# Measuring Changes in Corruption over Time*†

**Preliminary version**: This paper is under active development. Conclusions, arguments, and evidence may change as research progresses.

Maya Dalton[1] and Justin Esarey[2]

[1]Pennsylvania State University, Department of Political Science
[2]Wake Forest University, Department of Politics and International Affairs

December 1, 2023

**Abstract**

In reaction to prior criticism, country-level corruption indicators have adopted measurement strategies designed to accurately track changes in corruption within a country over time. We present evidence that these indicators still struggle to validly measure change in corruption. Many causal inference research designs, such as difference-in-difference and dynamic panel instrumental variable models, rely on such within-country changes to identify causal relationships. As a result, we argue that empirical findings about corruption based on these designs should be interpreted with caution. We present a new synthetic measure of within-country change in corruption, constructed using a factor score extracted from panel-adjusted variants of extant measures, that is more robust to bias and correlated measurement error in its components.

*We gratefully acknowledge the financial support for this project provided by the Wake Forest University Undergraduate Research and Creative Activities Center (URECA) and the WFU Department of Politics and International Affairs.

†Competing interests: The authors declare none.

*Tempora mutantur, nos et mutamur in illis.*

Johannes Nas

# Introduction

Every measure of corruption is constructed differently, but all face the fundamental challenge that corruption is a hidden behavior; its perpetrators do not want their largely illegal dealings recorded in a data set (Brooks et al., 2013, p. 27). Consequently, the "abuse of public power for private gain" (Hawken and Munck, 2008, pp. 74-75) is difficult to directly observe.[1] The unobservability of corruption creates a large number of persistent and difficult measurement problems, problems that have been a core concern of scholars and policymakers for nearly as long as measures of corruption have existed (e.g., Sampford et al., 2006; Brooks et al., 2013; Heywood and Rose, 2014). A recurring theme in this research is that influential country-level measures of corruption are not comparable over time. Transparency International's Corruption Perception Index (TI CPI, published since 1995) is a frequent target of this criticism (Andersson and Heywood, 2009; Brooks et al., 2013, p. 37; June et al., 2015, pp. 26-27), but similar issues have been raised for the International Country Risk Guide (ICRG) measure of the political risk of corruption as well as the World Bank Governance Indicators Control of Corruption estimate (WBGI CCE) (Treisman, 2007, p. 220; Knack, 2007; Hawken and Munck, 2008, p. 85; Standaert, 2015). Criticism has not stopped these measures from being widely used; the TI CPI alone has been cited about 8,240 times according to Google Scholar.[2]

Ensuring that cross-national measures of perceived corruption are comparable over time

[1] While there are alternative conceptualizations of corruption (e.g. Warren, 2004), its definition as "abuse of public office for private gain" has become relatively standard across the literature since being adopted by several prominent non-governmental organizations (June et al., 2015, p. 12).

[2] We produced this figure by searching with the prompt *"Corruption Perceptions Index" and intitle:corruption.* Google Scholar citation counts are estimates and therefore inexact.

should be a priority of scholarship in the field. Very basic descriptive inferences require such measures; without them, we cannot even answer simple questions such as whether democracies have gotten more corrupt (on average) over the last twenty years. Moreover, many causal inference techniques for observational data (such as difference-in-difference designs or dynamic panel data instrumental variable models) depend crucially on the accurate measurement of these changes. This concern is already recognized in published work; for example, Gründler and Potrafke's (2019, p. 2-3) study of the effect of corruption on economic growth begins:

> Studies using the CPI in panel data models ignored that the CPI was not comparable across countries and over time before 2012. In particular, including fixed period effects in panel data models does not solve the incomparability problem because the CPI in individual years before the year 2012 included data for different components and time periods to measure perceived corruption across continents. We believe that measuring corruption in the public sector by the CPI is suitable. However, one cannot conclude from previous studies that corruption decreases growth, because the earlier version of the CPI is not comparable across time.

They go on to use two-stage least squares (2SLS) analysis with lagged corruption scores as instrumental variables in a dynamic panel data model (Arellano and Bond, 1991; Blundell and Bond, 1998) to estimate the effect of corruption on growth. Changes in TI CPI score need to be valid measures of change in perceived corruption in order for their model to generate valid causal inferences.

As referenced in the above excerpt from Gründler and Potrafke (2019), some measures that were previously criticized have been changed to enable valid comparison of scores over time. For example, Transparency International adjusted its methodology in 2012 to make the TI CPI ratings comparable across different years (Transparency International, 2012). Some new measures were created in reaction to these past criticisms; among these new

measures is the Bayesian Corruption Index (BCI), an extension of the WBGI CCE with a revised methodology explicitly designed to capture trends over time (Standaert, 2015). Finally, in some cases, scholarship has responded to these critiques by arguing that they are wrong. As an example, Kaufmann, Kraay and Mastruzzi (2007) respond at length to eleven distinct criticisms of the WBGI, including four different reasons why WBGI CCE scores are alleged to be incomparable over time, and conclude that these critiques are "based on a misunderstanding of our aggregate indicators" and/or "entirely lacking in empirical support" (p. 30).

In this paper, we ask: are new and updated cross-national measures of corruption perception valid for comparisons over time? Can we use these measures to accurately study corruption's causes and effects using cross-national panel data? We argue that if within-country changes in the most commonly used measures of corruption perception are closely correlated with one another—that is if they perform well on an assessment of their construct validity[3]—then all capture the same concept. On the other hand, if there is little correspondence between different measures of corruption, then at most one of them—and perhaps none—is valid. In that case, we would conclude that using existing measures to study corruption would be challenging at best. Scholars of corruption already explain and defend the measures they use in published work using their construct validity (for examples, see Ko and Samajdar, 2010; Thomas, 2010; Langbein and Knack, 2010; Dincer and Gunalp, 2012; Tabish and Jha, 2012; Mondo, 2016; McMann et al., 2016; Gründler and Potrafke, 2019), and so our approach is a good match to the one already informally taken in the literature.

To study the construct validity of within-country trends in corruption, we use a data set containing information about 199 countries from 1980-2020. The key variables in our analysis are five influential measures of corruption: the TI CPI, WBGI CCE, BCI, ICRG,

[3]By construct validity, we mean "the degree to which [a] measure is related to other measures that theory requires them to be related to" (Kellstedt and Whitten, 2018, p. 123).

and the Varieties of Democracy (V-Dem) political corruption index (Coppedge et al., 2021). Our analysis shows that the five measures of corruption we study do not recover similar within-country changes in corruption over time. We first explain the methodology behind these measures, as each takes a different approach. Then, we describe a mathematical decomposition of time-series cross-sectional (TSCS) measures that identify change over time within units. We use that decomposition to construct *panel-adjusted* versions of each of the five measures of corruption we study; our panel-adjusted measures are designed to accurately track changes in perceived corruption within countries over time while ignoring systematic between-country differences or worldwide annual shocks. We find that panel-adjusted versions of the CPI, WBGI, BCI, V-Dem, and ICRG corruption measures are only weakly correlated and give very different answers to simple and important questions about the trajectory of corruption within countries.

