The promise and perils of data for anti-corruption efforts in international development work

By Daniel Berliner and Kendra Dupuy

Big and open data sources can empower development practitioners and aid recipients to reduce the harmful effects of corruption in development. Data on public procurement, campaign contributions, asset disclosure and other data types can be useful in this regard. Find out how you may apply such data for anti-corruption efforts, but be aware of challenges to using data effectively. This includes disclosure thresholds, technical complexities, a need to interpret the data, and a lack of consistent standards for dataset construction.

Main points

- Analysing detailed administrative data can enable anti-corruption practitioners to identify individual instances of corruption.
- It can also help governments, practitioners, and donors measure the prevalence of corruption and compare outcomes over time and across jurisdictions.
- Finally, it can also allow donors and practitioners to evaluate the impact of particular interventions on corruption more accurately.
- Key challenges to using data for these purposes include the credibility, clarity, and complexity of the data, but also the need to interpret the data to derive specific conclusions from it and differences in cultural and legal contexts that frustrate easy comparisons across jurisdictions.

Table of contents

About the authors

Daniel Berliner

Dr. Daniel Berliner is Assistant Professor of Political Science and Public Policy at the Department of Government, London School of Economics.

Kendra Dupuy

Dr. Kendra Dupuy is a political economist working on research about natural resource and energy management, the education sector, and civil society. Formerly a Senior Adviser at U4, she is currently an affiliated researcher at the Chr. Michelsen Institute.

Big data analysis as a tool for international development work

Increasingly, governments and citizens in developing countries as well as development agencies are using information technology to improve governance, enhance public service delivery, shape government-citizen relations, and reduce corruption. In particular, "big data" and "open data" on governance represent important new sources of information for anti-corruption work. However, we are just at the beginning phases of understanding how to best use these new data sources in anti-corruption work, as well as appreciating the challenges and limitations inherent in them.

Our focus in this brief is on the use of detailed administrative data that pertains to activities and individuals involved in governance, and that is suitable for use in identifying or evaluating corruption. Often, but not always, these data sources are *open* – meaning they can be freely used, reused, and redistributed by anyone, anytime, anywhere[1]. And often, but not always, these data sources are *big*. Big data has four characteristics:

1. Size or volume
2. The speed or velocity at which it is generated
3. The variety of forms in which the data is generated (ranging from photos to text messages to social media posts)
4. The complexity of the information being created, in terms of the ability to interconnect different data sources.[2]

However, we do not restrict our scope here to only big or open data. Often, governments and other organisations keep data out of the public realm. And often, highly relevant data for anti-corruption efforts pertain to "medium" numbers of observations, such as asset disclosures of just several thousand politicians. Yet in recent years, these forms of data have also become increasingly available or "open."

---

1. https://opendatacharter.net/principles/
2. Desouza and Jacob 2017

Our specific focus here is on administrative data that is sufficiently detailed. Data must be granular and broken-down to be of any use for anti-corruption purposes. We further focus on data from official sources, rather than data collected by private actors or crowd-sourced from the public, as such sources raise additional questions regarding sampling and representativeness.

Several global trends have contributed to the increasing accessibility and availability of data relevant for anti-corruption efforts. More and more governments have adopted e-government[3], transparency, and open data projects in recent years. New technology and global initiatives like the Open Government Partnership[4], the Open Contracting Partnership[5], and the Extractive Industries Transparency Initiative[6] have helped to make such projects possible. These developments are pushing governments around the world – including aid-receiving governments – to shift towards a more technology-centered way of working and interacting with citizens, including the collection, generation, and publication of many new data sources. Below, we review several of these sources with applications to anti-corruption efforts.

---

3. https://en.wikipedia.org/wiki/E-government
4. https://www.opengovpartnership.org/
5. https://www.open-contracting.org/
6. https://eiti.org/

Types of data for anti-corruption efforts

Public procurement data

As much of the world's corruption takes place through public contracts,[7] using data to better monitor procurement and detect corruption offers tremendous promise. For example, procurement data has been used to detect networks of politicians and companies at high risk of corruption,[8] measure corruption risk in contracting across European countries,[9] and link campaign contributions with "quid pro quo" contract awards in Brazil and Colombia.[10] In Paraguay, a new open procurement data platform led to public reporting on inflated education contracts, major protests, and the resignation of a minister.[11] As e-procurement systems become more common around the world, and standards like the Open Contracting Partnership become more prominent, the usefulness of these data sources will only increase.

