings underpinning some of the most prominent recent theoretical advances on corruption continue to be based mainly off perceptions data (e.g., Mungiu-Pippidi, 2015; Schwindt-Bayer and Tavits, 2016).

Over the past 10-15 years, many objective measures of corruption have emerged as alternatives to perceptions data (e.g., procurement red flags, abnormal asset growth, tax evasion), but violations contained in governmental audit reports are perhaps the most promising. Corruption is a multifaceted, complex phenomenon, and audits capture that complexity—everything from the aforementioned objective measures to document fraud, theft, nepotism, etc. Given that audits are specific to individuals or institutions, they also overcome the level-of-analysis problems that plague most country-based research using perceptions data (see Gingerich, 2013*a*).

To date, however, scholars using audit data have focused almost exclusively on the randomly assigned municipal audits in Brazil.[2] On the one hand, these data from Brazil are useful because they overcome selection/endogeneity problems and allow for causal estimation. On the other hand, scholars' constant use of these same data has generated knowledge with severe external validity challenges. Notably, Brazil is the only country in the world with randomized audits, and the country has a very unique "web" of institutions supporting its audit agency (Ferraz and Finan, 2018). Fortunately, many other countries disseminate subnational audit data, but researchers need a framework for discerning the validity of these

[1] So many studies rely on corruption perceptions data that it is impossible to mention them all here, but some of the most pertinent ones include Mauro (1995) on growth; Gerring and Thacker (2004, 2005), Kunicová and Rose-Ackerman (2005), and Lederman, Loayza and Soares (2005) on institutions; and Treisman (2000, 2007) on culture, economic development, and democracy.

[2] See, for example, Ferraz and Finan (2008, 2011), Ferraz, Finan and Moreira (2012), Melo, Pereira and Figueiredo (2009), Brollo et al. (2013), Pereira and Melo (2015), Brollo and Troiano (2016), Avis, Ferraz and Finan (2018), and Cavalcanti, Daniele and Galletta (2018).

data.

In this article, I demonstrate how researchers can profitably use subnational audit data to objectively measure corruption for observational, experimental, and quasi-experimental research. Three pillars underpin my framework for assessing the utility of audit data. First, the institution performing the audits must be independent from a country's executive branch, which can be assessed through legal and budgetary protections. Second, given that most countries' audit institutions have discretion to perform risk-based audits, the distribution of audits must not be biased against opposition party politicians—especially following close elections. Third, the most credible audit data will also be able to empirically show fairness in the intensity/dosage of audits. For example, opposition party politicians should not be subject to more intense or stringent audits than ruling-party politicians. Since relevant quantitative data will not always be available to assess intensity/dosage, it is at least necessary to qualitatively examine auditor codes of ethics/conduct and corresponding sanction systems.

I demonstrate the utility of my framework by collecting and analyzing a massive, new, micro-level dataset of subnational audit findings from India, Mexico, Honduras, and Guatemala. The specific data vary by country, but the panel dataset covers infractions committed, money stolen/missing, follow-up on audit recommendations, whistleblower complaints received, and relevant information on the sector and sub-sector of the corruption activity. Overall, the new data and framework proposed in this paper will help researchers undertake more objective analyses of governmental corruption around the world.

The paper proceeds as follows. In Section 1, I lay out the challenges to measuring corruption and explain why perceptions-based indicators cannot usefully advance new theory and knowledge. Section 2 describes the benefits of audit data in measuring corruption and the external validity challenges associated with over-relying on the Brazilian audit data. Section 3 details the framework and data central to this paper. Using the novel data that I collected from India, Mexico, Honduras, and Guatemala, in Section 4 I analyze the extent to which the data from each country meet the framework. The final section concludes.

1. Measuring Corruption with Observational, Perception-Based Data

Corruption entails "the misuse of public office for private gain",[3] and many analysts define the phenomenon to be much broader (e.g., Gingerich, 2013*b*; Mungiu-Pippidi, 2015). Irrespective of how one defines corruption, though, one thing is certain: by its very nature, corruption is a clandestine activity, so it is very difficult to measure.

The International Country Risk Guide (ICRG), Transparency International (Corruption Perceptions Index-CPI), and the World Bank (Control of Corruption score-Worldwide Governance Indicators) constructed the first widely-available corruption measures. Each of these indexes embarked on measuring corruption by aggregating and re-scaling survey data from businesspeople. Over time, as more data became available, Transparency International and the World Bank greatly diversified the data from which they constructed their corruption measures (e.g., Kaufmann, Kraay and Mastruzzi, 2011, 225). In the process, the measures gained significant conceptual intention,[4] which is especially useful because corruption is multidimensional concept with experience-near and experience-distant meanings.[5]

Although development of the perceptions indexes represented major steps forward in terms of measuring corruption, they never were suitable for theory development and testing (Hollyer, 2018). ICRG developed its index for the purposes of helping businesses make decisions about corruption risks related to foreign investment, so the measure is necessarily limited in terms of content validity.[6] Among its many limitations, Transparency International CPI data are not suitable for over time comparisons, which severely scholars ability to

[3] This is probably the most-commonly accepted definition of corruption. For more on the definition of corruption, see, for example, Rose-Ackerman and Palifka (2016).

[4] Intention is a synonym for connotation, meaning the "ensemble of characteristics and/or properties associated with, or included in, a given word, term, or concept" (Sartori, 1984, 24).

[5] "Experience-distant concepts are ones that specialists of one sort or another... employ to forward their scientific, philosophical, or and practical aims. Experience-near concepts, in contrast, are one[s] that someone might himself naturally and effortlessly use to define what he or his see, feel, think, imagine, and so on, which he would readily understand when similarly applied by others (Schaffer, 2016, 2)."

[6] "Content validity assesses the degree to which an indicator represents the universe of content entailed in the systematized concept being measured" (Adcock and Collier, 2001, 537).

use them for theory (Andersson and Heywood, 2009, 758). The World Bank's Control of Corruption score is perhaps the the most sophisticated of the three measures and is suitable for over time comparisons. Still, the measure has limited construct validity,[7] as well as suffers from information leakage,[8] halo effects,[9] and content opacity (Kurtz and Schrank, 2007*a*,*b*; Langbein and Knack, 2010; Thomas, 2010; Bersch and Botero, 2014; Gisselquist, 2014; Hollyer, 2018).[10]

Information leakage, halo effects, content opacity, and low construct validity are very problematic from the perspective of theory and knowledge-building (Hollyer, 2018). All of these deficits are related, too. For example, when participants respond to survey questions about corruption, it is difficult for them to exclude extraneous information and concepts that may overlap such as democracy (Hollyer, 2018). Additionally, as Kurtz and Schrank (2007*a*,*b*) show, countries' growth trajectories cloud survey respondents' answers on governance-related questions. Such problems compound as perceptions indexes include more sources, too, which makes the precise scope of perception-based measures necessarily opaque. With opaque measures, it is impossible to precisely verify the degree of construct validity: that is, whether concepts measure what they are supposed to measure (Trochim, 2006; Thomas, 2010). Which begs the question: how can a measure without construct validity be useful for theorizing? In short, its ability to do so is very limited, especially when investigating proximate and endogenous phenomena like democracy and corruption.

More recently, scholars have developed useful, Bayesian-based improvements to the

[7] Construct validity concerns whether the concepts measures what it is supposed to measure (Trochim, 2006). In particular, the Worldwide Governance Indicators have trouble with a component of construct validity called discriminant validity. It concerns whether measures are not associated with measures that they are not supposed to be associated with. The worldwide governance indicators lack discriminant validity because the correlation among indicators i so high that one may ask whether they are really different Langbein and Knack (2010); Thomas (2010).

[8] Information leakage refers to when a survey taker's perspective on one phenomenon is influence by something else.

[9] Halo effects specifically refer to when positive effects about something color or influence a person's opinion in a positive way about something else.

[10] Kurtz and Schrank (2007*a*,*b*) also critique the Worldwide Governance Indicators for systematic measurement error, sampling bias, and cultural biases, but those critiques are less convincing nowadays, particularly since Kaufmann, Kraay and Mastruzzi (2011) have included more sources and countries.

perception indexes (Bersch and Botero, 2014; Standaert, 2015; Coppedge et al., 2020).[11] The most sophisticated of these measures is the Varieties of Democracy (V-Dem) Project's hybrid approach (Marquardt and Pemstein, 2018; Coppedge et al., 2020). V-Dem uses both fact-based and perceptions data, and complements them with targeted measures to improve reliability and validity. Nevertheless, any index relying on perceptions-based measures cannot fully overcome the aforementioned mentioned trade-offs, because they still exist—albeit to a smaller degree (Hollyer, 2018, 118, 128).

2. Audit Findings as Objective Measures of Corruption, and the External Validity of the Brazil Studies

Generally, objective measures of any phenomenon, including corruption, do not capture as much conceptual intention as the perception-based ones, but audit-based measures are not as susceptible to such a drawback. The reason pertains to the diversity of violations that audit reports can reveal. Procurement red flags, ghost firms, theft, nepotism, spending misappropriations, and tax evasion are just some corrupt violations that audit reports examined for this study and others reveal (e.g., Ferraz and Finan, 2008; Brollo et al., 2013).

Since audit-based corruption measures, like any other objective measure, refer to actual instances of corruption, they also do not suffer from the same information leakage, halo effects, and construct validity challenges as perception-based measures. Accordingly, objective measures are more useful for theorizing and knowledge creation than perceptions-based measures (Hollyer, 2018). Perhaps equally as significant, objective measures such as audit violations can solve the level-of-analysis problem that plagues most country-level perceptions data: that is, the fact that within-country variation often overwhelms national-level

[11] The benefits and drawbacks of Bayesian statistics as compared to their frequentist counterparts greatly exceed the scope of this paper. In brief, the Bayesian measurement models of corruption such as Bersch and Botero (2014), Standaert (2015), and Coppedge et al. (2020) allow for less missing data, incorporation of prior data, and are more attuned to measuring intractable, unobservable concepts like corruption (Fariss, Kenwick and Reuning, 2020).

variation (Gingerich, 2013*a*, 505, 538). Aside from equivalence considerations,[12] the use of subnational data also tends to facilitate natural experiments and causal inferences—as opposed to correlational analyses with regression (Stanig, 2018, 150-151).