Finally, we present an alternative measurement model of corruption. Our model is similar to the unobserved components model behind the WBGI and BCI, but directly maps onto principal component analysis and thereby allows us to extract a common component of all five corruption perception variables corresponding to a latent concept that they all share. Most importantly, our approach imposes fewer assumptions on the structure of measurement error; it is robust to the presence of systematic bias among the measures and error correlation across them. While all five measures load onto a common factor in the expected way—that is, while all five measures seem to contain a signal corresponding to the perception of aggregate corruption—this factor explains comparatively little variance in the component scores. For all these reasons, we conclude that existing corruption perception measures lack construct validity for measuring changes in corruption within a country over time. There are substantively meaningful differences between these measures that are not ascribable to transient errors and that will affect statistical inference.

Although the field's interest in sub-national and sector-specific measures of corruption

is increasing (Heywood, 2014, p. 148), we agree with Mungiu-Pippidi and Fazekas (2020) that valid and reliable country-level measures remain vitally important for the study of corruption. We find their argument persuasive when they say (p. 24) that "the national context—the level where the rules of the game are set—is crucial." Exploiting sub-national variation in corruption is advantageous in many ways, but a great deal of institutional variation is between countries and over long time periods. If we think institutions affect corruption, time-series cross-sectional data is a particularly appropriate place to study those relationships. Furthermore, findings derived from within-country data have limited external validity until demonstrated in a cross-national context. The factor score corruption measure we create using principal component analysis, which extracts the common signal from all five panel-adjusted measures of perceived corruption, allows for a more valid assessment of within-country change in corruption compared to any of the individual component measures. Our measure does not answer every critique in the literature; for example, we do not address the claim that perception-based measures do not predict less subjective proxies for corruption (Donchev and Ujhelyi, 2014), a claim already disputed elsewhere (Charron, 2016). However, we do believe that our measure provides a strong response to one very important criticism of cross-national corruption measures, and we provide our factor score measurements as part of the replication data so that researchers can use and improve upon them in future work.

## Concept and Construct Validity for Measures of Corruption

The five quantitative, country-level TSCS measures of corruption that we study are constructed using different methodologies and data sources. This fact has served as a basis for criticism of their concept validity, particularly for measures that aggregate many disparate assessments into a single index like the WBGI CCE (Thomas, 2010; Oman and Arndt, 2006, p. 72). But it is well known that all these measures are highly correlated with one another, with pairwise correlation coefficients over $\rho = 0.9$ (Table 3 in Standaert, 2015, p. 789; see

also Ko and Samajdar, 2010 pp. 520-521). This is hopeful news for their construct validity and gives us reason to believe that, while different in some details, each measure captures the overall pervasiveness of corruption in a country.

Unfortunately, correlation among these corruption measures "is almost completely driven by their between-correlations (the correlation between the mean values for each country)" (Standaert, 2015, p. 788) and *not* changes in corruption within countries over time. Criticisms of the validity of within-country changes in corruption over time have focused on the idea that "variations in reported levels of corruption are as likely to be a product of [...] the methods that are used to create these measures, as they are to reflect actual levels of corruption" (Heywood and Rose, 2014, p. 508). For this reason, even if these measures are valid for cross-sectional studies, they may not be valid for changes in corruption within a country over time.

Our source for the five corruption measures we study are the WBGI dataset, the BCI dataset (via the Quality of Government data set or QOG from Teorell et al., 2019), the V-Dem data set, and the ICRG data set. Appendix A shows a full list of countries available in our data set. Some countries do not have all corruption scores available for certain years and some indicators are only available for segments of the time period; see Figure 6 in Appendix B for the availability of indicators over time.

The five corruption measures we study are indeed created using very different procedures. To highlight these differences, we describe each measure in depth in this section alongside some of the criticisms that have been leveled at them. We do not comprehensively cover all criticisms that have been made of perception-based, country-year corruption measures in our review; instead, we focus on criticisms that specifically pertain to the validity of these measures to track changes in corruption over time.[4] These descriptions list the original

[4]The descriptions in this section are similar to descriptions in the online appendix of Esarey and Dalton (2023); these papers were written at the same time using (some of) the same variables.

scales of the variables, but we have rescaled them to range from 0-100 (with larger numbers indicating more corruption) for all our analyses.

## Transparency International's Corruption Perceptions Index (TI CPI)[5]

The CPI is an extremely influential indicator of corruption widely used by scholars and policymakers. According to Galtung (2006, p. 106), "The impact of the CPI has been considerable. It has been credited as a factor that gave the issue of corruption 'greater international prominence' (Florini, 1998).... The CPI has facilitated a qualitative shift in the journalistic writing and public discourse on corruption.... This interest and awareness of the CPI extends well beyond the business and financial press." However, the use of the CPI has become more controversial over time (Bello y Villarino, 2021; Heywood and Rose, 2014). As its name explicitly says, the CPI measures *perception* of corruption and not personal experience with corruption or convictions for corruption-related crimes. The CPI is constructed by averaging at least three (but as many as thirteen) different corruption scores taken from perception-based surveys and expert assessments of corruption in a given country. The CPI targets perceived corruption in the public sector within a country and compiles relevant data from multiple, independent sources.

As noted in our introduction, the CPI has been a frequent target of criticism in part because its methodology did not always enable accurate tracking of corruption trends within countries over time. In response to that criticism, Transparency International altered its methodology in 2012 to ensure that consistent and contemporaneous sources were used year after year (Transparency International, 2012); for scores prior to 2012, the underlying sources of data changed from year to year and information from several prior years was used to create the target year's score. Like all perception-based measures, TI CPI relies on judgments made

[5]Information about the CPI has been paraphrased from Transparency International (2016) and Transparency International (2020)

by experts and survey respondents; these sources may have conflicting understandings among themselves about what corruption means or may have conceptualizations of what corruption means that do not match the definition prevalent in the country itself. The CPI's native scale ranges from 0 (most corrupt) to 100 (least corrupt) and is available from 1995 onward.