Campaign contributions data

When financial contributions are made to campaigns for corrupt purposes, they may be made implicitly or explicitly in exchange for past or future considerations. Relevant data sources – such as the Campaign Finance website for U.S. elections[12] – thus offer the potential for linking these contributions with their "quid pro quo," often by studying them in conjunction with data on public contracts or hiring. For example, Boas et al (2014)[13] show that campaign finance contributors specialising in public works contracts receive a boost in contracts post-election when they donate to a winning ruling party candidate's campaign.

---

7. Lewis-Faupel et al 2016

8. Fazekas and Tóth 2014

9. Charron et al 2017

10. Arvate et al 2013; Ruiz 2018

11. Paraguay's transparency alchemists. How citizens are using open contracting to improve public spending
https://medium.com/open-contracting-stories/paraguays-transparency-alchemists-623c8e3c538f

12. http://campaign-finance.org/

13. https://www.journals.uchicago.edu/doi/abs/10.1017/s002238161300145x

Asset disclosure data

A third of countries worldwide require public data on individual financial assets. These data are useful for identifying politicians whose assets increased dramatically while in office. Researchers in India have carried out analyses of such data and have found that election winners' assets increased by up to 5% compared to runners-up.[14]

Budget and spending data

Data on government budgets and spending can be another powerful anti-corruption tool. Comparisons between funds allocated and results on the ground often provide a means of identifying funds lost to corruption in sectors such as the health and education sectors.[15] Broader analyses of resource allocation patterns can also identify policy biases in favour of certain groups.

Several countries, cities, and municipal governments have made their budgets and government spending information publicly available. Examples include the Government of the United States; the Oakland, California city government; and the Government of Nepal. Initiatives like the International Budget Partnership also contribute to efforts to make more – and better – data of this sort publicly available.

Bureaucratic personnel and responsiveness data

Governments are increasingly making data on the behaviour and characteristics of public servants publicly available. Such data can be particularly useful in identifying patterns of patronage hiring or political favouritism.[16] This also includes data on bureaucratic responsiveness (how bureaucrats carry out their duties), for instance on responses to freedom of information requests in Mexico;[17] police event records;[18] as well as on public servant social networks.[19]

---

14. Fisman et al 2013; Bhavnani 2012
15. Olken 2007
16. Bersch et al 2017; Brierly 2018; Colonnelli 2018; Pierskalla and Sacks 2018.
17. Bagozzi et al 2016
18. Carton et al 2016
19. Fafchamps and Labonne 2017

Aid data

Development agencies themselves now make substantial data on their activities available, with help from standards like the International Aid Transparency Initiative (IATI)[20] and projects like AidData.org[21]. Several bilateral aid agencies that are U4 partners are also moving towards making more of their (often, big) data open, including Sweden[22] and Denmark[23]. Multilateral financial institutions like the World Bank are also employing more technological tools to improve data collection around corruption in aid projects as well as to enhance government accountability. For instance, the World Bank-funded Andhra Pradesh Rural Poverty Reduction Project used village-level mobile phone booths to enable people to report on funding misuse as part of its overall complaint handling strategy.

Transnational contracts and ownership data

The previous examples have all been from the public (government) sector. But the data revolution has also impacted the private sector. For instance, the Extractive Industries Transparency Initiative[24], an international, multi-stakeholder voluntary effort in the oil, mining, and gas sectors, requires sector operators (often private companies) to publish what they pay to governments in the form of royalties, taxes, and other payments, and for governments to publish the payments they receive from sector operators. EITI now collects and publishes a variety of country-level data on the extractive industries, and is pushing for disclosure of beneficial ownership of extractive industries companies. By 2020, all EITI countries will be required to disclose the owners of extractive industries companies who benefit from contracts in their countries. This initiative has recently received backing in law in the United Kingdom, which in 2016 required that UK-incorporated companies disclose the ultimate beneficial owners of a company.[25]