On the subject of innovative corruption measurement techniques that use natural experiments at the right level of aggregation, Ferraz and Finan's (2008) study of randomized municipal audits in Brazil pioneered an entire research agenda.[13] From the perspective of measurement, these objective data on corrupt acts like procurement fraud or over-invoicing from Brazil measure true corruption, not preceptions of it (Ferraz and Finan, 2008, 710). Also, due to the fact that the audits are randomly assigned, they obviate the many endogeneity concerns that corruption typically engenders (Escaleras, Anbarci and Register, 2007). In turn, from the perspective of knowledge creation, these data enabled Ferraz and Finan (2008) to conclusively show—for the first time—that information about politicians' corruption levels can hurt their re-election prospects. Other important results stemming from the Brazilian audit infractions data include that: re-election incentives condition politicians' corruption levels (Ferraz and Finan, 2011); getting audited nudges subnational government entities to reduce their future corruption levels—although effects are sectorally heterogeneous (Avis, Ferraz and Finan, 2018; Zamboni and Litschig, 2018); malfeasance revelations affect the types of candidates that parties put forth on party lists (Cavalcanti, Daniele and Galletta, 2018); and female politicians are less corrupt than their male counterparts (Brollo and Troiano, 2016).

The above studies are all very precisely estimated with objective data, but how much external validity do they have? In other words, how certain can scholars be that the results apply equally to other settings, time periods, units (e.g., countries), mechanisms, as well as different operationalizations of the treatment and outcomes (Findley, Kikuta and Denly, 2021)? For practical, substantive, and methodological considerations, the corruption

[12] For more on equivalence, see Locke and Thelen (1995) and Stegmueller (2011).

[13] In this article, I focus on audits. However, there were many other innovations to measuring corruption using, for example, objective data from procurement (e.g., Broms, Dahlström and Fazekas, 2019; Fazekas and Kocsis, 2020), asset declarations (e.g., Eggers and Hainmueller, 2009; Fisman, Schulz and Vig, 2014), and taxes/duties (e.g., Rijkers, Baghdadi and Raballand, 2017; Naritomi, 2019).

literature needs to move beyond Brazil.

2.1. Substantive and Practical External Validity Concerns with the Brazil Studies

Brazil is the only country in the world with randomized subnational audits. In turn, that means Brazil is the only country where scholars can be certain that partisan considerations do not influence who gets audited and at what time. Unequivocally, that is very useful from the perspective of causal inference. By the same token, even countries with Weberian bureaucracies like Sweden are subject to partisan pressures (Dahlström and Holmgren, 2019). Accordingly, understanding the extent of impartiality in the distribution of subnational audits is part and parcel of answering the question of whether audits deter political corruption.[14]

Practically, most countries' supreme audit institutions (SAIs) need to use their own discretion to choose where to audit through risk-based audits. High-corruption areas, or ones for which SAIs receive whistleblower complaints, simply need more attention than low corruption areas. Similarly, it is practically unreasonable for large, populous cities with large budgets to always have the same probability of receiving an audit as villages with smaller budgets. Simple randomization is not always fair or useful without some sort of blocking strategy or stratified sampling,[15] which is something that the Brazilian government has not implemented.

At least three additional substantive features make the external validity of the Brazil studies limited. First, the country's media facilitates collective action by exposing corrupt

[14] Olken (2007), Duflo, Hanna and Ryan (2012), and Alm, Martinez-Vazquez and McClellan (2016) show that audits work in other contexts, but Dizon-Ross, Dupas and Robinson (2017) and Dhaliwal and Hanna (2017) have also shown that audits demotivate qualified staff, resulting in many qualified staff to leave their positions.

[15] Block randomization refers to a randomizing strategy that entails dividing treatment and control groups into equally-sized groups. Stratification refers to sampling on the basis of strata or groups (Gerber and Green, 2012).

politicians.[16] Indeed, both Ferraz and Finan (2008) and Cavalcanti, Daniele and Galletta (2018) show that their results are stronger when there is greater local media presence, which is something that not all relatively corrupt countries enjoy. Second, the country has compulsory voting that contributes to relatively large voter turnout and more engagement in politics than in many countries. Third, at least since 1988, the country has a "web" of institutions (i.e., Comptroller General, Federal Audit Court, Regional Audit Courts, Public Ministry, Federal Police) that promote horizontal accountability in anti-corruption (Ferraz and Finan, 2018, 255).[17] Clearly, not all countries with challenges controlling corruption have such a robust institutional presence. That is especially significant because there is scholarly consensus that institutions are the fundamental cause of development, growth, and corruption (Acemoglu and Robinson, 2012; Fisman and Golden, 2017, 18).

2.2. Methodological Challenges with the External Validity of the Brazil Studies

Methodologically, there are concerns with over-relying on one country as well. To understand why, first examine how the samples of the Brazil Studies in Table 1 differ from their population and target inference counterparts. As is apparent, the gaps are often quite wide to make general statements. To precisely understand how these gaps can result in biased inference, consider the simple difference-in-means estimator, $\hat{\delta}_S$. Assuming a binary treatment, Findley, Kikuta and Denly (2021) show that it is possible to decompose $\hat{\delta}_S$ using the potential outcomes framework as follows:[18]

$$\hat{\delta}_S = \delta_P + b_{S1} + b_{S2} + b_P + b_V. \tag{1}$$

[16] For more on how the media plays a crucial role in exposing corruption and fostering electoral accountability, see Besley and Burgess (2002), Snyder and Strömberg (2010), and Larreguy, Marshall and Snyder (2020).
[17] Horizontal accountability refers to the ability of the bureaucracy to exert checks and balances on itself (O'Donnell, 1998).
[18] For an overview of the potential outcomes framework, refer to Imbens and Rubin (2015).

Table 1: External Validity Dimensions of the Brazil Studies

Dimension	Sample(s)	Population(s)	Target(s)
Mechanism	Local media, election proximity, politician quality (education level)		
Settings	Comptroller General audit lotteries program on use of federal transfers	Regular municipal audits, audits of other programs, experiments, or other types designs	All countries in the world
Treatment	Municipalities receiving an audit, mayor ability to run for reelection, being a female mayor	Subnational unit receiving an audit, politician ability to run for reelection, being a female politician in Brazil	Getting audited, reelection, being a female politician in any country
Outcome	Corrupt infractions or misappropriations in use of federal transfers	Subnational political corruption in Brazil	Subnational political corruption in all other countries
Unit	Municipalities receiving an audit, municipalities with politicians seeking reelection or all municipalities in Brazil below certain populations	All municipalities in Brazil	All subnational political corruption in countries with reelection
Time	2004-a few years before the publication of each respective study	2004-present	All years

Note: "Brazil Studies" refer to Ferraz and Finan (2008, 2011), Brollo et al. (2013), Brollo and Troiano (2016), and Avis, Ferraz and Finan (2018). Other studies that employ the same data, but the above studies correspond to those using corruption as a dependent variable.

In Equation (1), δ_P corresponds to the Population Average Treatment Effect (PATE), which is the ideal quantity of theoretical interest;[19] b_{S1} refers to the selection bias;[20] b_{S2} is the difference in the strength of the treatment effects between the treatment and control groups—i.e., within-sample effect heterogeneity; b_P is the bias due to non-random sample selection; and b_V refers to variable heterogeneity bias—i.e., in treatments, outcomes, and mechanisms.

[19] Not all (quantitative) studies will aim to estimate the PATE. However, given that the ultimate goal of social science is to make inferences beyond the data at hand (King, Keohane and Verba, 1994, 8,34), having an estimate correspond to the population average treatment effect (PATE) is an implicit goal in social science.

[20] Selection bias corresponds to the bias due to lack of random assignment (Angrist and Pischke, 2008).

The only bias that randomization in the assignment of Brazilian municipal audits data solves is b_{S1}. The extent to which randomization solves b_{S2} depends on the size of the treatment and control groups in the sample. Given that the Brazilian Comptroller General only randomly audits 30-60 municipalities per year out of a total of $5,570$ municipalities, it is difficult to argue that $b_{S2} = 0$ in the Brazil Studies. The unique institutional features described in Section 2.1 enter into Equation (1) through b_P and b_V—i.e., what Findley, Kikuta and Denly (2021) term "external validity bias". Under many circumstances, it is feasible that the external validity bias can overwhelm the internal validity bias, represented by b_{S1} and b_{S2}. In short, while the Brazil studies represent excellent contributions to the corruption literature, random assignment is not sufficient for external validity or general knowledge creation more broadly. Scholars still need more evidence from other countries, time periods as well as settings with potentially different treatments, outcomes, and mechanisms to assess the extent to the which results generalize. Notably, these future studies will need to take into account how politics conditions the results.

3. Assessing the Utility of Non-Randomized Audit Data

There is a dearth of political corruption studies employing audit data outside of Brazil.[21] One reason likely pertains to the lack of easily-accessible audit data. A second reason is that scholars do not have systematized ways of discerning the validity of audit data, especially when audits are not randomly assigned. In the absence of randomly assigned audits, it is feasible that presidents or prime ministers can use their control of the executive branch, which includes the bureaucracy, to unfairly audit political opponents. The risk is particularly high in poor developing democracies and countries with low horizontal accountability[22]—that is,

[21] Bobonis, Cámara Fuertes and Schwabe (2016), Larreguy, Marshall and Snyder (2020), and Ajzenman (2020) are, to my knowledge, the only audit studies outside of Brazil using administrative data—outside of studies on individual one-off programs, such as Di Tella and Schargrodsky (2003), Olken (2007), and Nikolova and Marinov (2017).

