

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

Diego Romero, Jitender Swami and Erik Wibbels

Utah State University and University of Pennsylvania

Objective

Corruption is associated with everything from poor governance outcomes and low-quality public services to poor economic growth. It can also have negative effects on civic space by reducing political participation and general trust. However, corruption scandals have also motivated powerful protest movements and civic action in countries as diverse as South Africa, Guatemala and Moldova in recent years. We provide new high-frequency data for 53 countries that allows for a detailed understanding of when, where and how corruption scandals impact civic space.

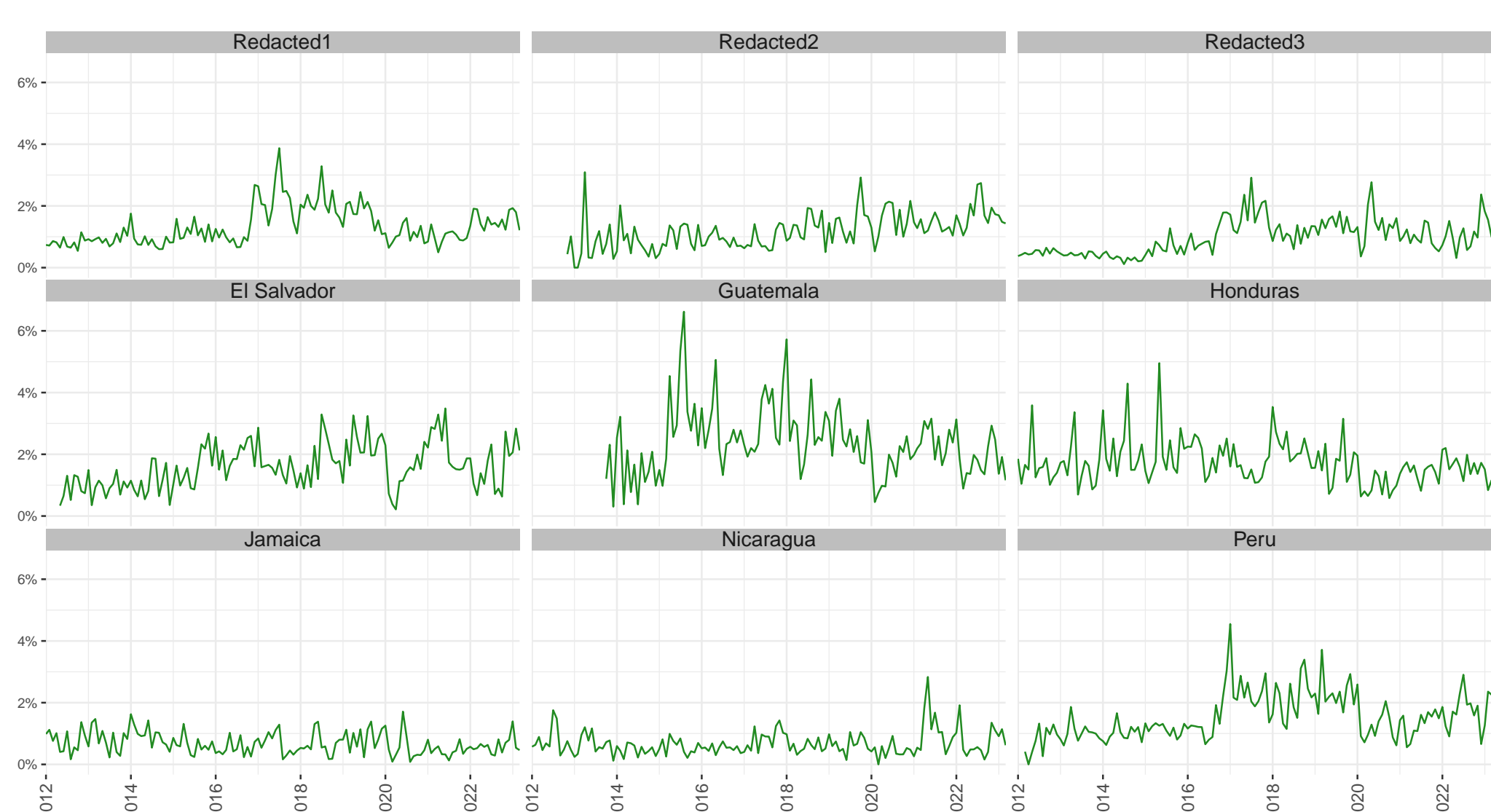
The Challenge

Despite the centrality of corruption scandals to civic space dynamics in many countries, our capacity to understand its impact 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?