## World Bank Group's Worldwide Governance Indicators (WBGI)[6]

The WBGI consists of six aggregated indicators of governance—control of corruption, voice and accountability, rule of law, government effectiveness, political stability, and regulatory quality—compiled from 30 data sources. The WBGI's sources include many different surveys (of the general population and of selected sub-groups) and expert assessments, including some of the other measures of corruption we study here (such as the ICRG political risk of corruption measure). It utilizes an Unobserved Components Model (UCM) to construct six aggregated indicators of governance and estimate margins of error for each indicator. Of the six indicators, our interest is in their measure of *control of corruption*, defined by Kaufmann, Kraay and Mastruzzi (2010, p.4) as "the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as 'capture' of the state by elites and private interests."

Some have argued that the six WBGI indicators are not empirically or conceptually distinct from each other (Langbein and Knack, 2010). In addition, the corruption measure combines many disparate concepts of corruption into an amorphous and unidentifiable aggregate (Apaza, 2009, p. 141), a criticism that may be applicable to many country-level omnibus measures (such as the TI CPI). The WBGI's creators reject these claims (Kaufmann, Kraay and Mastruzzi, 2007) in part based on the design of their UCM. The WBGI has also been criticized because of the large standard errors associated with its governance estimates; the uncertainty of these estimates makes it difficult to find statistically significant

[6]Information about the WBGI has been paraphrased from Kaufmann, Kraay and Mastruzzi (2010).

differences in corruption scores within countries (Bello y Villarino, 2021). The WBGI's original scale ranges from -3 (least control over corruption, or highly corrupt) to 3 (most control over corruption, or least corrupt), and is available for 1996, 1998, 2000, and 2002 onward.

## Bayesian Corruption Index (BCI)[7]

The BCI is another index of the perception of overall corruption (abuse of public power for private gain) within a country. It is built using an extension of the WBGI's Unobserved Components Model and based on 17 different surveys of countries' inhabitants, business executives, and government officials. Unlike the WBGI, the BCI implements a state-space model that allows for both the overall level of corruption worldwide and any individual country's level of corruption to change over time; by contrast, the WBGI assumes that global corruption levels remain constant over time. The underlying source data of the BCI are entered without any ex-ante imputations, averaging, or other manipulations, thereby avoiding selection biases introduced through the modeling choices of the index's creator. The BCI ranges between 0 (least corrupt) to 100 (most corrupt) in countries and is available from 1984 onward (a larger time span than the CPI and WBGI).

## Political Risk Service's International Country Risk Guide (ICRG)[8]

The International Country Risk Guide provides political, economic, and financial risk ratings to inform businesses about potential risks to their firms when operating within certain countries. The corruption measure is a panel of experts' assessment of the *risk to businesses and foreign investors* that corruption presents in a given country; it may therefore be biased in favor of the attitudes, priorities, and viewpoints of businesses rather than individual citizens or other organizations. Furthermore, the ICRG's concept of corruption includes

[7]Information about the BCI has been paraphrased from Standaert (2015).
[8]Information about the ICRG has been paraphrased from The PRS Group (2020)

concepts such as the prevalence of bribery, patronage, nepotism, extortion, and suspicious ties between business and politics, some of which are excluded from the other measures; the V-Dem measure of political corruption, for example, excludes patronage and clientelism (Lindberg, Lo Bue and Sen, 2022). Because the ICRG may give different weights to the various types of corruption going on in a state between years, it may be difficult to compare the measure year-over-year even though it is designed for this purpose (Knack, 2007, p. 261). As produced by the ICRG, the measure ranges between 0 (low political risk from corruption) and 6 (high political risk from corruption) and is available for 1994 onward.

## Varieties of Democracies (V-Dem) Political Corruption Index[9]

The V-Dem project collects 470 measures related to democratic governance, each of which is built using structured subjective assessments by country experts. Their index of political corruption is based on averaging information from four subsidiary measures: (i) the public sector corruption index, (ii) the executive corruption index, (iii) a measure of legislative corruption, and (iv) a measure of judicial corruption. Each of these four sub-indexes is in turn created from the output of an item response theory (IRT) model combining many experts' assessments about different aspects of corruption in the targeted sector. This IRT model is designed to allow meaningful comparisons of country-level corruption (or any of the four subsidiary measures) over time. The resulting composite measure of political corruption ranges from 0 to 1, with 0 indicating low corruption, and is available in our data between 1980 and the present; the V-Dem data set goes back considerably further, with historical assessments (made by scholars in the present day) going all the way back to 1789.

[9]Information about the V-Dem has been paraphrased from Coppedge et al. (2021)

## Construct Validity Check: Corruption in China and the United States

The inconsistency of these five measures of corruption can be illustrated with a simple descriptive example. Consider Figure 1, which shows corruption measurements over time for China (panel 1a) and the United States (panel 1b). For this graph, we subtracted the value for each country-year from the country's overall mean on that measure to emphasize differences in trajectory of within-country change over this time period. While the measures clearly share commonalities, from year to year they often disagree on the direction of change. For example, in China, the ICRG indicates a relatively stable corruption level between 2005 and 2018. But the other measures indicate a sharp decline in corruption. For the United States, the V-Dem and BCI measures indicate stability in corruption levels over the last forty years. By contrast, the WBGI shows dramatic growth in corruption between the years 2000 and 2010.

