

High-Frequency Evidence on Corruption in 53 Countries: New Data from the MLP Project*

Diego Romero[‡], Jitender Swami[‡] and Erik Wibbels[‡]



August 8, 2023

Corruption is associated with everything from poor governance outcomes and low-quality public services to poor economic growth and political instability. Corruption can have negative effects on civic space as it reduces political participation and generalized trust. However, corruption scandals have also motivated powerful protest movements and civic action in countries as diverse as Brazil, Guatemala and Moldova in recent years.

Despite the centrality of corruption scandals to civic space dynamics in many countries, our capacity to understand when, why and where corruption elicits civic responses is sharply limited by constraints inherent in standard corruption data. Most such data, whether in the form of expert (e.g., Transparency International’s Corruption Perceptions Index) or citizen (e.g., Afrobarometer) surveys, provide annual snapshots that preclude answering key questions such as: When do corruption scandals evoke protests, civic activism, legal changes and/or the collapse of governments? And what are the implications of corruption for the evolution of civic space more generally?

In this report we introduce a new big data approach to measuring corruption that allows researchers and analysts to address these kinds of questions. Our corruption measure relies on the Machine Learning for Peace project’s infrastructure, which has collected and classified over 90 millions articles published by international and local newspapers every day from 2012 until last month for nearly 60 countries. By measuring the share of monthly news reporting on corruption, we provide data on its salience. This measure does a good job of identifying corruption scandals and provides a tool for monitoring corruption in near real-time.

To elucidate one potential use of the data, we analyze the relationship between corruption scandals and anti-corruption protests across 52 countries. We complement the cross-national evidence with case studies of Guatemala and Ghana, two cases with relatively high incidences of corruption, but where civic responses have varied a lot. Our key findings are:

- In many countries there is enormous month-to-month variation in the incidence of

*Report prepared for the INSPIRES Machine Learning for Peace project.

[‡] University of Pennsylvania.

reporting on corruption. This fact underscores a key advantage of our measure of corruption.

- There is also enormous variation in the average level of reporting on corruption across countries. It is unclear if this reflects differences in the actual practice of corruption or in the freedom of media to report on corruption across countries.
- Overall, there is a positive, but modest, relationship between corruption and protest movements. There is interesting cross-national variation in the extent to which civil society mobilizes protests in response to corruption.

We conclude with some reflections on the usefulness of our data for analysts of corruption and for policymakers intent on monitoring and responding to corruption in realtime.

Corruption and Civic Space

In addition to its negative impact on economic growth, administrative capabilities and the quality of public services (Mo 2001; Gründler and Potrafke 2019), corruption can also have significant effects on civic space. For instance, corruption is associated with lower levels of generalized trust (You et al. 2018) and undermines trust in – and the legitimacy of – the actors and institutions of government (Anderson and Tverdova 2003). Indeed, cross-country evidence suggests that increases in perceptions of corruption are associated with lower willingness to participate in anti-corruption activism (Peiffer and Alvarez 2016).

On the other hand, in countries as diverse as Guatemala (Flores 2019) and Nigeria (Smith 2014) corruption has been a central issue motivating civic mobilization and mass protests. Indeed, corruption scandals have brought down numerous governments in the past decade: in 2015, Moldovan Prime Minister Valeriu Strelet received a vote of no confidence after months of anti-corruption protests; in 2016, Brazil’s Dilma Rousseff was impeached after a corruption scandal; and South Africa’s Jacob Zuma resigned in 2018 amid anti-corruption protests. Thus, it seems that corruption can have divergent effects on civic space – it can either demobilize civic actors or it can mobilize them and contribute to big political changes.

One key obstacle to understanding when, where and why corruption has such varied effects on civic space is the lack of dynamic, high frequency corruption data. Previous research, as well as corruption monitoring efforts by watchdogs, typically rely on three broad types of data to track corruption: annual indices, which typically rely on surveys of citizens, firms and experts; audit data, which detail specific instances of corruption; or administrative data (e.g., procurement records) that can be used to infer specific instances of corruption or to explore patterns of malfeasance. These methods suffer from a number of limitations,¹ but most importantly, they do not produce data at high enough frequencies (e.g., surveys and/or audits tend to be once a year) to be able to analyze how civil society or the legal system responds to revelations of corruption. To the extent corruption scandals have been at the heart of civic space dynamics in many countries, that is a serious limitation.