---

20. https://www.aidtransparency.net/

21. AidData.org

22. https://openaid.se/aid/2017/

23. http://openaid.um.dk/en

24. https://www.eiti.org

25. See Library of Congress. Disclosure of beneficial ownership for information about other countries with beneficial ownership legal requirements.
https://www.loc.gov/law/help/beneficial-ownership/chart.php

Uses of data for anti-corruption efforts

All of these types of data offer many different types of uses. We suggest several possible approaches to using data for anti-corruption, organised by their *goal* and their *scope*, as summarised in Table 1. Across all of these, the users may be government entities, law enforcement or judicial authorities, individual members of the public, civil society groups, journalists, commercial bodies, or international organisations, or bilateral aid donors.

Some users will simply aim to *identify individual instances of corruption*, either to pursue legal remedy, to inform the public, or to bring political or media attention to a case. The ability to search these data sources for specific politicians, projects, or companies can be very important tools for citizens, journalists, and both governmental and non-governmental investigators. Even just identifying instances of corruption risk can support investigative efforts, or even enable automated compliance checks in contexts like public procurement.

Other users will aim to *measure the prevalence of corruption* more broadly, potentially involving *comparisons across jurisdictions or over time*. Such uses are generally more demanding in terms of technical and statistical skills, as well as knowledge of legal and procedural contexts. Nonetheless, governments, international organisations, and the media regularly seek out this sort of information. Tracking changes over time and benchmarking them against other data are particularly useful both for donors and for domestic audiences. Of course, such comparisons require both that data is available over time or across jurisdictions and that the data sources share common structure and definitions to make comparison possible and meaningful.

Combining comparisons over time and across entities with information on particular anti-corruption efforts, may help evaluate the effectiveness of such efforts.

Finally, some users can further seek to build on these measures in order to *identify the impact of a particular anti-corruption intervention*. Combining

comparisons over time and across entities with information on where and when particular anti-corruption efforts took place, policymakers may be able to better evaluate the effectiveness of such efforts. In the specific case of randomised controlled trials, this means using anti-corruption data to generate outcome variables for measurement and comparison across treated and untreated units for an intervention, such as trainings, audits, or community monitoring efforts.

Depending on the goal, each of these types of uses can also proceed with different scopes of data sources used. Some uses limit themselves to *individual entries* in single data sources, such as the assets of one specific politician, or the value of one specific contract. Other uses employ the *analysis of a full dataset*, such as the identification of suspicious patterns across all procurement data in a given jurisdiction, or the measurement of rates of suspected election fraud across all precincts in a given locale. Finally, the most promising uses generally involve *cross-referencing across multiple sources* of data. For example, patterns of politicians receiving campaign contributions from firms that later receive public works contracts may be indicative of underlying corruption or even constitute proof of illegal activity, depending on the legal context. Similarly, beneficial ownership data in global settings allows users to link it with other sources of information on financial flows that might together help better identify instances of tax evasion or transnational corruption.

Table 1: Possible approaches to using data for anti-corruption

Goal of analysis	Scope of analysis		
	Individual data entries	Analysis of a full dataset	Cross-referencing multiple datasets
Identifying instances of corruption	Investigating specific individuals, firms, contracts, etc.	Identifying anomalous patterns or extreme values	Identifying instances of quid-pro-quo
Measuring prevalence of corruption	Combining multiple specific investigations	Measuring frequency of improper or suspicious values	Measuring frequency of improper or suspicious linkages
Comparison across entities or time		Comparing values across entities in order to produce rankings; or over time in order to assess trends.	Comparing linkages across entities in order to produce rankings; or over time in order to assess trends.
Evaluating impacts on corruption		Comparing values across treated/untreated units or before/after intervention	Comparing linkages across treated/untreated units or before/after intervention

The "five Cs": Key challenges of data for anti-corruption efforts

While open and big data sources represent a major advance in identifying and preventing corruption, these new data sources also come with challenges that that donors, advocates, and users must consider. Below, we outline "five C's" of these challenges.