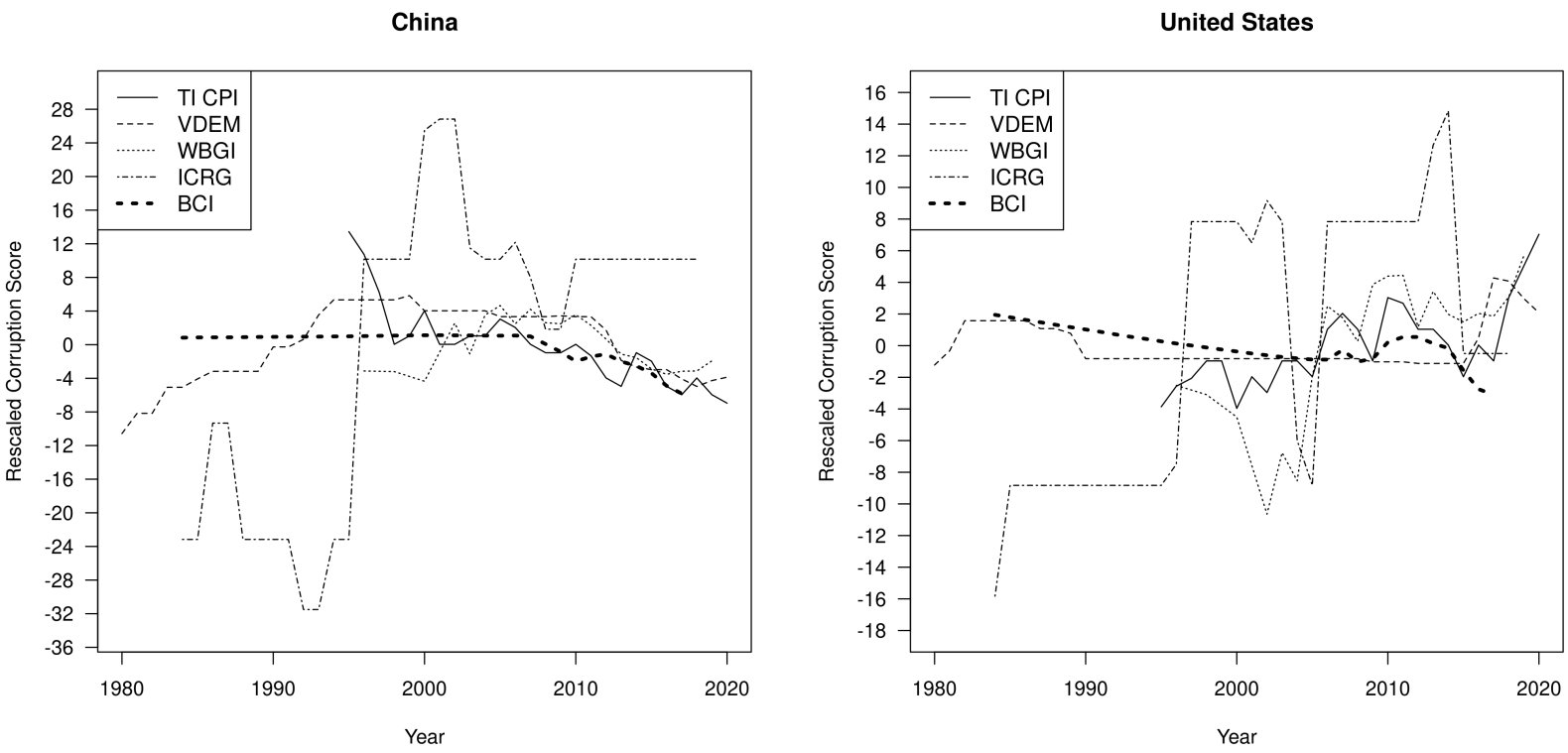
The substantive upshot of this divergence is that it is difficult to answer even relatively basic descriptive questions about influences on corruption in a country because different measures do not agree. Consider one such question: was the Obama administration more or less corrupt than other US presidential administrations between 1980 and 2020? To answer that question, we can estimate a relatively simple time series model on corruption data from the United States:

$$\text{US corruption}_t = \beta_0 + \beta_1 \times \text{year}_t + \beta_2 \times \text{Obama}_t + \varepsilon_t \tag{1}$$

The variable *Obama* is a dummy for whether the observation $t$ occurs between the years of 2009 and 2016, inclusive.[10] We need to consider the potential non-stationarity of this time series to avoid the potential for spurious correlation; although corruption measures are

[10] As this is a single time series and we are estimating a trend term, there is no need to employ panel-adjusted dependent variables for this model (described in the next section) as the results would not be different.

Figure 1: **Two examples of divergent corruption trajectories**



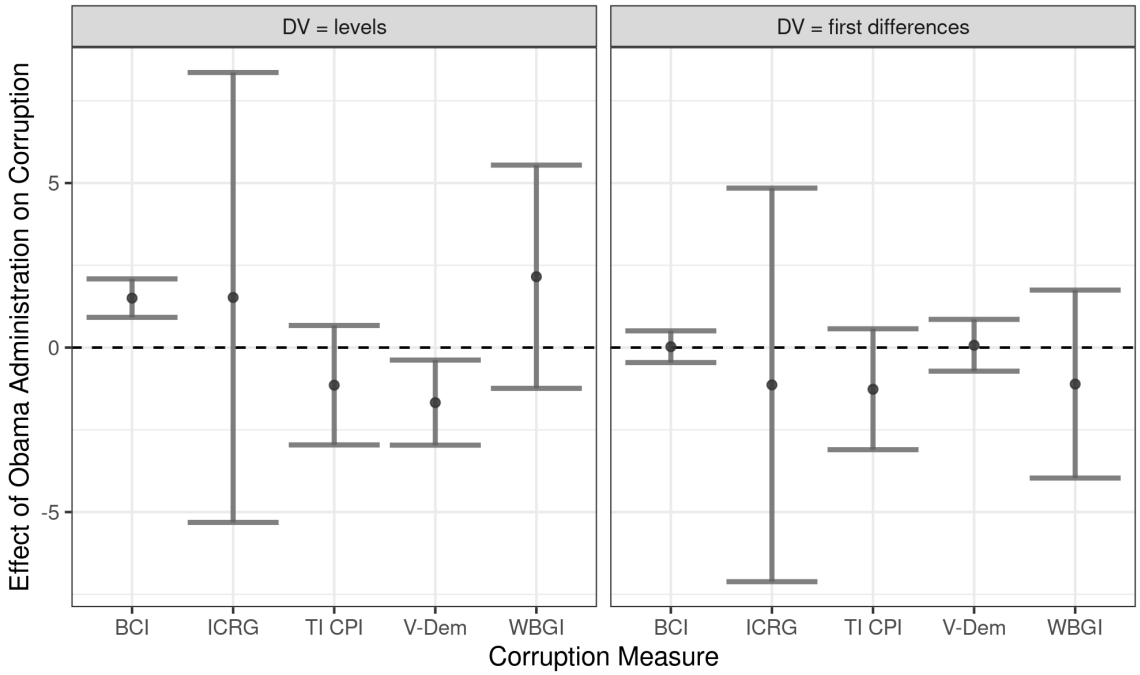

(a) China

(b) United States

Figure 1a depicts the trajectory of six corruption perception measures (the five observed panel-adjusted variables) from 1980 to 2020 for China. Figure 1b depicts the trajectory of the same corruption measures from 1980 to 2020 for the United States of America. All measures shown are centered on the country mean. Variables are scaled so that more positive numbers mean greater perceived corruption.

bounded and therefore by definition have finite mean and variance (unlike non-stationary series), the augmented Dickey-Fuller and Phillips-Perron tests still fail to reject the null of non-stationarity for this series (although the KPSS test fails to reject the null of stationarity for the same series).[11] As a robustness check, we also present results from a first-difference model to eliminate any potential unit root.