¹For instance, survey-based measures may suffer from issues related to under-reporting due to social desirability bias or fear of retaliation, as well as complications related to reliance on peoples’ perceptions (e.g., Bohn 2012; Little and Meng 2023). Audit and procurement data require considerable public sector capacity, so data availability and quality vary a lot across countries.

Our Approach

We provide an original, high-frequency measure of corruption produced via the machine-coding of newspaper articles. The Machine Learning for Peace project (MLP) collects and classifies newspaper articles published by high-quality international sources and reputable local newspapers from nearly 60 countries. We rely on a fine-tuned large language model to classify each article into 20 different civic space events.² In the case of corruption, we apply a small keyword corpus to increase classification accuracy and allows the machine to correctly classify articles that bear on investigations, court cases or arrests that are related to corruption.³ The final corruption measure is the monthly share of all news articles on a country that are related to corruption. This ratio tells us how frequently corruption is mentioned relative to the total volume of news in a given month. While this method does not directly allow us to code *individual* corruption events, it does provide information on the *magnitude* of the corruption event(s) in a given month. Spikes in this measure are a particularly good proxy for corruption scandals.

The Corruption Data

Below we graph the monthly incidence of corruption for 52 countries from 2012-2023 (Figures 1 to 4). We organize countries by region and put them on a common scale to ease comparisons across countries. Interestingly, there is considerable cross-regional variation in the incidence of corruption in the media with relatively more corruption reported in African and Latin American countries than in European and Asian ones.

As a descriptive matter we want to emphasize two key findings. First, there is a great deal of cross-country variation in the incidence of corruption in the media. On average, countries with higher levels of reporting on corruption tend to also be those with higher levels of corruption perceptions as measure by standard sources like Transparency International. Looking within regions, the countries with particularly high rates of reporting include Kenya, South Africa and Nigeria in Africa, Guatemala in LAC, Ukraine in Europe, and Malaysia in Asia; and the corresponding low corruption cases are Benin (Africa), Jamaica (LAC), and Armenia (Europe). In Asia, the low average scores over time for Bangladesh and Cambodia reflect a limitation with our approach to measuring corruption, namely that it relies on press reporting. Both countries are known to have high levels of corruption, but they have very low levels of press freedom, which almost certainly reduces the capacity of journalists to investigate and report on corruption.

This aspect of our data limits its usefulness for *cross-country* comparisons, but it points to its major strength and our second key finding, i.e. there is a great deal of variation in corruption reporting across time within countries. This is true even in repressed media environments, because even modest amounts of increased reporting on corruption are reflected in the time-series. Take, for instance, Nicaragua (in Figure 4) which has amongst the most repressive media environments in the world. Even there, where corruption stories typi-

²For a full list of event types and more information about the MLP pipeline, visit our website.

³The keywords we employ are: assets, bribe, budget, corrupt, embezzlement, enrichment, fraud, irregularities, laundering, procurement and their derivations.

cally represent less than 1% of reporting, significant corruption events are evident in several months in 2021 and 2022. Indeed, significant spikes in corruption-related reporting across months within countries are associated with corruption scandals irrespective of whether baseline corruption reporting is high or low. Thus, the data lends itself to answer key questions bearing on dynamics within countries. For instance: When and why do corruption scandals elicit protests sometimes but not others?