Credibility

The first challenge is the credibility of data. Information may be inaccurate or incomplete. Data from government sources may be strategically missing or misleading. Analysts must consider the organisational and political processes behind the data. Have the relevant actors fully complied with data collection or disclosure requirements? Are the relevant categories applied correctly? If there are thresholds above which disclosure is required – for example for public contracts or gifts to officials – is there reason to expect strategic behaviour to stay below the threshold?

*Analysts must consider the organisational and political
processes behind the data.*

Some of these issues may be matters of limited technical capacity or human
error. But there may also be strategic decision-making behind *which* datasets
are made public and which are not, or behind which types of information are
included in, or absent from, disclosure requirements. It is essential to consider a
political economy perspective[26] and question why political actors would
willingly expose themselves to additional scrutiny.

Examples of such credibility problems abound. Scholars have found strategic
sorting around key thresholds in cases such as procurement publicity.[27] The
initial disclosures filed in the United Kingdom's much-hailed beneficial
ownership registry were found to contain large numbers of errors, including
impossible numeric values and "over 500 spellings of 'British' ".[28] In Mexico,
responses to information requests are sometimes officially recorded as "the
information has been delivered electronically" when the actual response is a
refusal to provide the information.[29]

---

26. Berliner 2014; Berliner and Erlich 2015;

27. Palguta and Pertold 2017; Fazekas and Toth 2017

28. Learning the lessons from the UK's public beneficial ownership register. OpenOwnership and Global
Witness Joint Briefing. 2017.
https://www.globalwitness.org/en/campaigns/corruption-and-money-laundering/learning-lessons-uks-
public-beneficial-ownership-register/

29. Fox and Haight 2011

Clarity

The second challenge is the clarity of data. Large government data sources frequently suffer from issues of format and presentation that can make it difficult, if not impossible, to actually use them for purposes of pursuing accountability. These include unclear or undefined variable names, missing identification codes that are necessary for cross-referencing with other data sources, and inconsistent spelling. Sometimes data is not sufficiently granular to be relevant for anti-corruption uses, such as when individual records are aggregated by sector or year. Moreover, information may not be provided in machine-readable format, for example when data is contained in tables in PDF documents. In other cases, data is housed in confusing or time-consuming interfaces that make it difficult to find, search, or to know what any given file actually contains. For example, in Ukraine, civil society groups noted difficulties in using a company ownership registry, including that the lack of open format meant the data was "cumbersome for analysis," and could not "be linked with information from other databases and registries."[30]

As an illustration, consider two hypothetical countries, both with publicly available data sources on campaign contributions, public contracts, and local government spending. In the first country, these data sources all use an identical set of codes or consistently spelled names to identify politicians, firms, and localities. But in the second country, no unique codes exist to identify any of these entities, and entries in each data source use different reference codes and formats for presenting names and other identifiers, in some cases making it impossible to differentiate entities with similar names, or to match the same entity with alternate spellings. These seemingly mundane issues of clarity will actually produce high barriers – often prohibitively high – to the use of data for accountability in the second country.

Complexity

The third challenge is the complexity of data. Effective usage of data for accountability requires both a high level of technical skill and a high level of

---

30. Ukraine civil society pinpoints obstacles to beneficial ownership clarity. Natural Resource Governance Institute.
https://resourcegovernance.org/blog/ukraine-civil-society-pinpoints-obstacles-beneficial-ownership-clarity

familiarity with details of the political, legal, and procedural context at hand. Users that lack one of these are likely to face substantial barriers to using open and big data sources for anti-corruption work, and they are highly likely to draw incorrect or misleading conclusions.

Uses that go beyond simply identifying individual instances of corruption will require familiarity with statistical methods and concepts, and many uses require work in statistical software environments like Stata, R, or Python. Such skills are particularly relevant where data is messy and requires substantial data cleaning or processing in order to be usable. However, technical skills alone are not on their own sufficient. Using data for anti-corruption purposes requires a high degree of legal and contextual familiarity, both in order to know what to analyse, what the results mean, and what the legal implications might actually be. Such context can be difficult for non-specialists to learn. For example, Paraguay's procurement data portal initially found that the "only people really using the information were suppliers, economists and computer scientists." [31]

Often, the twin challenges of complexity mean that only large and well-resourced organisations can afford to pursue anti-corruption data analyses. In some cases, no individuals or organisations have yet attempted to use available data for anti-corruption purposes. Thus it is important to build the capacity of actual and potential users of data. There are also substantial opportunities available for creating innovative teams that combine different types of specialists.