Figure 2: **Estimated effect of the Obama administration (2009-2016) on corruption in the United States**




The left panel shows the coefficient and 95% confidence interval for the *Obama* variable in the model of equation 1 using the dependent variable indicated on the $x$-axis. Each measure theoretically ranges between 0-100 across all countries and time periods, with larger values indicating more corruption. The right panel shows the same coefficient and 95% CI for the same model with the first-differenced dependent variable.

Figure 2 presents the estimated coefficient and 95% confidence interval for the *Obama* variable in the model of equation 1. The left panel uses dependent variables in levels, while

[11] We conducted these tests using the `aTSA` library (Qiu, 2015).

the right panel uses the first difference as the dependent variable. The particular corruption perception measure used as the dependent variable is indicated on the $x$-axis of each graph.

Disturbingly, when studying the level of perceived corruption (the left panel of Figure 2), the five measures give very different indications of the extent of corruption in the Obama administration. The BCI, ICRG, and WBGI all report that corruption was *higher* on average during the Obama administration compared to the rest of the time period, with the relationship reported by the BCI statistically significant at the $\alpha = 0.05$ level (two-tailed). The CPI and V-Dem measures find that corruption during the Obama administration was on average *lower* than typical, with the V-Dem result also being statistically significant at the $\alpha = 0.05$ level (two-tailed). Thus, if we interpret statistically insignificant results as being null findings (somewhat misleadingly; see Rainey, 2014), three of five measures report no difference between the Obama administration and other US presidencies, one finds that the Obama administration was more corrupt, and the last finds that the Obama administration was less corrupt. None of the models using first-differenced corruption as a dependent variable finds a statistically significant relationship between corruption and the Obama administration, although the signs of the estimates still do not agree.

## Panel Decomposition of Corruption Measures

Why do we see these discrepancies between corruption measures when studying within-country changes over time? Although the correlation between the measures we study is strong, this is attributable to differences in corruption *between* countries and not changes in corruption within countries over time (Standaert, 2015). To show this, and to create a measure specifically targeted at *within* country changes in corruption, we extract the average differences in corruption *between* countries as well as any global trends or shocks in corruption over time. We hereinafter refer to these measures as *panel-adjusted*.

Consider a theoretical decomposition of corruption measure $Y_{it}$ in a country $i$ at time $t$

over a time period $t \in \{1...T\}$:

$$Y_{it} = g_i(t) + \varepsilon_{it} \tag{2}$$

Each country $i$'s corruption level is given by a country-specific function $g_i(t)$ that represents the trajectory of corruption over time $t$ plus an added stochastic component $\varepsilon_{it} \sim f(\mu, \sigma^2)$ with mean $\mu = 0$ and variance $\sigma^2$ that represents random influences on corruption and/or pure measurement noise. We can extract between-country differences in the corruption measure by rewriting equation 2 as:

$$Y_{it} = \frac{1}{T}\sum_{t=1}^{T} g_i(t) + \left(g_i(t) - \frac{1}{T}\sum_{t=1}^{T} g_i(t)\right) + \varepsilon_{it} \tag{3}$$

$$= A_i + \gamma_i(t) + \varepsilon_{it} \tag{4}$$

where $A_i$ represents country $i$'s average corruption over the time period $t \in \{1...T\}$ and $\gamma_i(t)$ is the de-meaned function representing its trajectory over time. We speculate that the measurements of corruption we are studying (TI CPI, WBGI, BCI, V-Dem, and ICRG) can distinguish countries' average levels of corruption $A_i$ from one another more easily than countries' changes in corruption over time.

Suppose there are any common global shifts in corruption, either due to measurement noise or a genuine change in the overall level of corruption worldwide. In that case, we will need to remove these impacts from the measure as well if we are trying to study country-specific net changes in corruption.[12] Removing that component is important if (as is most common) we are studying influences on corruption that vary from country to country and are

[12]This may or may not be necessary according to the objectives of the study and our assumptions about the measurement process. If we believe that overall global trends in corruption measurement are indicative of a time-varying bias in corruption ratings, then we would certainly want to remove that trend before analysis (similar to an argument made about measures of democracy made by Little and Meng, 2023). On the other hand, if we believe that there are average global changes in corruption and these changes are relevant to the causal mechanism we were studying, we may wish not to extract global trends from the data.

not system-wide. It is also important to determine how much our corruption measures are able to distinguish country-specific changes in corruption excluding overall global changes. We therefore rewrite equation 4 as:

$$Y_{it} = A_i + \frac{1}{N}\sum_{i=1}^{N}\gamma_i(t) + \left(\gamma_i(t) - \frac{1}{N}\sum_{i=1}^{N}\gamma_i(t)\right) + \varepsilon_{it} \tag{5}$$

$$= A_i + P_t + \psi_i(t) + \varepsilon_{it} \tag{6}$$

where $P_t$ is the global average corruption measure for time $t$ and $\psi_i(t)$ represents the remaining variance in country-specific corruption.

We are able to estimate the components of equation 6 with a fixed effects model:

$$y_{it} = \sum_{i=1}^{N}\hat{\alpha}_i I_i + \sum_{t=1}^{T}\hat{\pi}_t J_t + \hat{\omega}_{it} \tag{7}$$

where $y_{it}$ is the observed value of a measure of corruption for country $i$ in year $t$, $\hat{\alpha}_i$ is the average value of the corruption measure across time in country $i$ (and a measure of $A_i$), $\hat{\pi}_t$ is the average corruption value across countries in year $t$ (and a measure of $P_t$), $\mathbf{I} = \{I_1, I_2, ..., I_N\}$ and $\mathbf{J} = \{J_1, J_2, ..., J_T\}$ are vectors of dummy variables for countries and years respectively, and all remaining variance in the corruption measure is in $\hat{\omega}_{it}$. Thus $\hat{\omega}_{it}$ is a measure of $\psi_i(t) + \varepsilon_{it}$, country-specific corruption plus (possibly systematic) error. These estimates are consistent as long as $\omega_{it}$ is uncorrelated with country and year (Wooldridge, 2010, pp. 300-301). Consequently, $\hat{\omega}_{it}$ is what we term a *panel-adjusted* measure of corruption for country $i$ at time $t$ excluding between-country variation and time-specific global shocks.[13]

With this model, we can estimate $\hat{\omega}_{it}$ for the CPI, BCI, ICRG, WBGI, and V-Dem corruption measures using least-squares dummy variables regression.[14] As before, all corruption

[13] $\alpha_i$ and $\pi_t$ may be defined relative to a reference category (without loss of generality); for example, if an overall intercept coefficient is estimated as a part of equation 7 then $\alpha_1$ and $\pi_1$ may be fixed at zero as is typical in panel dummy variable models.