Figure 1: Africa: Reporting on Corruption



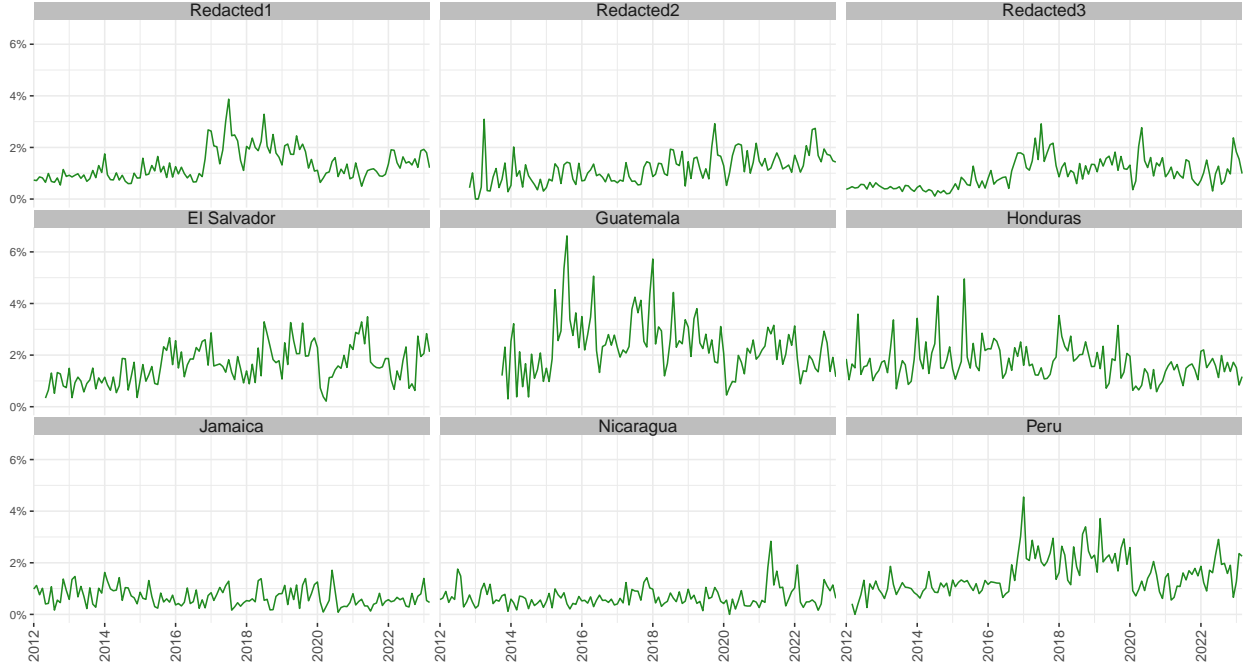
Figure 2: Asia: Reporting on Corruption



Figure 3: Europe and Eurasia: Reporting on Corruption



Figure 4: Latin America and The Caribbean: Reporting on Corruption



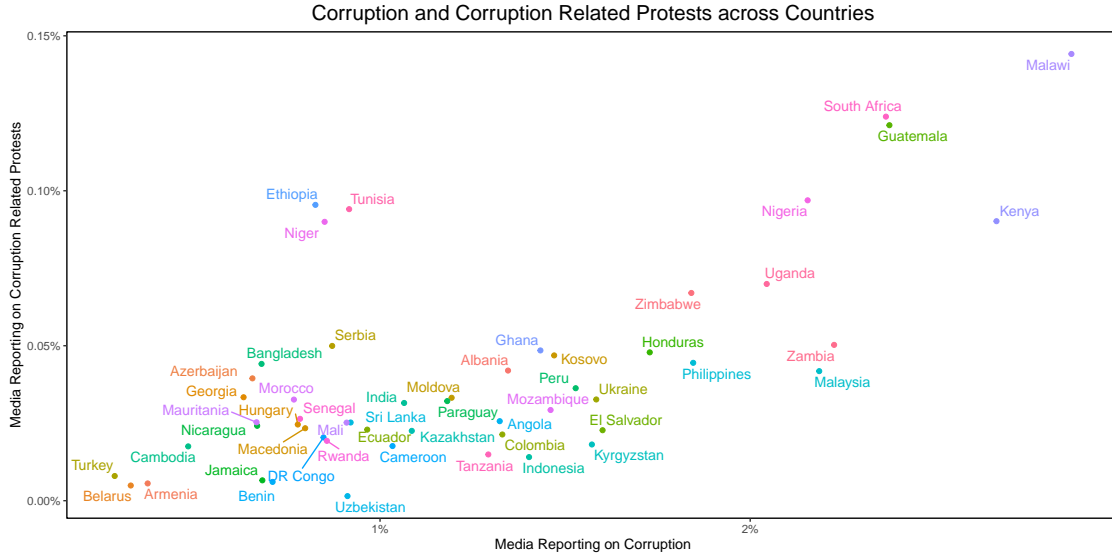
The Relationship between Corruption and Protests