Conclusiveness

The fourth challenge is conclusiveness: the ability to know whether a pattern is actually indicative of corruption or not. How does an analyst know what patterns actually indicate corruption, unless they already have some validated "ground truth"? Otherwise, seemingly suspicious patterns in the data may actually have perfectly mundane procedural explanations.

---

31. Paraguay's transparency alchemists. How citizens are using open contracting to improve public spending
https://medium.com/open-contracting-stories/paraguays-transparency-alchemists-623c8e3c538f

As noted by a previous analysis of open data for anti-corruption[32], "open data shows only formal aspects" of records like procurement, yet these are "usually correct in corruption schemes," in order to divert scrutiny. Confidently identifying instances of corruption using large-scale data often requires first comparing that data with already-verified instances from some other source of "ground truth" information, such as legal cases or in-depth media investigations. Indeed, conventional machine-learning approaches are based on the identification of patterns that match an already-verified "training set" of cases. Cantu (2014)[33], for example, first hand-coded over 1,000 vote tallies before using these to train a convolutional neural network classifier to identify likely aggregation fraud in over 50,000 other tallies. Alternately, some specific combinations of data may allow analysts to draw reasonable conclusions as to patterns that would only be present in the case of corruption – such as between campaign contributions and public contracts following elections.

Yet in some cases, the same data patterns might suggest very different conclusions in different contexts or to different analysts. Both Fazekas and Kocsis (2017)[34] and Popa (2018)[35] use "big data" on European public procurement with an explicit goal of understanding corruption patterns. But both face the challenge that the same discretionary procedures might indicate corruption risk in some contexts, yet the reverse in others. Thus Fazekas and Kocsis (2017) see the use of different discretionary procedures as "red flags" in different countries, and Popa (2018) finds several discretionary procedures as indicative of a *stronger* institutional environment that enables public officials to be entrusted with greater procedural flexibility. Without the ability to systematically compare "suspicious" indicators with confirmed cases of corruption, such inconclusiveness can be difficult to resolve.

---

32. https://www.transparency.it/wp-content/uploads/2015/09/2015-TACOD-REPORT.pdf

33. https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12097

34. https://www.cambridge.org/core/journals/british-journal-of-political-science/article/uncovering-highlevel-corruption-crossnational-objective-corruption-risk-indicators-using-public-procurement-data/8A1742693965AA92BE4D2BA53EADFDF0

35. https://www.cambridge.org/core/journals/european-political-science-review/article/what-do-good-governments-actually-do-an-analysis-using-european-procurement-data/3F6A04C8F80A05708C03636A1DFFCA37

Another example comes from Bagozzi et al's 2016 analysis of access-to-information requests in Mexico[36]. Although their semi-supervised learning[37] approach can identify clusters of requests that receive suspicious patterns of denials, it still requires in-depth contextual knowledge to understand whether these are appropriate applications of legal exemptions or a misuse of discretion to avoid disclosure.

Comparability

The final challenge is comparability. Even where the other challenges are successfully overcome or ameliorated in any given context, data may still not be appropriate for comparison *across* governments or organisations, or over time. This is particularly relevant given that the goals of many donors and advocacy groups are to produce global measures or rankings of corruption. Varying legal frameworks, policy models, and even data structures across countries or other jurisdictions tend to hinder comparability, or even render it impossible. Details often vary as to the availability of key variables, the features that are measured, how the relevant policy domain operates, and the legal details of what constitutes corruption. These differences often mean that any country-level summary measures derived from national-level data sources cannot be directly compared.

In the case of public procurement, for example, Fazekas and Kocsis (2017) have been able to produce comparable measures of corruption risk across European countries only because of the shared legal, policy, and data structures created by a common EU standard. Nevertheless, organisations like Open Contracting[38] are making progress in promoting common standards more broadly.