[14] All analyses are conducted using R 4.2.3 (R Core Team, 2023), in this case with the basic `lm` function.

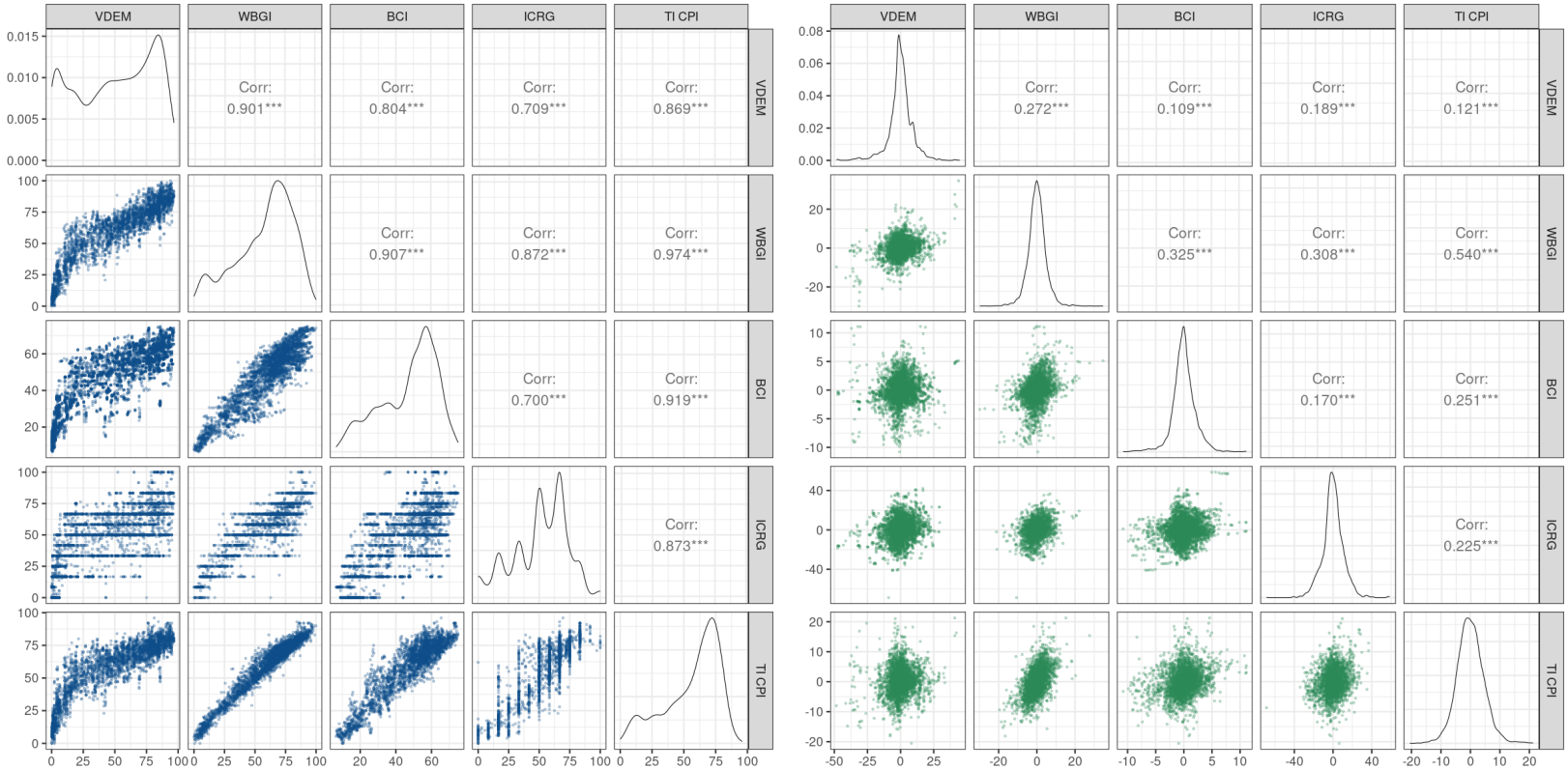
measures are set to a 0-100 scale with larger numbers indicating more corruption. We then extract the estimated residuals from the model, $\hat{\omega}_{it}$, to create the new panel-adjusted measure of corruption with between-country differences and worldwide time trends removed.

Figure 3 reports correlations among the raw corruption measures and panel-adjusted scores using the residuals from the fixed effects model in equation 7.[15] As expected, the raw annual measures are highly correlated with one another (as reported by Standaert, 2015), whereas the correlation between the panel-adjusted measures is weak (median $\hat{\rho} = 0.238$). That is, the typical corruption measure only explains about 6% of the variation in any other corruption measure once between-country differences and worldwide trends are removed. The low correlation between panel-adjusted corruption measures indicates that they cannot agree on how much a country's corruption level changes over time.

The relatively weak correlation among panel-adjusted corruption measures, and their disagreement in answering simple substantive questions about corruption, leads us to question whether the CPI, BCI, ICRG, WBGI, and V-Dem measures all correspond to the same understanding or concept of "corruption." We answer this question by proposing a relatively simple measurement model for the latent concept of corruption that will, in turn, imply a methodology allowing us to decide whether these five measures map onto the same latent concept. Although similar to the Unobserved Components Model laid out by Kaufmann, Kraay and Mastruzzi (2010), our approach does not assume any structure for non-corruption-related variance in the measures; in particular, error terms among different measures can be biased and/or correlated. We verify this methodology with a simulation study that generates data from our theoretical model of corruption measurement and demonstrates that it can accurately recover latent corruption from biased and noisy measures.

[15] For more detailed correlation figures between raw measures and fixed effect residual scores, see Table 2 in Appendix C.

Figure 3: **The correlation among raw and panel-adjusted measures of corruption**




(a) Raw Measures

(b) Panel-Adjusted Measures

Figure 3a depicts the correlation between pairs of raw measures of corruption (named at the top and right panels). Figure 3b depicts the correlation between panel-adjusted measures. The stars indicate statistical significance: $(* = p < 0.1, ** = p < 0.05, *** = p < 0.01)$.