Here we provide an example of one important use of the data, i.e. for learning about the relationship between corruption and civic mobilization. As a preliminary step in that direction, Figure 5 plots the average level of reporting on corruption against reporting on corruption-related protests.⁴ If there is a high correlation between corruption reporting and corruption-related protest reporting, this would provide suggestive evidence that they are closely related.

The figure shows that there is, indeed, a positive relationship between the two. Countries like Malawi, Guatemala and South Africa have both lots of corruption and lots of corruption-related protests. Nevertheless, the relationship is far from linear. There are many countries toward the bottom middle of the graph (Tanzania, Indonesia, Kyrgyzstan, etc.) with considerable reporting on corruption but quite little protest activity in response. There is also a cluster of countries – Ethiopia, Tunisia and Niger – where corruption-related protests have been quite high despite relatively modest reporting on corruption itself. All told, the figure presents a series of puzzles for future research.

⁴We distinguish corruption-related protests from other protests by applying a small set of keywords to the MLP protest data.

Figure 5: The Relationship Between Corruption and Corruption-Related Protests



The summary figure above necessarily smooths over a lot of within-country variation in reporting on corruption. To provide some insight into how our MLP corruption data can be useful for within country analysis, we briefly discuss two case studies – Guatemala and Ghana. Both countries have had regular corruption scandals, but while those have elicited regular rounds of protests in Guatemala, they have mostly not in Ghana.

Guatemala. Corruption has been one of the most salient topics in Guatemalan politics over the past two decades. In the 2007-2019 period, the fight against corruption was spearheaded by the *International Commission Against Impunity in Guatemala* (CICIG, in Spanish) – a special commission backed by the United Nations– and the General Attorney’s Office (Ministerio Público, MP) it helped reform. Over the span of 12 years, CICIG helped in the filing of 120 cases, including against two ex-presidents and a sitting president and vice-president.⁵ When then President Morales suspended CICIG in 2019, most anti-corruption efforts were abandoned.

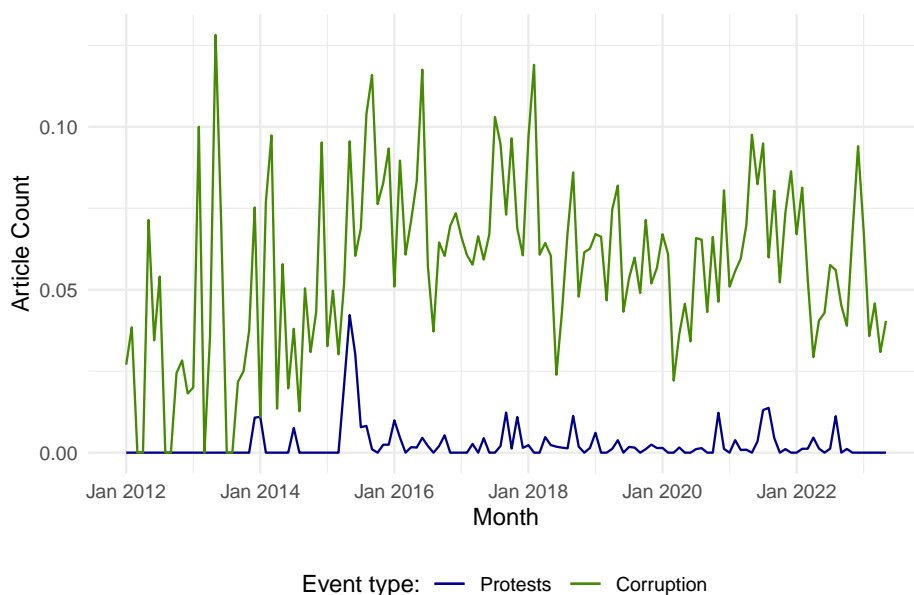
Figure 6 shows how the salience of corruption and corruption-related protests changes through time. The figure shows two important trends. First, reporting on corruption climbed after the corruption cases of early 2015 and the large protests that ensued. Second, when corruption reporting peaks it often coincides with anti-corruption protests. Indeed, almost all major corruption scandals of the period under study (2012-2023) were followed by protests. The largest of these scandals was the “La Linea” case, which implicated the sitting president and vice-president in a vast network of customs fraud. The case was made public in phases starting in April 2015, along with a few other corruption scandals, including one involving the vice-president and the Social Security Institute. The protests that ensued were the largest in recent history and contributed to the resignation of both the president and vice-president. Likewise, the attempt by Congress to reduce sentences for corruption in the fall of 2017 (known since then as the “pacto de corruptos”) and the attempts to expel CICIG in 2018 were both met by protests. In November 2021, people went to the streets of Guatemala City