[22] Horizontal accountability concerns the ability to keep checks on itself—i.e., the executive branch. For new data on horizontal accountability, refer to Lührmann, Marquardt and Mechkova (2020) For more on the poor performance of poor democracies, see Keefer (2007*a*,*b*).

exactly the types of countries where corruption tends to be high as well.

In this section, I attempt to address both of the above gaps in the literature. First, I provide a new, three-pillar framework for assessing the utility of audit data. The framework is especially useful for when audits are not randomly assigned, but all three of its pillars are also relevant for experiments or natural experiments that randomly assign audits. Second, I provide a wealth of new, micro-level audit data to demonstrate the utility of the framework and facilitate further audit-based corruption research using these objective data.

3.1. A New Framework

The first pillar underpinning the framework is that the institution performing the audits must be ***legally independent from a country's executive branch***. Generally, it is possible to verify that independence through legal and budgetary provisions. As highlighted by the International Organization of Supreme Audit Institutions (2019), the strongest legal provisions regarding audit independence are those in a country's constitution. In particular, an independent audit institution will have tenure security for the agency head, budget autonomy, and discretion to audit without securing approval from politicians. For example, if a country's President or Prime Minister can easily fire and replace the audit agency head at will, it will be difficult for the audit process to be apolitical. The same is true for budgets: if audit agencies receive smaller outlays after undertaking unpopular audits, then independence becomes more elusive. That is why there is less risk if audit agencies report to Congress or a parliament instead of a president or prime minister. This way, there is less possibility for a single individual to unduly influence the audit process.

Second, the ***distribution, frequency, or selection*** of audits must not be biased against opposition party politicians—especially following close elections. Although many observers may think that randomization immediately solves this second pillar, some forms of random assignment are more credible than others. Complete random assignment,[23] block

[23] In complete random assignment, units are randomly assigned into equally-sized treatment and control groups.

randomization, or a well-justified, stratified sampling strategy[24] can help meet the pillar for experimental studies. Simple randomization,[25] such as in the Brazil studies, is generally sufficient but less credible.

For observational studies to meet this second pillar on audit distribution fairness, quantitative analysis is necessary. If all units (e.g., states, municipalities) do not receive the same number of audits for the given time interval, then it is necessary to perform a maximum likelihood regression (see King, 1998). When the dependent variable refers to a count of the number of audits received during an interval, then a poisson or negative binomial regression is germane. If not all units receive an audit during the particular interval, a logistic regression model is appropriate. In either scenario, especially given that SAIs' non-randomly assigned audits are supposed to be "risk-based", it is useful to control for population (more populated places will generally have more vulnerable revenue), past corruption (either lags or cumulative sums), and revenue amounts.

Regardless of what the maximum likelihood regressions show, it is necessary to complement the above analyses with a regression discontinuity design (RDD). Specifically, the latter needs to follow the setup of Brollo and Nannicini (2012) and examine whether party alignment between the executive and lower-level government units predicts audit frequency after close elections. The RDD not only allows for quasi-causal estimation but is particularly relevant because countries' presidents may wish to use their control of the bureaucracy to target audits at electorally vulnerable opposition party politicians. If either the maximum likelihood regressions or the RDD consistently show that opposition/unaligned subnational units are more likely to receive audits, then it is clear that the data are biased.

The third pillar of the framework pertains to audit ***intensity/dosage***. More precisely, the third pillar regards whether all units in the sample (municipalities, politicians, etc.) receive audits that are similarly stringent. If, for example, politicians aligned with the

[24] In stratiefied random sampling, the researcher devises a strategy to sample

[25] By simple randomization, I mean independent random draws (think: coin flips) to decide whether a unit is assigned to the treatment or control group. In simple randomization, there is no guarantee that the treatment and control groups are equally-sized.

president receive less stringent than those given to opposition party politicians, then it is difficult to assert that the audit process is fair. By the same token, it is necessary to note that some more stringent types of audits, such as forensic audits, may be necessary to complement financial or compliance audits under some circumstances.

In many ways, the third criterion on intensity/dosage is the hardest to verify and might be the easiest to ignore. Not coincidentally, I did not find a single published study that attempted to examine audit intensity/dosage. One reason is likely that not every country will have the requisite quantitative data on the number of auditors sent to a place, duration of the audit, amount of money of audited as percent of the total, etc. When audit duration data are available, researchers can analyze them through various survival/event history analysis techniques.[26] Auditor counts can be analyzed with poisson or negative binomial regression, and shares of money audited can be fruitfully examined with linear or beta regression.[27]

Researchers can undertake relevant qualitative analysis of audit intensity/dosage as well. In fact, given that quantitative data on audit intensity/dosage will not always be accessible, qualitative data will often provide the only recourse assess audit intensity/dosage. Particularly relevant qualitative data include those related auditor codes of ethics/conduct. They are most credible when their scope is quite broad, especially regarding conflicts of interest, and there are frequent internal or external audits with consequences for auditor noncompliance.[28] Some very skilled qualitative researchers may also undertake interviews of auditing staff, but adequately convincing readers that social desirability bias[29] is not driving auditors' responses will be challenging. Ethnographic studies of auditors may be the most credible qualitative method to assess dosage. However, the skills required to avoid

[26] As Blossfeld, Golsch and Rohwer (2007) emphasize, it is to difficult to theorize about the functional form of event history models, so it is generally best to examine a wide range of event history methods. Future versions of the manuscript will provide further details.

[27] Note: there is no agreed upon method for beta regression with fixed effects, which is why scholars generally use linear regression when estimating a fixed effects model with a fractional dependent variable. For more on information, see Papke and Wooldridge (2008).

[28] External audits are especially important in instances of high state capture.

[29] Social desirability bias refers to when respondents are unwilling to admit to socially unacceptable behavior and thus provide inaccurate responses to enumerators.

Hawthorne effects[30] will be prohibitively high for most researchers.[31]

3.2. New Subnational Audit Data

Table 2: Summary of Original Data Collected

	Honduras	**Guatemala**	**Mexico**	**India**
Administrative Level	Municipal	Municipal	Municipal	State
Years Covered	2002-2018	2004-2018	2000-2018	2004-2018
Number of Audits	900	3,500	2,300	1,100
Sector	No	No	Yes	Yes
Sub-sector	No	No	No	Yes
Infractions	Yes	Yes	Yes	Yes
Precise Details of Infractions	Yes	No	Yes	Partial
Money Stolen/Missing	No	Yes	Yes	Yes
Follow-up on Audit Recommendations	No	No	No	Yes
Whistleblower Complaints	No	Yes	No	No
Bureaucrat Sanctions	No	Yes	No	No
Money Audited	No	No	Yes	Yes
Money Outside Audit	No	No	Yes	No

Table 2 summarizes the original data that I collected and cleaned with the help of a large team of research assistants. With respect to case selection, I restricted my initial searches for subnational audit data to developing countries that are members of the International Organization of Supreme Audit Institutions (INTOSAI). I focused on INTOSAI members because it has specific provisions in its Mexico and Lima Declarations regarding audit independence and data transparency (International Organization of Supreme Audit Institutions, 1977, 2007). Due to my language constraints, I also needed to limit my search to developing countries that provide documents in English, French, or Spanish.

From the above universe, I selected Guatemala (2004-2018), Mexico (2007-2018), Honduras (2002-2014), and India (2006-2018). Both the SAIs of Guatemala and Mexico provide municipal-level audit data on infractions and the amount of lost money associated with those

[30] Hawthorne effects refer to the when participants change their behavior because they know that they are under study.

[31] The extensive set of skills and techniques required to undertake a credible ethonographic study fall outside this scope of the present study.

Table 3: ASF Findings/Corrective Actions by Municipality (2007-2018)

Variable/Statistic	Mean	Max
Infractions	3.73	47
Money missing (deflated to 2013 Mexican pesos)	6,714,158.00	774,237,624.00
Money missing (log)	7.50	20.47
Share of total money audited	0.81	1.00
Financial irregularity with required compensatory action	0.39	10
Performance recommendations	0.16	7
Punitive noncompliance with regulations	1.65	44
Recommendation	4.90	67
Report of crime	0.06	4
Formal request for clarification	0.15	9
Statement of financial irregularity	1.47	16
Tax evasion or financial regulatory noncompliance	0.09	5

Note: data only refer to municipalities audited in each given year

infractions on their websites. Honduras' SAI only provides the municipal audits reports. Therefore, a team of research assistants and I inspected each infraction in very available report for corruption using a typology loosely based on the World Bank's Anti-Corruption Guidelines (see World Bank, 2016). India's state-level audits are very comprehensive. Like Honduras, though, India does not provide data in tabular format, so it was necessary to read all of the audit reports and extract the data from various tables and paragraphs in the reports.

Most countries in the sample provide data on infractions and missing money associated with those infractions, but the larger dataset contains even more data that will be of interest to corruption researchers. India, for example, provides yearly data on the number of infractions and money amounts that are outstanding from the previous years. For its part, Guatemala provides data on whistleblower complaints received and the number of bureaucratic charge reports filed. The most detailed classification of infractions, however, come from Mexico's SAI (see Table 3).

4. Analysis of the Country Cases

In this Section, I apply the framework in Section 3.1 to analyze the extent to which researchers can profitably use the non-randomized, subnational audit data from India, Mexico, Guatemala, and Honduras. As above, I analyze the legal independence of the audit institution from the executive branch, the extent of partisan distributional bias, and the audit intensity/dosage.