## Measurement Model and Analysis

Suppose that we have $K$ many observed panel-adjusted measures; in our case, $K = 5$. We begin by assuming that each panel-adjusted measure $k$ is a composite of latent corruption ($\psi_{it}$ from equation 6) and extraneous variance $\xi_{itk}$:

$$\omega_{itk} = \delta_k \psi_{it} + \xi_{itk} \tag{8}$$

This implies that, if we knew the true values of $\delta_k$, $\omega_{itk}$, and $\xi_{it}$ we could recover $\psi_{it}$. We do have an estimate for $\omega_{itk}$, our panel-adjusted measure of corruption for each $k$, that we can plug in:

$$\hat{\omega}_{itk} = \delta_k \psi_{it} + \xi_{itk} - \varepsilon_{itk} \tag{9}$$

$$\hat{\omega}_{itk} = \delta_k \psi_{it} + \nu_{itk} \tag{10}$$

where in the second line we set $\nu_{itk} = \xi_{itk} + \varepsilon_{itk}$. Rewriting equation 10 in matrix form for all observations in the data set, we get:

$$\underset{NT \times K}{\hat{\Omega}} = \boldsymbol{\psi} \boldsymbol{\delta}^T + \boldsymbol{\nu} \tag{11}$$

where $\hat{\Omega}$ is the $NT \times K$ matrix of panel-adjusted corruption measures, $\boldsymbol{\psi}$ is an $NT \times 1$ vector of the latent true value of corruption, $\boldsymbol{\delta}$ is a $K \times 1$ vector mapping latent corruption into observable measures, and $\boldsymbol{\nu}$ is the $NT \times K$ matrix of remaining variance in $\hat{\Omega}$ not explained by $\boldsymbol{\psi}$.

As a model of the observed panel-adjusted corruption scores, equation 11 is simply a restatement of principal components analysis (Bro and Smilde, 2014, pp. 2816-2817). PCA chooses the vector $\boldsymbol{\delta}^T = \{\delta_1, \delta_2, ..., \delta_K\}$ in equation 11 that maximizes the variance of $\boldsymbol{\psi}$

subject to the constraint that $\sum_{k=1}^{K} \delta_k = 1$ (without loss of generality). Of course, there is no *a priori* guarantee that the $\hat{\boldsymbol{\psi}}$ estimated by PCA is "really" latent corruption. But because all our measures are designed to target the same concept (corruption), if all the observed variables load strongly onto a single dimension in the expected way it is reasonable to infer that this latent dimension *is* corruption and the degree to which $\hat{\boldsymbol{\psi}}$ explains all the observed measures of corruption is a measure of the construct validity of those measures. Although we have written equation 11 with only one common factor $\boldsymbol{\psi}$, if the matrix $\hat{\Omega}_{NT\times K}$ is of full rank then singular value decomposition of $\hat{\Omega}_{NT\times K}$ can produce $K$ many orthogonal unobserved component dimensions in descending order of the proportion of variance in $\Omega_{NT\times K}$ that each explains by making $\boldsymbol{\psi}$ into an $NT \times K$ matrix (with the $K$ columns corresponding to $K$ principal components) and $\boldsymbol{\delta}$ into a $K \times K$ matrix.

We perform probabilistic PCA analysis[16] of both the raw (unadjusted) corruption measures[17] as well as our panel-adjusted measures of corruption. If all measures do not load onto a single dimension or that dimension does not explain most of the variance, the measures may:

1. target different notions of corruption;
2. include concepts other than corruption, or;
3. might be highly contaminated by measurement error.

As we noted above, our measurement model is similar to the Unobserved Components Model of Kaufmann, Kraay and Mastruzzi (2010). However, to enable estimating the components of their model via averaging, Kaufmann, Kraay and Mastruzzi (2010) make the

[16] We use probabilistic PCA (PPCA) to account for the missingness of some corruption measures for some country years. Specifically, we use the implementation of PPCA in the `pcaMethods` package (Stacklies et al., 2007).

[17] PPCA is performed on the raw measures after they have been all been rescaled to range between 0 and 100 with larger numbers indicating more corruption. Additionally, as is standard for PCA analysis, all measures are further rescaled to set their means at zero and their standard deviations at one before PPCA is performed.

additional assumptions that errors across sources are independent ($E\left[\nu_{itj}\nu_{itk}\right] = 0$ for $j \neq k$) and homoskedastic within a source ($E\left[\nu_{itj}^2\right] = \sigma_j^2$). PCA, by contrast, does not assume any structure for $\boldsymbol{\nu}$ other than that it is orthogonal (i.e., unrelated) to $\boldsymbol{\psi}$. Thus we believe our approach improves upon Kaufmann, Kraay and Mastruzzi (2010) by being more robust to the probable dependence among measures across time and space pointed out by prior criticisms of the WBGI CCE and other TSCS country-level corruption measures.

## Simulation Study

To verify that our PCA methodology can accurately recover within-country changes in latent corruption over time using biased and noisy measures whose error terms are correlated, we simulate the process with known parameters and verify that PCA can recover those parameters. Each simulated country $i \in \{1, 2, ..., N\}$ has a true latent corruption value $\psi_{it}$ at time $t \in \{1, 2, ..., T\}$ given by:

$$\begin{aligned} \psi_{it} &= A_i + \tau_{\psi(i)} t \\ A_i &\sim U[-4, 4] \\ \tau_{\psi(i)} &\sim \phi(\mu = 0.3, \sigma = 0.3) \end{aligned} \tag{12}$$

However, there are $J$-many latent features $\lambda_j$ for $j \in \{1...J\}$ that might contaminate measures of corruption; this represents the possibility that the people and organizations that construct measures may make biased assessments of corruption using irrelevant factors (such as institutional history or cultural stereotypes) or could be using a conceptualization of corruption that is different from the target (for example, petty or grand corruption only

instead of overall corruption). For the simulation, we set $J = 2$ and make:

$$\lambda_{(ij)t} = \Lambda_{ij} + \tau_{\lambda(ij)} t \tag{13}$$

$$\Lambda_{ij} \sim \phi(\mu = 0, \sigma = 0.1)$$

$$\tau_{\lambda(ij)} \sim \phi(\mu = 0, \sigma = 0.6)$$

These are "pure" bias components in that they are completely unrelated to the target corruption dimension $\psi$; components that capture a different conceptualization of corruption would be correlated with but not identical to $\psi$. This construction is meant to mirror our theoretical idea that there are large differences in corruption between countries and that (on average) biases are zero, but within-country changes in corruption are much smaller than trends in the bias components.

Measures of corruption are constructed by presuming that each is a weighted average of both the target concept $\psi$ and the bias dimensions $\{\lambda_1, \lambda_2\}$. So for country $i$ at time $t$, the $k$th measure $M_{(it)k}$ is:

$$M_{(it)k} = p_{\psi(k)}(\psi_{it}) + \sum_{j=1}^{2} p_{j(k)}(\lambda_{(ij)t}) + \varepsilon_{it}$$

where $p_{\psi(k)} + p_{1(k)} + p_{2(k)} = 1$. For the simulation we create six measures, all of which set $p_\psi = 0.7$. The first two measures set $p_1 = 0.3$ and $p_2 = 0$, the second two set $p_1 = 0$ and $p_2 = 0.3$, and the final two set $p_1 = p_2 = 0.15$. Thus, we presume that corruption measures are "good" in the sense that they largely capture the target component of corruption but still contain substantial bias and noise. This also builds in the fact that extraneous variance in the vector of measures at time $t$ can be systematically related across measures owing to shared biases among subsets of the measures.