---

36. https://www.benjaminbagozzi.com/uploads/1/2/5/7/12579534/poltext2016-finalproceedings.16-21.pdf

37. https://en.wikipedia.org/wiki/Semi-supervised_learning

38. http://www.open-contracting.org

Policy and advocacy specialists have confronted this challenge in other domains as well. In the realm of access-to-information, advocates have produced comparable measures of *de jure*[39] *information rights in law*[40], but never any measure of *de facto* compliance with those rights in practice. Despite substantial demand for such a measure, it has remained impossible to arrive at a common measurement scheme across countries, given widely varying differences in procedures, legal requirements, and monitoring regimes.

Even within countries, analysts must be wary of how differences across public authorities may confound attempts at comparison. Data from one ministry or province may appear to show more irregularities than another, in some cases not because the true degree of corruption or non-responsiveness is higher, but rather because that ministry or province faces a fundamentally different "caseload" of projects, contracts, requests, etc.

Recommendations for donors

In order for data to have a positive impact on reducing corruption in development aid, donors must continue to support efforts to generate, store, use, and publicly release data, both within their own agencies but also within the countries to which they provide support.

Developing country governments need support to digitise public records and to make their data publicly available. This requires technology infrastructure and training.

Developing country governments need support to digitise public records and to make their data publicly available; doing so requires technology infrastructure and training, as well as regulation of technology services and personnel.

39. http://www.rti-rating.org
40. http://www.rti-rating.org

Technology hubs are emerging in many developing countries, and these could be leveraged to provide innovative solutions for data generation and storage, as well as capacity building and training of citizens, civil society groups, and government representatives.

Donors can mainstream requirements for the use and public release of data into their programmes and projects, and they can further support civil society organisations and businesses in their efforts to create applications that enable the public to generate and use data. Important attention must be paid, beyond the first-order creation of open data sources, to the more complex issues of format, verification, and usability.

Finally, donors should encourage partnerships between experts on the legal and procedural context and experts in the computational and statistical sciences in order to best leverage the potential of data for accountability. This should include pressure to look beyond the creation of new data portals and the disclosure of new types of data to the details of accessibility and verification.

References

Arvate, Paulo, et al. (2013). Campaign donation and government contracts in Brazilian states. https://cmicro.fgv.br/sites/cmicro.fgv.br/files/arquivos/WP_7_2013.pdf. São Paulo School of Economics Center for Applied Microeconomics, Working Paper 7/2013.

Bagozzi, Benjamin E., et al. (2016). Predicting government (non)responsiveness to freedom of information requests with Supervised Latent Dirichlet Allocation. https://www.benjaminbagozzi.com/uploads/1/2/5/7/12579534/poltext2016-finalproceedings.16-21.pdf. Proceedings of the International Conference on the Advances in Computational Analysis of Political Text, Dubrovnik, Croatia, 14-16 July 2016.

Berliner, Dan. (2014). The political origins of transparency. https://www.journals.uchicago.edu/doi/abs/10.1017/s0022381613001412. The Journal of Politics, 76(2): 479-491.

Berliner, D. and Erlich, A. (2015). Competing for transparency: Political competition and institutional reform in Mexican states. https://www.cambridge.org/core/journals/american-political-science-review/article/competing-for-transparency-political-competition-and-institutional-reform-in-mexican-states/38ACAECD760A3FC00FD9E2753D5C56BC. American Political Science Review, 109(1): 110-128.

Bersch, K., et al. (2017). "State capacity, bureaucratic politicization, and corruption in the Brazilian state. https://onlinelibrary.wiley.com/doi/abs/10.1111/gove.12196". Governance, 30(1): 105-124.

Bhavnani, R. R. (2012). Using asset disclosures to study politicians' rents: An application to India. http://siteresources.worldbank.org/EXTABCDE/Resources/7455676-1315933592317/8143947-1335963402037/8622235-1336401580364/Session-4-Rikhil_Bhavnani.pdf.

Boas, T. C., et al. (2014). The spoils of victory: Campaign donations and government contracts in Brazil. https://www.journals.uchicago.edu/doi/abs/10.1017/s002238161300145x ". The Journal of Politics, 76(2): 415-429.