4.1. India

Legal Basis for Independence from the Executive Branch. The Comptroller and Auditor General (CAG) is the institution responsible for conducting governmental audits in India. The Auditor General leads the CAG and serves a term of six years. The relevant legislation protecting the Auditor General from political interference dates back to Audits and Accounts Order of 1936, which the Companies Act of 1956 and Duties, Powers, and Conditions (DPC) Act of 1971 reinforce. Articles 148-151 of the Indian Constitution further protect the independence of the CAG. Although the president appoints the Auditor General, the Parliament determines the salary and office requirements of the Auditor General. Additionally, the process for removing the Auditor General is akin to that of a Supreme Court Justice, and the Auditor General is not eligible for further political office after his or her term at the CAG ends. Given that the constitution grants the Auditor General even more autonomy in terms of day-to-day operations, it is clear that the Auditor General has significant independence from the executive branch to carry out unbiased auditing.

Frequency, Selection, and Distribution of Audits. Each year, India's CAG publishes audit reports for each of India's 29 states on its website. The audits reports are very comprehensive, often spanning hundreds of pages. Although the audits from 2004-2018 cover 70 different "sectors", there are audit reports every year for state finances, revenue, civil, commercial, economic, and social sectors, among others. Within these sectoral reports, the CAG generally undertakes in-depth analysis of at least one different sub-sector each year.

Accordingly, analysts can use these state-level audits data from India without concern of distributional bias along party lines.

Intensity/Dosage. The CAG has an extensive code of ethics for all of its auditors (Comptroller and Auditor General of India, 2012). [More analysis, including quantitative data, coming soon.]

4.2. Honduras

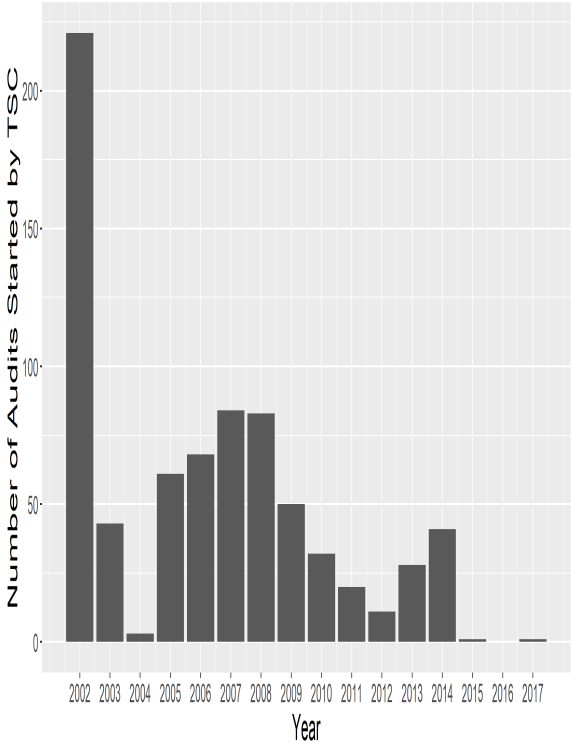
Legal Basis for Independence from the Executive Branch. The Supreme Tribunal of Accounts (TSC, *Tribunal Superior de Cuentas*) is the institution responsible for government audits in Honduras. Chapter III (Articles 222-227) of the Honduran Constitution and the Organic Law of the TSC (2002, revised 2011) provide the legal basis of the TSC. Overall, its legal basis is strong. Notably, the TSC reports to Congress, not the President, so the chance that any one individual can control the TSC is low. That is especially the case because the head of the TSC rotates annually over a three-year term, with the heads being selected uniquely by Congress; and the Organic Law of the TSC specifically excludes numerous politicians, such as the President, from being a member of the TSC. From an operational autonomy perspective, the TSC's is independence similarly robust. The Organic Law of the TSC supersedes all other laws, except those in the Constitution, so the TSC can engender compliance with its operations, and the Organic Law grants the TSC broad scope to do so.

The one area where the autonomy of the TSC is weaker pertains to its budget. Per the Law of Municipalities, the TSC is supposed to receive a budget akin to one percent of municipal revenues in the previous year. In 2011, the government revised the Organic Law of the TSC in order to allow for contributions from various other sources as well, including foreign aid. However, according to author interviews with various mayors as well as a TSC transparency request response, the budget is a recurring a challenge in Honduras more broadly, and the TSC is no exception.

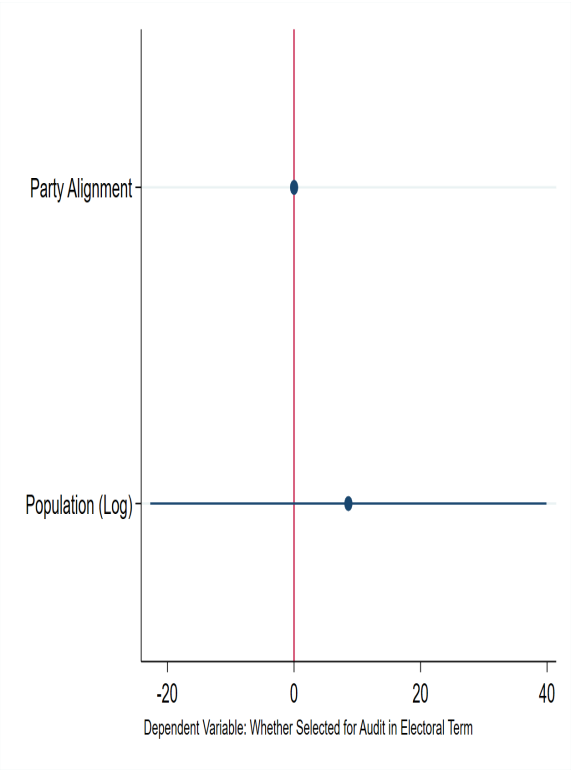
Frequency/Selection of Audits. The TSC performs numerous different types of audits

Figure 1: Main Results: Honduras' Municipal Audit Distribution

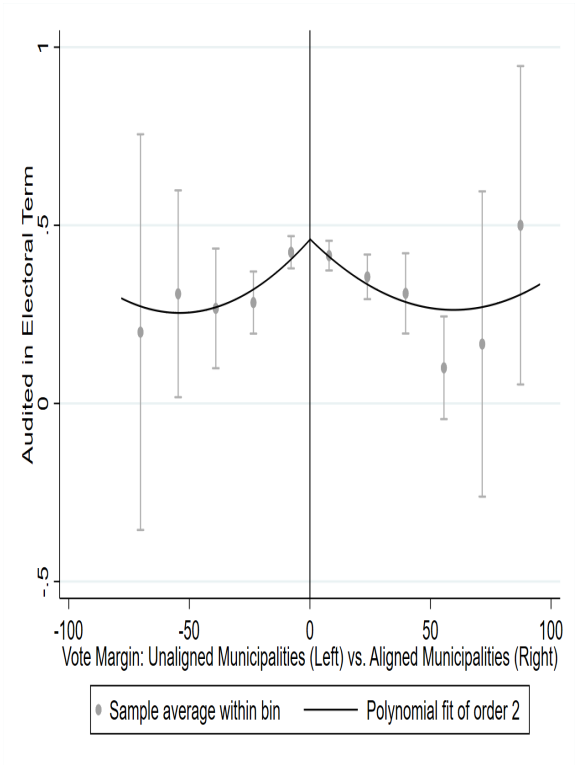
(a) Audits Started by the TSC Over Time



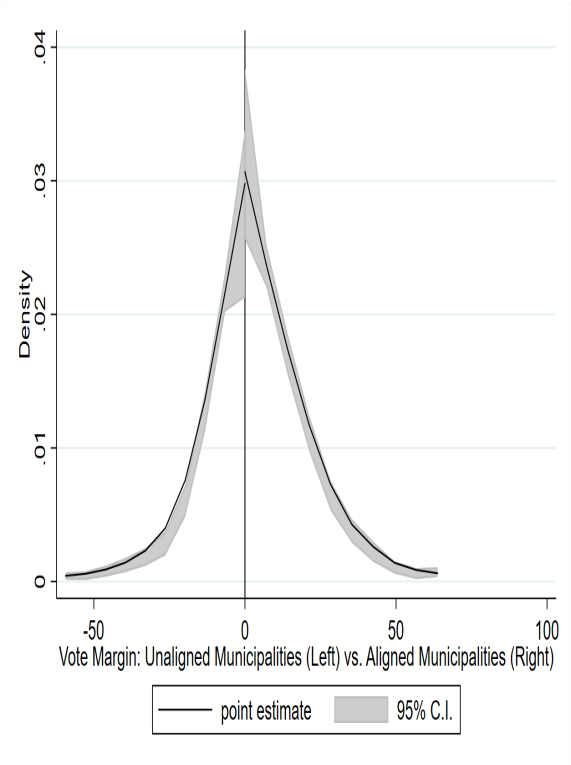
(b) Logit Model with Municipality and Term FE



(c) RD Results (Term)



(d) McCrary Density Plot for RDD



of various institutions in Honduras but, as compared to the other countries in this paper, it performs relatively relatively fewer municipal audits. Honduras has 298 municipalities but only in 2002, when the TSC became a formal SAI, did the country undertake more than 200 audits in a year (see Figure 1a). Since then—and particularly over the course of Juan Orlando Hernandez's two-term presidency (2014-present)—the number of municipal audits undertaken by the TSC has dropped steadily. A caveat is that many audits take place over the course of multiple years, and reports only become available years after the audits ends. Accordingly, the number of audits during the Hernandez Presidency will increase slightly as the TSC releases more multi-year audits reports on its website, but the decline is still significant.