Our Approach

We introduce a new big data approach to measuring corruption that allows researchers and analysts to address these kinds of questions. The Machine Learning for Peace project (MLP) collects and classifies newspaper articles from high-quality international and local sources across over 53 countries. A fine-tuned large language model classifies more than 90 million articles into civic space qualities, including a label for articles related to corruption. By measuring the share of monthly news reporting on corruption in each country, 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. Figure 1 graphs the monthly incidence of corruption for several countries in Latin America as an example (“redacted” refers to countries with which the USG has sensitive relations). To elucidate one potential use of the data, we analyze the relationship between corruption scandals and anti-corruption protests across countries, including case studies of Guatemala and Ghana, two countries where civil society has responded to corruption very differently in recent years.

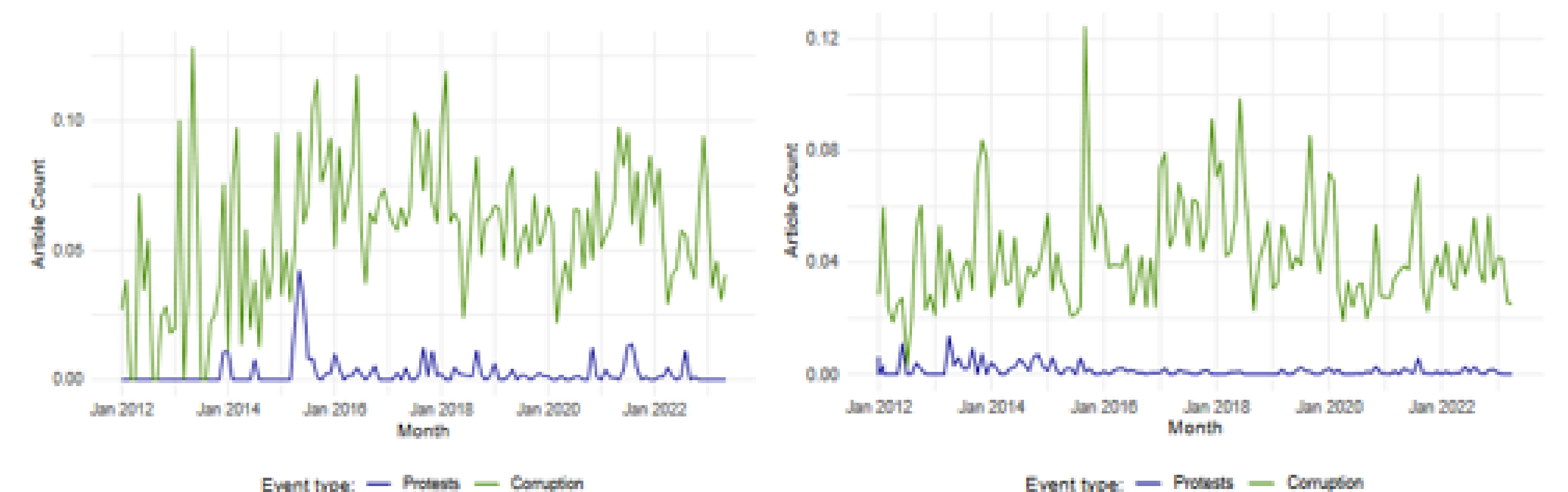
Figure 1: The Incidence of Corruption Across LAC Cases



Findings

We have three key findings. First, as Figure 1 shows there is a great deal of cross-country variation in the incidence of corruption as reported in the media. On average, countries with higher levels of reporting on corruption (like Guatemala) also tend to be those with higher levels of corruption perceptions as measure by standard sources like Transparency International. The low level of reporting on a country like Nicaragua points to a limitation of our data, i.e. that country’s tightly controlled press 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 over 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. Even in Nicaragua, where corruption stories typically represent less than 1% of reporting, significant corruption events are evident in several months in 2021-22. 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. Third and as a result, the data lends itself to answer key questions bearing on dynamics within countries. As an example, Figure 2 shows the relationship between corruption and corruption-related protests in Guatemala and Ghana. While corruption scandals are often associated with protests in Guatemala, they are not in Ghana. Building on those our cases, in our report we discuss how 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. This might help explain why there is a positive, but weak, relationship between corruption and protest movements across our 53 countries.

Figure 2: Corruption and Protests in Guatemala (left) and Ghana (right)



Implications

Our full report showcases one of the many applications of the novel measure of corruption developed by the MLP research team. That measure overcomes several limitations of standard corruption data and, therefore, 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?

Additional Information

- Email: mlpeace-devlab@sas.upenn.edu
- [Click here to read the full report](#)

Acknowledgements

This study is part of the (INSPIRES) project funded by the United States Agency for International Development (USAID) Center for Democracy, Human Rights, and Governance.