⁵<https://www.wola.org/analysis/cicigs-legacy-fighting-corruption-guatemala/>

to protest a budget plan which many saw as a vehicle to increase corruption, and in 2022 a politically-motivated investigation against the last remaining anti-corruption prosecutor resulted, once again, in mass protests.

Before 2015, corruption scandals did not lead to protests with the same frequency. For instance, the capture of Carlos David De León Argueta, former general attorney and one of the first high-profile arrests made by anti-corruption prosecutors did not lead to anti-corruption protests. Civil society saw this arrest as evidence that the fight against corruption was beginning to bear fruit. In contrast, the magnitude and scope of the scandals unveiled from 2015 onward gave citizens a glimpse of the extent and enormous costs of corruption.

Figure 6: Guatemala: Reporting on Corruption and Corruption-Related Protests



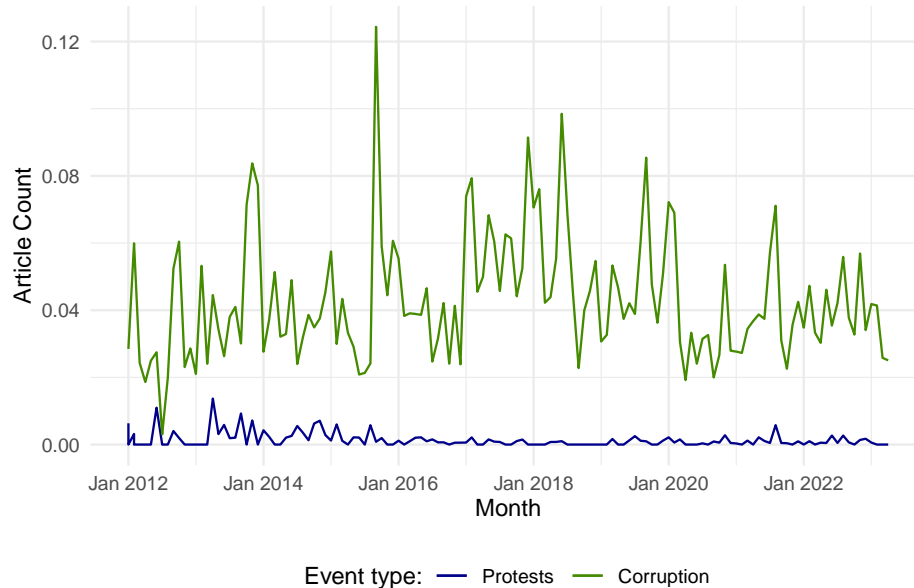
Ghana. Over the past two decades, Ghana has shown lower levels of corruption compared to other African countries. Indeed, it has seen a recent decline in corruption, including a ten-point improvement in Transparency International’s corruption perception Index in 2022.⁶ Nevertheless, there have been regular corruption scandals involving everything from the misuse of welfare funds to suspicious contracting around COVID-19 vaccines. Despite those scandals, public outrage and civic protests have been relatively muted.