Regardless, the preliminary logit model in Figure 1b shows that partisan motivations are not driving which municipalities receive audits. I only examine mayor-president party alignments because Honduras' president appoints governors from his/her own party, so there is no relevant political variation at the department (state-equivalent) level. Consistent with the framework detailed in Section 3.1, I run a regression discontinuity design along the lines of Brollo and Nannicini (2012) as well. It uses random variation in close elections—using an automatically derived bandwidth following Calonico, Cattaneo and Titiunik (2014)—to assess whether party alignment between the president and mayors is driving audit allocation decisions. Overall, I find the same pattern in Figure 1b and Figure 1c: party alignment does not affect auditing decisions. As shown in Figure 1d, the regression discontinuity analysis passes the McCrary (2008) density test, too, so we can be confident that random variation in close elections is smoothly distributed, and that there are no signs of electoral fraud that could skew the results.

Intensity/Dosage. Unlike the audit data from other countries examined in this paper, Honduras' TSC provides start and end dates for each of its audits. The issue is that the start and end dates often only differ by one day, and the audits often last for multiple years, indicating that only the start dates are meaningful for analysis. In future iterations of the paper, I will conduct some event-history analysis to check for potential biases.

On the qualitative side, the TSC has a very broad code of ethics for its staff (Tribunal Superior de Cuentas de Honduras, 2018), which I will examine in more detail in further iterations of the paper.

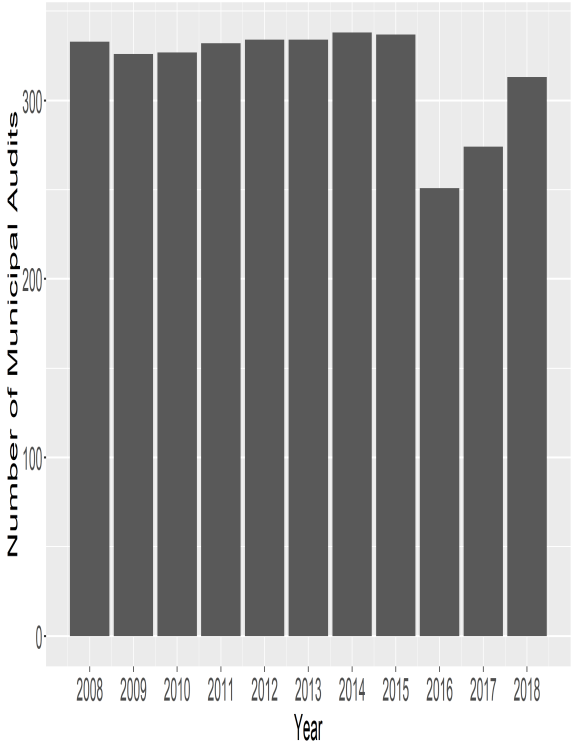
4.3. Guatemala

Legal Basis for Independence from the Executive Branch. The National Audit Office (*Contraloría General de Cuentas*) is the body in charge of undertaking audits within Guatemala. Articles 232-236 of Guatemala's 1985 Constitution provides the basis for the office's independence and ability to audit all uses of public funds throughout the country. Notably, these articles stipulate that Congress, not the President, elects the Comptroller General of Accounts (*Controlador de Cuentas*) to non-renewable, four-year terms. Removing the Comptroller General of Accounts is also uniquely within the purview of the Congress. It can only remove the Comptroller General of Accounts by majority vote only for reasons pertaining to "negligence, crime, and lack of aptitude."

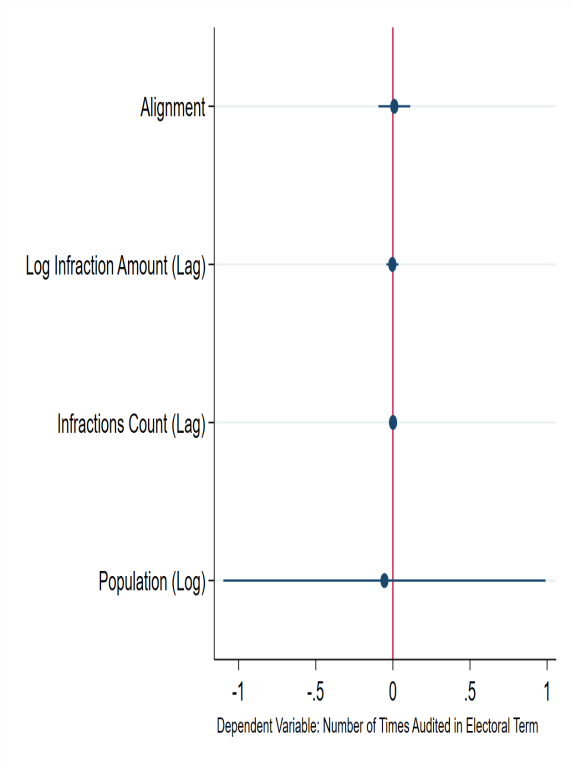
Frequency/Selection of Audits. Each year, the National Audit Office audits circa 320 of Guatemala's 340 municipalities (See Figure 2a). Because there is not much variation is which municipalities are audited each year, in such circumstances it is more appropriate to run a count model, such as poisson regression, instead of a logit model. As shown in Figure 2b, past corruption (lagged infractions and log amounts attached to those infractions), partisanship, and demographics (population) are all similarly poor predictors of how many audits a municipality receives in a given year. Like Honduras, Guatemala also does not have political variation at the department (state-equivalent) level, so it is possible to analyze mayor-president party alignments directly. When switching to a regression discontinuity design in Figure 2c, results similarly are poor due to the lack of variation—more precisely, alignment is not a statistically significant predictor of audits. Since the regression discontinuity estimates pass the McCrary (2008) density test as well (see Figure 2d), it is safe to conclude that the audit distribution is fair.

Figure 2: Main Results: Guatemala's Municipal Audit Distribution

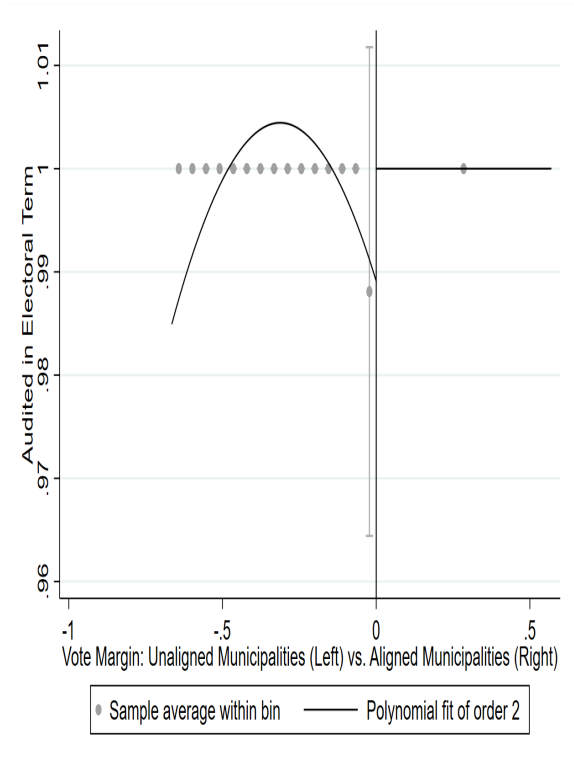
(a) Number of Audits Conducted by Year



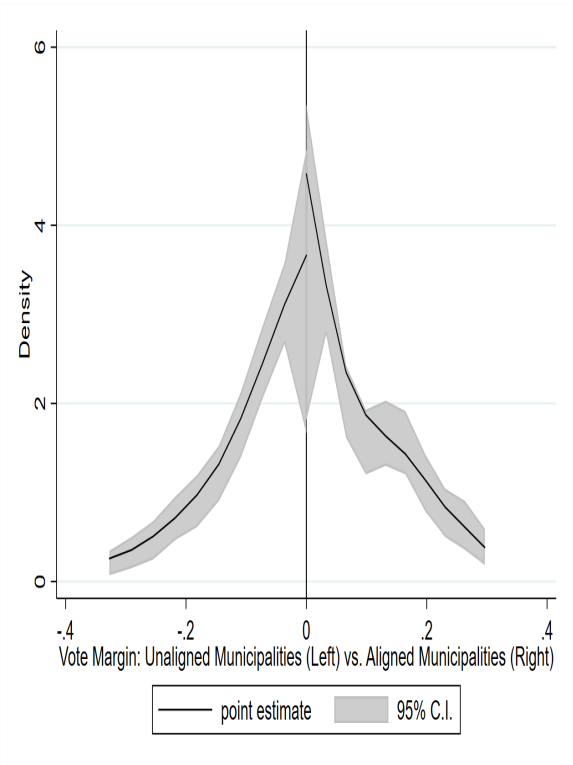
(b) Poisson Model with Municipality and Term FE



(c) RD Results for Audit Count



(d) McCrary Density Plot for RD



Intensity/Dosage. Guatemala's National Audit Office does not provide quantitative data for assessing audit intensity/dosage, but it does have an extensive code of ethics/conduct (Contraloría General de Cuentas de Guatemala, 2018). In future iterations of the paper, I will analyze that code of ethics/conduct in more detail.