Brierly, S. (2018). Patronage, meritocracy and political party machines.. https://www.sarahbrierley.com/uploads/2/7/7/1/27711295/brierley_merit_machine_9jun18.pdf

Cantu, F. (2014). Identifying irregularities in Mexican local elections. https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12097. American Journal of Political Science, 58(4): 936-951.

Carton, S., et al. (2016). Identifying police officers at risk of adverse events. https://dl.acm.org/citation.cfm?id=2939698. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17 August 2016, San Francisco, CA.

Charron, N., et al. (2017). Careers, connections, and corruption risks: Investigating the impact of bureaucratic meritocracy on public procurement processes. https://www.journals.uchicago.edu/doi/abs/10.1086/687209. The Journal of Politics, 79(1): 89-104.

Colonnelli, E., et al. (2018). Patronage and selection in public sector organizations.. https://static1.squarespace.com/static/5596b3d9e4b0dd603214b252/t/5ba96bfd24a694c2396c4345/1537829888293/CPT_2018.pdf

Desouza, K.C., and Jacob, B. (2017). Big data in the public sector: Lessons for practitioners and scholars. http://journals.sagepub.com/doi/abs/10.1177/0095399714555751?journalCode=aasb. Administration & Society, 49(7): 1043-1064.

Fafchamps, M. and Labonne, J. (2017). Do politicians' relatives get better jobs? Evidence from municipal elections. https://academic.oup.com/jleo/article/33/2/268/2980942. The Journal of Law, Economics, and Organization, 33(2); 268-300.

Fazekas, M. and Kocsis, G. (2017). Uncovering high-level corruption: Cross-national objective corruption risk indicators using public procurement data. https://www.cambridge.org/core/journals/british-journal-of-political-science/article/uncovering-highlevel-corruption-crossnational-objective-corruption-risk-indicators-using-public-procurement-data/8A1742693965AA92BE4D2BA53EADFDF0. British Journal of Political Science, 1-10.

Fazekas, M., and Tóth, I. J. (2014). New ways to measure institutionalised grand corruption in public procuremen. https://www.u4.no/publications/new-ways-to-measure-institutionalised-grand-corruption-in-public-procurement?t. U4 Anti-Corruption Resource Centre, Issue 2014:9. Bergen: CMI.

Fisman, R., et al. (2013). The private returns to public office. https://www.journals.uchicago.edu/doi/abs/10.1086/676334. Journal of Political Economy, 122(4): 806-862.

Fox, J. and Haight, L. (2011). Mexico's transparency reforms: Theory and practice. http://pub-jschol-stg.escholarship.org/content/qt50q0m31z/qt50q0m31z.pdf?t=lmy8b8. Research in Social Problems and Public Policy, 19, 353-379.

Lewis-Faupel, S. et al. (2016). Can electronic procurement improve infrastructure provision? Evidence from public works in India and Indonesia. https://www.aeaweb.org/articles?id=10.1257/pol.20140258 . American Economic Journal: Economy Policy, 8(3): 258-283.

Olken, B. (2007). Monitoring corruption: Evidence from a field experiment in Indonesia. https://www.journals.uchicago.edu/doi/10.1086/517935. Journal of Political Economy, 115(2): 200-249.

Palguta, J. and Pertold, F. (2017). Manipulation of procurement contracts: Evidence form the introduction of discretionary thresholds. https://www.aeaweb.org/articles?id=10.1257/pol.20150511. American Economic Journal: Economic Policy, 9(2): 293-315.

Pierskalla, J. and Sacks, A. (2018). Teacher hiring in Indonesia.. http://pubdocs.worldbank.org/en/391721528721184345/personnel-politics-elections.pdf

Popa, M. (2018). What do good governments actually do? An analysis using European procurement data. https://www.cambridge.org/core/journals/european-political-science-review/article/what-do-good-governments-actually-do-an-analysis-using-european-procurement-data/3F6A04C8F80A05708C03636A1DFFCA37. European Political Science Review, 10(3): 369-391.

Ruiz, N. A. (2018). The power of money: The consequences of electing a donor funded politician. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3123592.