As shown in Figure 7, there does not seem to be a relationship between the salience of corruption in the media and mass protests, particularly since 2015. Interestingly, Ghana adopted a significant legal change – the National Anti-Corruption Action Plan (NACAP) – in 2014. The plan gave civil society organisations and the media a significant role in the implementation of the NACAP.⁷

⁶<https://www.transparency.org/en/countries/ghana>

⁷See (<https://chraj.gov.gh/nacap/>)

Figure 7: Ghana: Corruption and Corruption-Related Protests



One plausible interpretation is that the post-2015 reduction in corruption-related protests is attributable to the active role played by the media and CSOs in formal anti-corruption institutions, such as the Commission for Human Rights and Administrative Justice and the Ghana Anti-corruption Coalition (Rahman 2018). Indeed, it is telling that one of the few scandals that did trigger public demonstrations was pre-NACAP, i.e. the 2013 GYEEDA Scandal. The outcry over the misappropriation of funds intended for youth welfare was enough to spark significant protests. On the other hand, the strength of institutionalized civic responses to corruption revelations are evident across the post-NACAP years, including during the 2015 Judicial Bribery Scandal, the 2015 Bus Branding Scandal, and the 2020-21 COVID Funds Mismanagement scandal. Civil society actors and the media played a prominent role across each of these scandals,⁸. It could be that the efficacy of these organized civic efforts, coordinated via the NACAP, obviate the incentive and need for public protests to promote government responses to corruption scandals.

Conclusion

A key finding in our exploration of the relationship between corruption and mass mobilization has to do with the vital role of a vibrant and well-organized civil society in the fight to control corruption. Civil society can bring about change either through institutionalized channels, as in the case of Ghana, or through the organization of mass protests, as in the case of Guatemala. The success of one strategy versus the other crucially depends on the responsiveness of elected officials.

⁸See, for instance, <https://www.refworld.org/docid/574c49079.html>, <https://citifmonline.com/2016/02/occupy-ghana-goes-to-court-over-bus-branding-scandal/>,

<https://www.bbc.com/news/world-africa-34452768>,

This report showcases one of the many applications of the novel measure of corruption developed by the MLP research team. That measure provides analysts with a near realtime tool for monitoring corruption, and it provides policymakers with a tool for responding to corruption scandals in a timely way. Indeed, our data offers the potential to answer several crucial policy questions: How do repressive media laws impact reporting of malfeasance? What the electoral cost of corruption scandals? And what is relationship between corruption and democratic backsliding more broadly?

References

- Anderson, Christopher J, and Yuliya V Tverdova. 2003. "Corruption, political allegiances, and attitudes toward government in contemporary democracies." *American journal of political science* 47 (1): 91–109.
- Bohn, Simone R. 2012. "Corruption in Latin America: understanding the perception-exposure gap." *Journal of Politics in Latin America* 4 (3): 67–95.
- Flores, Walter. 2019. "Youth-Led Anti-Corruption Movement in Post-Conflict Guatemala: 'Weaving the Future'?"
- Gründler, Klaus, and Niklas Potrafke. 2019. "Corruption and economic growth: New empirical evidence." *European Journal of Political Economy* 60:101810.
- Little, Andrew, and Anne Meng. 2023. "Subjective and Objective Measurement of Democratic Backsliding." *Available at SSRN 4327307*.
- Mo, Pak Hung. 2001. "Corruption and economic growth." *Journal of comparative economics* 29 (1): 66–79.
- Peiffer, Caryn, and Linda Alvarez. 2016. "Who will be the "principled-principals"? Perceptions of corruption and willingness to engage in anticorruption activism." *Governance* 29 (3): 351–369.
- Rahman, K. 2018. "Overview of corruption and anti-corruption in Ghana: U4 Anti-Corruption Helpdesk, CHR." *Michelsen Institute*.
- Smith, Daniel Jordan. 2014. "Corruption complaints, inequality and ethnic grievances in post-Biafra Nigeria." *Third World Quarterly* 35 (5): 787–802.
- You, Jong-sung, et al. 2018. "Trust and corruption." *The Oxford handbook of social and political trust*, 473–496.