4.4. Mexico

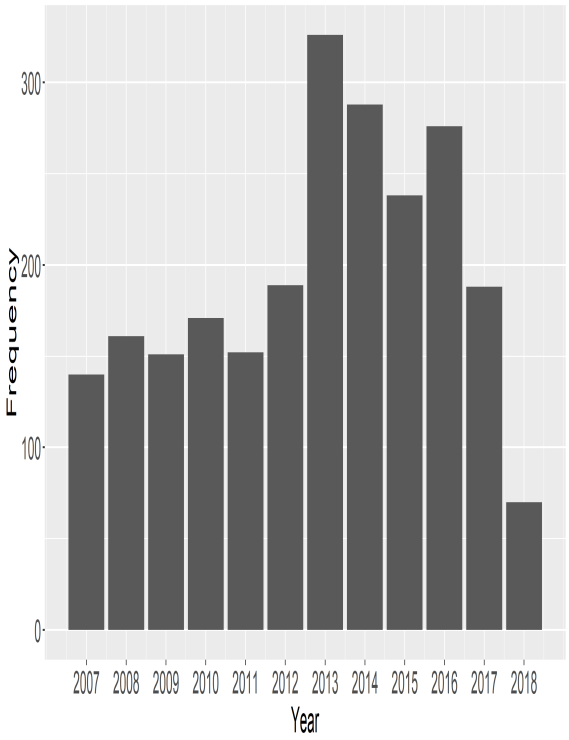
Legal Basis for Independence from the Executive Branch. The Supreme Audit Institution of the Federation (ASF, *Auditoría Superior de la Federación*) is the institution responsible for government audits in Mexico. The ASF receives its mandate directly from Articles 74, 79, and 113 of the Mexican Constitution and reports to the Chamber of Deputies, not the President. The Chamber of Deputies also elects the head of ASF (*Auditor Superior*), who serves a term of 8 years that may be renewed once.

Frequency/Selection of Audits. The ASF does not audit all of Mexico's 2,400 municipalities each year. Instead, the ASF conducts risk-based audits, which are decided upon by the ASF planning commission in consultation with the Chambers of Deputies. In contrast to Guatemala, for which electoral term regressions are more appropriate, year-wise regressions are more appropriate for Mexico. The reason is that, until 2018, Mexico's mayoral, gubernatorial, and presidential elections did not take place at the same time. Therefore, it is most appropriate to analyze partisan alignment patterns on a yearly basis. Unlike Guatemala and Honduras, Mexico has governors from different political parties, so it is necessary to account for that variation. In Figure 3b, a logit model with municipality and year fixed effects suggests that no type of alignment configuration or demographics has any influence on the audit distribution. In the corresponding regression discontinuity design model using three-way party alignment in Figure 3,[32] there is a slight jump on the right-side of the plot, indicating that aligned municipalities are more likely to receive an audit. That jump, however, is not statistically significant, and the McCrary (2008) density test in Figure 3d does not pass. It is thus difficult to conclude that there is any bias in the audit distribution,

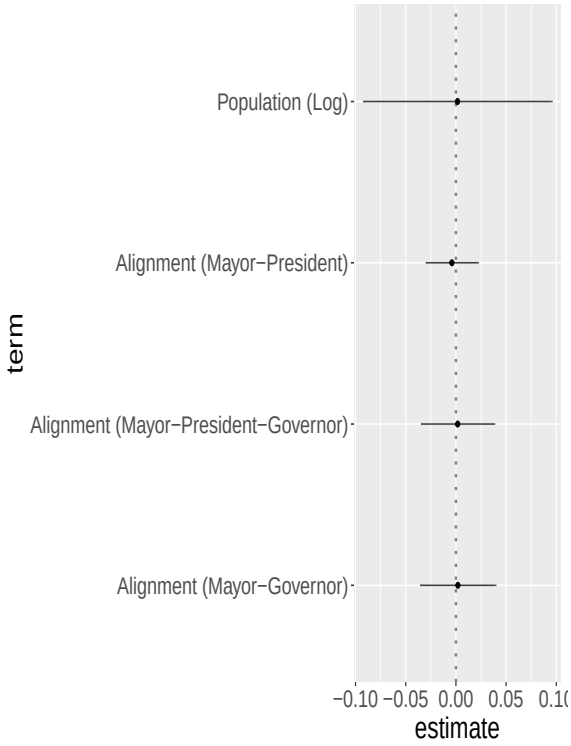
[32] Note: results are similar for the other alignment configurations.

Figure 3: Main Results: Mexico's Municipal Audit Distribution

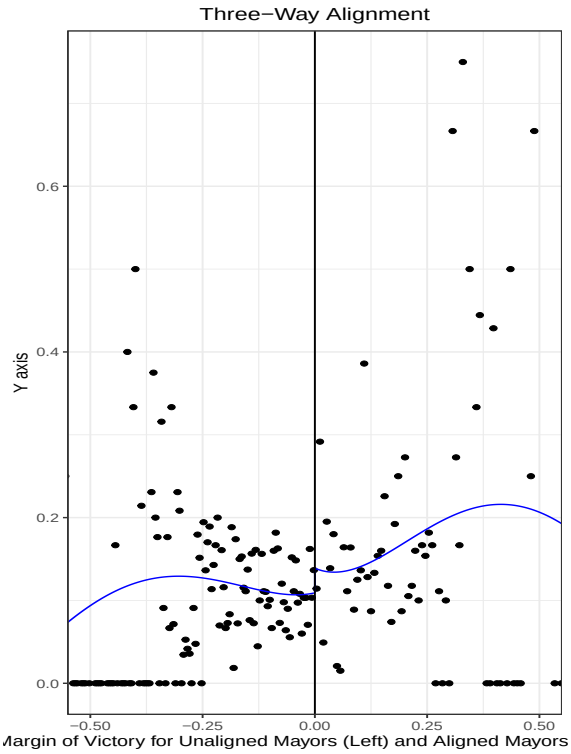
(a) Municipalities Receiving Audits by Year



(b) Logit Model with Municipality and Year FE



(c) RD Results (Year)



(d) McCrary Density Plot for RD

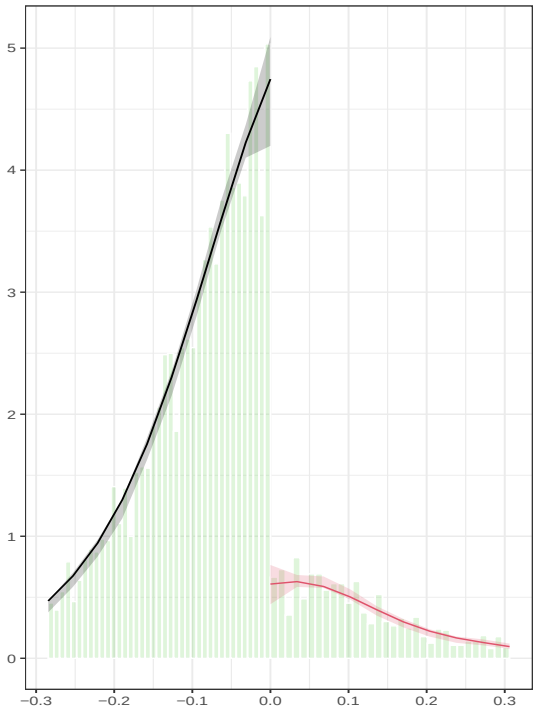
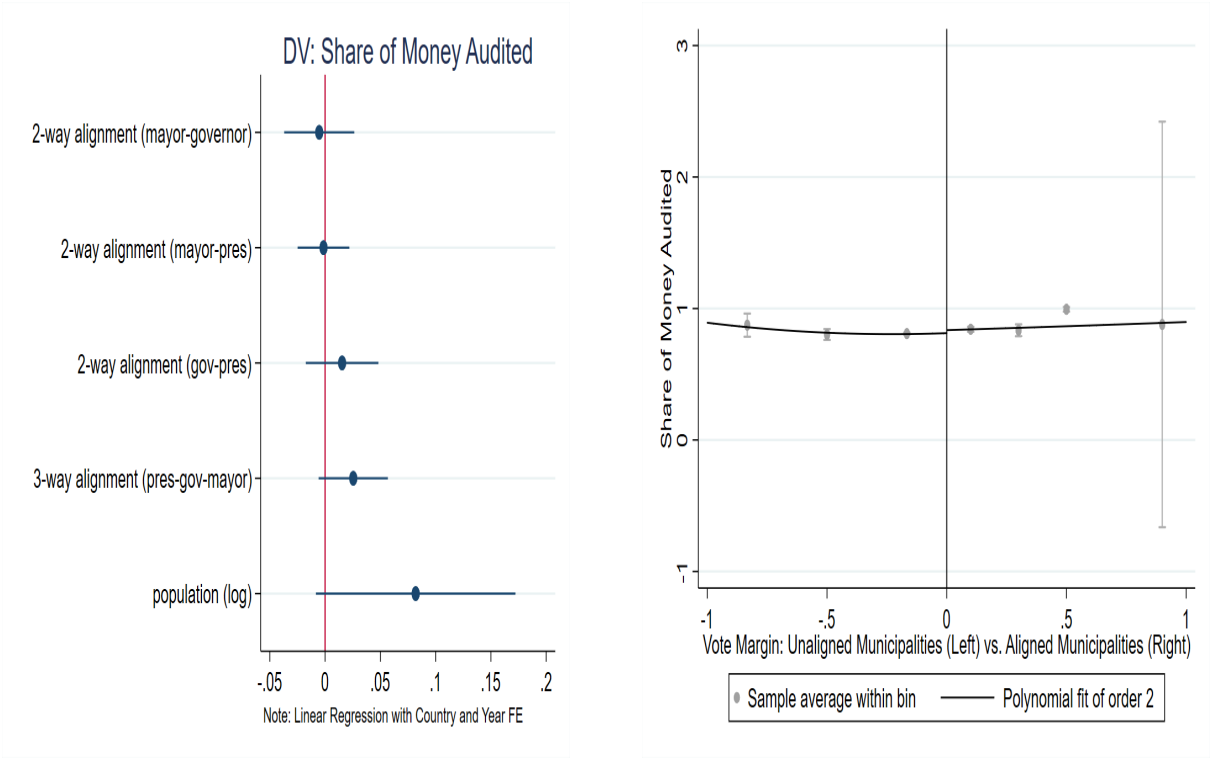


Figure 4: Mexico's Audit Intensity/Dosage

(a) Linear Regression Model with Municipality and Year FE

(b) RDD Results (3-Way Alignment)



though I will perform additional analyses in further iterations of the paper.

Intensity/Dosage. Mexico provides credible quantitative data from which to assess the intensity/dosage of audits. For each audit, Mexico's ASF provides the amount of money in Mexican pesos in the audit sample as well as the overall amount of money that it could have audited. If membership in an opposition party predicts a higher share of funds audited, it suggests that opposition parties are receiving tougher audits. As results from Figure 4 suggest, that does not appear to be the case for Mexico. Both the linear regression model and regression discontinuity approach reach the same conclusion.

Mexico's ASF also has an extensive Integrity Policy, comprising a code of ethics, code of conduct, and directives on conflicts of interest (Auditoría Superior de la Federación de México, 2013). In future iterations, I will analyze the Integrity Policy in further detail.

5. Conclusion

Audit data do not provide the only, more objective alternative to measuring corruption with perceptions data. For example, scholars have convincingly measured corruption using data from the stock market,[33] procurement,[34] asset declarations,[35] taxes,[36] and customs duties.[37] However, audit data standout from the above measures for a simple reason: audits reveal a greater diversity of corrupt activities. Ghost firms, theft, nepotism, fraud, violations related to all of the above measures, and many other types of corruption are all within the purview of audits.

The problem with audit-related corruption measures to date is that, with very few exceptions, scholars have focused almost exclusively on the municipal audit data from Brazil. The numerous Brazil studies using these data have produced some very useful causal findings, notably due to the random assignment of audits to individual municipalities. Nevertheless, the external validity of these studies remains a concern due Brazil's unique institutional features and the fact that Brazil is the only country in the world with randomized audits.

In this article, I put forth new subnational audit data from Mexico, Guatemala, Honduras, and India as well as a new framework for assessing whether researchers can profitably use these audit data and others to analyze corruption. For the framework's first pillar on legal independence from the executive branch, all of the countries in the sample have sufficient independence. Similarly, there do not appear to be any issues with any country in the sample on the second pillar, relating to the partisan distributions of the audits. Analysis of the framework's third pillar on audit intensity/dosage is still ongoing at the time of this writing, but preliminary analyses do not show any signs of bias.

[33] See, for example, Fisman (2001), Faccio (2006), and Fisman and Wang (2015)

[34] See, for example, Bandiera, Prat and Valletti (2009), Mironov and Zhuravskaya (2016), Fazekas (2017), Fazekas, Cingolani and Tóth (2018), Baltrunaite (2019), Broms, Dahlström and Fazekas (2019), and Fazekas and Kocsis (2020).

[35] See, for example, Eggers and Hainmueller (2009) and Fisman, Schulz and Vig (e.g., 2014).

[36] See, for example, Kleven et al. (2011), Khan, Khwaja and Olken (2016), Bilicka (2019), and (Naritomi, 2019).

[37] See, for example, Fisman and Wei (2004, 2009), Sequeira and Djankov (2014), and Rijkers, Baghdadi and Raballand (2017).

Given the results thus far, the above data and framework will help analysts undertake sophisticated analyses of corruption across the world. In the process, researchers will move beyond Brazil, use perceptions-based data less, and better understand the true causes and consequences of corruption.

References

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

Adcock, Robert and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95(3):529–546.

Ajzenman, Nicolás. 2020. "The Power of Example: Corruption Spurs Corruption." *American Economic Journal: Applied Economics* (Forthcoming).

Alm, James, Jorge Martinez-Vazquez and Chandler McClellan. 2016. "Corruption and Firm Tax Evasion." *Journal of Economic Behavior and Organization* 124:146–163.

Andersson, Staffan and Paul M. Heywood. 2009. "The Politics of Perception: Use and Abuse of Transparency International's Approach to Measuring Corruption." *Political Studies* 57(4):746–767.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, New Jersey: Princeton University Press.

Auditoría Superior de la Federación de México. 2013. "Política Institucional de Integridad de la ASF.".

Avis, Eric, Claudio Ferraz and Frederico Finan. 2018. "Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians." *Journal of Political Economy* 126(5):1912–1964.

Baltrunaite, Audinga. 2019. "Political Contributions and Public Procurement: Evidence from Lithuania." *Journal of the European Economic Association* 18(2):541–582.

Bandiera, Oriana, Andrea Prat and Tommaso Valletti. 2009. "Active and Passive Waste in Government Spending: Evidence from a Policy Experiment." *American Economic Review* 99(4):1278–1308.

Bersch, Katherine and Sandra Botero. 2014. "Measuring Governance: Implications of Conceptual Choices." *European Journal of Development Research* 26(1):124–141.

Besley, Timothy and Robin Burgess. 2002. "The Political Economy of Government Responsiveness: Theory and Evidence from India." *Quarterly Journal of Economics* 117(4):1415–1451.

Bilicka, Katarzyna Anna. 2019. "Comparing UK Tax Returns of Foreign Multinationals to Matched Domestic Firms." *American Economic Review* 109(8):2921–2953.

Blossfeld, Hans-Peter, Katrin Golsch and Götz Rohwer. 2007. *Event History Analysis with Stata*. New York: Psychology Press: Taylor and Francis Group.

Bobonis, Gustavo J., Luis R. Cámara Fuertes and Rainer Schwabe. 2016. "Monitoring Corruptible Politicians." *American Economic Review* 106(8):2371–2405.

Brollo, Fernanda and Tommaso Nannicini. 2012. "Tying Your Enemy's Hands in Close Races: The Politics of Federal Transfers in Brazil." *American Political Science Review* 106(4):742–761.

Brollo, Fernanda, Tommaso Nannicini, Roberto Perotti and Guido Tabellini. 2013. "The Political Resource Curse." *American Economic Review* 103(5):1759–1796.

Brollo, Fernanda and Ugo Troiano. 2016. "What Happens When a Woman Wins an Election? Evidence from Close Races in Brazil." *Journal of Development Economics* 122:28–45.

Broms, Rasmus, Carl Dahlström and Mihály Fazekas. 2019. "Political Competition and Public Procurement Outcomes." *Comparative Political Studies* 52(9):1259–1292.

Calonico, Sebastian, Matias D. Cattaneo and Rocío Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82(6):2295–2326.

Cavalcanti, Francisco, Gianmarco Daniele and Sergio Galletta. 2018. "Popularity Shocks and Political Selection." *Journal of Public Economics* 165:201–216.

Comptroller and Auditor General of India. 2012. "Code of Ethics for the IAAD Office of the Comptroller and Auditor General of India.".

Contraloría General de Cuentas de Guatemala. 2018. "Código de Ética de Contraloría General de Cuentas.".

Coppedge, Michael, John Gerring, Adam N. Glynn, Carl Henrik Knutsen, Staffan I. Lindberg, Daniel Pemstein, Brigitte Seim, Svend-Erik Skaaning and Jan Teorell. 2020. *Varieties of Democracy: Measuring Two Centuries of Political Change*. Cambridge: Cambridge University Press.

Dahlström, Carl and Mikael Holmgren. 2019. "The Political Dynamics of Bureaucratic Turnover." *British Journal of Political Science* 49(3):823–836.

Dhaliwal, Iqbal and Rema Hanna. 2017. "The Devil is in the Details: The Successes and Limitations of Bureaucratic Reform in India." *Journal of Development Economics* 124:1–21.

Di Tella, Rafael and Ernesto Schargrodsky. 2003. "The Role of Wages and Auditing during a Crackdown on Corruption in the City of Buenos Aires." *Journal of Law and Economics* 46(1):269–292.

Dizon-Ross, Rebecca, Pascaline Dupas and Jonathan Robinson. 2017. "Governance and the Effectiveness of Public Health Subsidies: Evidence from Ghana, Kenya and Uganda." *Journal of Public Economics* 156:150–169.

Duflo, Esther, Rema Hanna and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102(4):1241–1278.

Eggers, Andrew C. and Jens Hainmueller. 2009. "MPs for Sale? Returns to Office in Postwar British Politics." *American Political Science Review* 103(4):513–533.

Escaleras, Monica, Nejat Anbarci and Charles A. Register. 2007. "Public Sector Corruption and Major Earthquakes: A Potentially Deadly Interaction." *Public Choice* 132(1-2):209–230.

Faccio, Mara. 2006. "Politically Connected Firms." *American Economic Review* 96(1):369–386.

Fariss, Christopher J., Michael R. Kenwick and Kevin Reuning. 2020. Measurement Models. In *SAGE Handbook of Research Methods in Political Science and International Relations*, ed. Luigi Curini and Robert J. Franzese. London: Sage Publications chapter 20, pp. 353–370.

Fazekas, Mihály. 2017. "Red Tape, Bribery and Government Favouritism: Evidence from Europe." *Crime, Law and Social Change* 68(4):403–429.

Fazekas, Mihály and Gábor Kocsis. 2020. "Uncovering High-Level Corruption: Cross-National Objective Corruption Risk Indicators Using Public Procurement Data." *British Journal of Political Science* 50(1):155–164.

Fazekas, Mihály, Luciana Cingolani and Bence Tóth. 2018. Innovations in Objectively Measuring Corruption in Public Procurement. In *Governance Indicators: Approaches, Progress, Promise*, ed. Helmut Anheier, Matthias Haber and Mark Andreas Kayser. Oxford: Oxford University Press chapter 7, pp. 154–185.

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.

Ferraz, Claudio and Frederico Finan. 2011. "Electoral Accountability and Corruption: Evidence from the Audits of Local Governments." *American Economic Review* 101(4):1274–1311.

Ferraz, Claudio and Frederico Finan. 2018. Fighting Political Corruption: Evidence from Brazil. In *Institutions, Governance, and the Control of Corruption*, ed. Kaushik Basu and Tito Cordella. Cham, Switzerland: Palgrave Macmillan chapter 9, pp. 253–284.

Ferraz, Claudio, Frederico Finan and Diana B. Moreira. 2012. "Corrupting Learning: Evidence from Missing Federal Education Funds in Brazil." *Journal of Public Economics* 96(9-10):712–726.

Findley, Michael G., Kyosuke Kikuta and Michael Denly. 2021. "External Validity." *Annual Review of Political Science* .

Fisman, Raymond. 2001. "Estimating the Value of Political Connections." *American Economic Review* 91(4):1095–1102.

Fisman, Raymond, Florian Schulz and Vikrant Vig. 2014. "The Private Returns to Public Office." *Journal of Political Economy* 122(4):806–862.

Fisman, Raymond and Miriam A. Golden. 2017. *Corruption: What Everyone Needs to Know*. Oxford: Oxford University Press.

Fisman, Raymond and Shang-Jin Wei. 2004. "Tax Rates and Tax Evasion: Evidence from Missing Imports in China." *Journal of Political Economy* 112(2):471–496.

Fisman, Raymond and Shang-Jin Wei. 2009. "The Smuggling of Art, and the Art of Smuggling: Uncovering the Illicit Trade in Cultural Property and Antiques." *American Economic Journal: Applied Economics* 1(3):82–96.

Fisman, Raymond and Yongxiang Wang. 2015. "Corruption in Chinese Privatizations." *Journal of Law, Economics, and Organization* 31(1):1–29.

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

Gerring, John and Strom C. Thacker. 2004. "Political Institutions and Corruption: The Role of Unitarism and Parliamentarism." *British Journal of Political Science* 34(2004):295–330.

Gerring, John and Strom C. Thacker. 2005. "Do Neoliberal Policies Deter Political Corruption?" *International Organization* 59(01):233–254.

Gingerich, Daniel W. 2013*a*. "Governance Indicators and the Level of Analysis Problem: Empirical Findings from South America." *British Journal of Political Science* 43(July 2013):505–540.

Gingerich, Daniel W. 2013*b*. *Political Institutions and Party-Directed Corruption in South America: Stealing for the Team*. New York: Cambridge University Press.

Gisselquist, Rachel M. 2014. "Developing and Evaluating Governance Indexes: 10 Questions." *Policy Studies* 35(5):513–531.

Hollyer, James R. 2018. Measuring Governance: Objective and Perceptions-Based Governance Indicators. In *Governance Indicators: Approaches, Progress, Promise*, ed. Helmut Anheier, Matthias Haber and Mark Andreas Kayser. Oxford: Oxford University Press chapter 5, pp. 103–133.

Imbens, Guido W. and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. New York: Cambridge University Press.

International Organization of Supreme Audit Institutions. 1977. "The Lima Declaration.".

International Organization of Supreme Audit Institutions. 2007. "Mexico Declaration on Supreme Audit Institutions (SAI) Independence.".

International Organization of Supreme Audit Institutions. 2019. "Independence of Supreme Audit Institutions.".
URL: *https://www.intosai.org/fileadmin/downloads/focus_areas/independence/EN_0219_Handout_Indep*

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi. 2011. "The Worldwide Governance Indicators: Methodology and Analytical Issues." *Hague Journal on the Rule of Law* 3(2):220–246.

Keefer, Philip. 2007*a*. "Clientelism, Credibility, and the Policy Choices of Young Democracies." *American Journal of Political Science* 51(4):804–821.

Keefer, Philip. 2007*b*. The Poor Performance of Poor Democracies. In *Oxford Handbook of Comparative Politics*, ed. Carles Boix and Susan C. Stokes. Oxford: chapter 36, pp. 886–909.

Khan, Adnan Q., Asim Ijaz Khwaja and Benjamin A. Olken. 2016. "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors." *Quarterly Journal of Economics* pp. 219–271.

King, Gary. 1998. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Ann Arbor, Michigan: University of Michigan Press.

King, Gary, Robert Keohane and Sydney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, New Jersey: Princeton University Press.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen and Emmanuel Saez. 2011. "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark." *Econometrica* 79(3):651–692.

Kunicová, Jana and Susan Rose-Ackerman. 2005. "Electoral Rules and Constitutional Structures as Constraints on Corruption." *British Journal of Political Science* 35(May):573–606.

Kurtz, Marcus J. and Andrew Schrank. 2007*a*. "Growth and Governance: A Defense." *Journal of Politics* 69(2):563–569.

Kurtz, Marcus J. and Andrew Schrank. 2007*b*. "Growth and Governance: Models, Measures, and Mechanisms." *Journal of Politics* 69(2):538–554.

Langbein, Laura and Stephen Knack. 2010. "The Worldwide Governance Indicators: Six, One, or None?" *Journal of Development Studies* 46(2):350–370.

Larreguy, Horacio, John Marshall and James M. Snyder. 2020. "Publicizing Malfeasance: When Media Facilitates Electoral Accountability in Mexico." *Economic Journal* .

Lederman, Daniel, Norman V. Loayza and Rodrigo R. Soares. 2005. "Accountability and Corruption: Political Institutions Matter." *Economics and Politics* 17(1):1–35.

Locke, Richard M. and Kathleen Thelen. 1995. "Apples and Oranges Revisited: Contextualized Comparisons and the Study of Comparative Labor Politics." *Politics and Society* 23(3):337–367.

Lührmann, Anna, Kyle L. Marquardt and Valeriya Mechkova. 2020. "Constraining Governments: New Indices of Vertical, Horizontal, and Diagonal Accountability." *American Political Science Review* 114(3):811–820.

Marquardt, Kyle L. and Daniel Pemstein. 2018. "IRT Models for Expert-Coded Panel Data." *Political Analysis* 10(4):431–456.

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

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142(2):698–714.

Melo, Marcus André, Carlos Pereira and Carlos Mauricio Figueiredo. 2009. "Political and Institutional Checks on Corruption: Explaining the Performance of Brazilian Audit Institutions." *Comparative Political Studies* 42(9):1–28.

Mironov, Maxim and Ekaterina Zhuravskaya. 2016. "Corruption in Procurement and the Political Cycle in Tunneling: Evidence from Financial Transactions Data." *American Economic Journal: Economic Policy* 8(2):287–321.

Mungiu-Pippidi, Alina. 2015. *The Quest for Good Governance: How Societies Develop Control of Corruption*. New York: Cambridge University Press.

Naritomi, Joana. 2019. "Consumers as Tax Auditors." *American Economic Review* 109(9):3031–3072.

Nikolova, Elena and Nikolay Marinov. 2017. "Do Public Fund Windfalls Increase Corruption? Evidence From a Natural Disaster." *Comparative Political Studies* 50(11):1455–1488.

O'Donnell, Guillermo. 1998. "Horizontal Accountability in New Democracies." *Journal of Democracy* 9(3):1112–1126.

Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115(21):200–249.

Papke, Leslie E. and Jeffrey M. Wooldridge. 2008. "Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates." *Journal of Econometrics* 145(1-2):121–133.

Pereira, Carlos and Marcus André Melo. 2015. "Reelecting Corrupt Incumbents in Exchange for Public Goods: Rouba mas faz in Brazil." *Latin American Research Review* 50(4):88–115.

Rijkers, Bob, Leila Baghdadi and Gael Raballand. 2017. "Political Connections and Tariff Evasion: Evidence from Tunisia." *World Bank Economic Review* 31(2):459–482.

Rose-Ackerman, Susan and Bonnie Palifka. 2016. *Corruption and Government: Causes, Consequences, and Reform*. Second ed. New York: Cambridge University Press.

Sartori, Giovanni. 1984. *Social Science Concepts: A Systematic Analysis*. Beverly Hills: Sage Publications.

Schaffer, Frederic Charles. 2016. *Elucidating Social Science Concepts: An Interprevist Guide*. New York: Routledge.

Schwindt-Bayer, Leslie A. and Margit Tavits. 2016. *Clarity of Responsibility, Accountability, and Corruption*. New York: Cambridge University Press.

Sequeira, Sandra and Simeon Djankov. 2014. "Corruption and Firm Behavior: Evidence from African Ports." *Journal of International Economics* 94(2):277–294.

Snyder, James M. and David Strömberg. 2010. "Press Coverage and Political Accountability." *Journal of Political Economy* 118(2):355–408.

Standaert, Samuel. 2015. "Divining the Level of Corruption: A Bayesian State-Space Approach." *Journal of Comparative Economics* 43(3):782–803.

Stanig, Piero. 2018. Considerations on the Method of Constructing Governance Indices. In *Governance Indicators: Approaches, Progress, Promise*, ed. Helmut Anheier, Matthias Haber and Mark Andreas Kayser. Oxford: Oxford University Press chapter 6, pp. 134–153.

Stegmueller, Daniel. 2011. "Apples and Oranges? The Problem of Equivalence in Comparative Research." *Political Analysis* 19(4):471–487.

Thomas, Melissa A. 2010. "What Do the Worldwide Governance Indicators Measure?" *European Journal of Development Research* 22(1):31–54.

Treisman, Daniel. 2000. "The Causes of Corruption: A Cross-National Study." *Journal of Public Economics* 76(3):399–457.

Treisman, Daniel. 2007. "What Have We Learned About the Causes of Corruption from Ten Years of Cross-National Empirical Research?" *Annual Review of Political Science* 10(1):211–244.

Tribunal Superior de Cuentas de Honduras. 2018. "Código de Conducta Ética del Tribunal Superior de Cuentas de Honduras.".

Trochim, William. 2006. "The Research Methods Knowledge Base, 2nd Edition.".
URL: *http://www.socialresearchmethods.net/kb/*

World Bank. 2016. "Guidelines on Preventing and Combating Fraud and Corruption in Projects Financed by IBRD Loans and IDA Credits and Grants.".

Zamboni, Yves and Stephan Litschig. 2018. "Audit Risk and Rent Extraction: Evidence from a Randomized Evaluation in Brazil." *Journal of Development Economics* 134(April):133–149.