Finally, we vary the covariance structure among the six simulated measures to assess the

robustness of our measurement strategy to correlated measurement error. Specifically, for country $i$ at time $t$, the six-element vector of error terms $\boldsymbol{\varepsilon}^T = \{\varepsilon_{it1}, \varepsilon_{it2}, ..., \varepsilon_{it6}\}$ is distributed:

$$\boldsymbol{\varepsilon} \sim \Phi\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = 0.5 \times \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & & \ddots & & \vdots \\ \rho & \rho & \cdots & \rho & 1 \end{bmatrix}$$

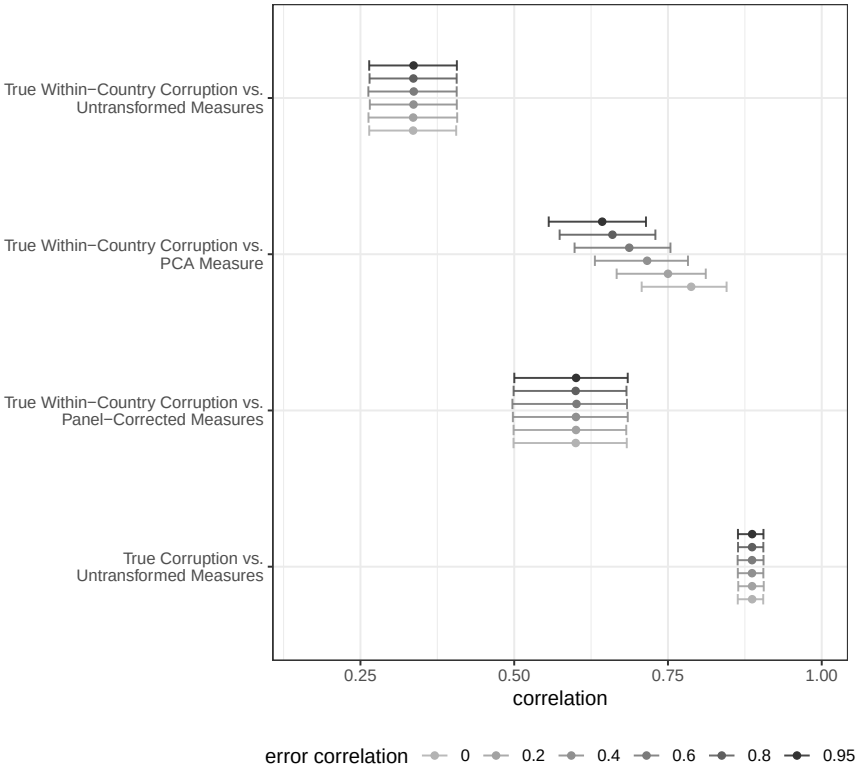
We conduct separate simulations setting $\rho \in \{0, 0.2, 0.4, 0.6, 0.8, 0.95\}$.

We simulate this data generation process 10,000 times for $N = 100$ countries and $T = 10$ time periods, thereby constructing 10,000 simulated data sets. We then construct panel-adjusted versions of all six measures in the data and create a PCA-based measure of latent corruption using those six panel-adjusted measures. This enables us to determine whether our procedure is accurate in a simulated situation where the true within-country changes in corruption are known (unlike in real data, where they are not directly observable).

Figure 4 shows the results of our simulation. The figure presents 95% simulation intervals for the correlation between within-country corruption and both raw measures (row 1), the PCA factor score extracted from panel-adjusted measures (row 2), and the panel-adjusted measures (row 3). Simulated within-country corruption is only weakly correlated with simulated raw measures of corruption (before panel correction). Panel-adjusted corruption scores and PCA-based scores in our simulation are both much more strongly correlated with within-country corruption, with the exact degree of improvement for PCA-based scores dependent on the level of correlation in error among the raw measures; the PCA-based measure is typically correlated with within-country corruption at $\approx 0.88$ when measurement errors are uncorrelated ($\rho = 0$). Finally (and unlike the panel-adjusted measures) there is a strong

average correlation between raw measures of corruption and true corruption scores that include both between-country and within-country variance (row 4), just we found among the actual (non-simulated) measures of corruption in Figure 3.

Figure 4: **Simulating the recovery of within-country corruption scores using panel correction and PCA**




Results from 10,000 simulated data sets of measurement of latent within-country corruption. Each data set simulates 100 countries over 10 years. Error bars indicate 95% simulation intervals. The color of the bar indicates the correlation among error terms for the five simulated raw measures of corruption.

With these simulated results reinforcing the robustness of our methodology to common criticisms of TSCS country-level corruption measures, we now turn to deploying this methodology on actual corruption measures (the CPI, BCI, ICRG, WBGI, and V-Dem measures) taken from the aforementioned country-year data studying 199 countries from 1980-2020.

## Results from Principal Components Analysis

Table 1 shows factor loadings for all corruption measures on the first two principal components (PC1 and PC2) produced by PPCA on the five measures of corruption in our country-year data; the row labeled $R^2$ displays the proportion of variance in corruption scores that is explained by each principal component. Among raw annual corruption scores (the first two columns in Table 1), 90% of the variance in corruption measures is accounted for by a single dimension (PC1); all corruption scores load positively on this dimension.[18] Although the dimensions extracted by PPCA do not have an intrinsic interpretation, the fact that all corruption measures load positively on a single dimension suggests that they all map onto a single concept: *aggregate corruption*.

For the panel-adjusted measures of corruption, PPCA identifies a PC1 dimension (shown in the second two columns of Table 1) with factor loadings extremely similar to the PC1 dimension for raw corruption scores. However, PC1 only explains 39.5% of the variance in the within-country change measures of corruption as compared to 90% for the raw measures. The second dimension extracted by PPCA (PC2) explains much more variance for the panel-adjusted measures compared to the second PPCA dimension for raw corruption scores. This finding suggests that corruption measures *do* track a common component of change in corruption within a country over time, but do so less accurately than they track between-country differences in corruption. Furthermore, there are other (substantively unidentified) common factors shared by these measures that explain a considerable portion of their within-country variance: PC2 explains over 21% of the variance in corruption for the panel-adjusted scores.

[18] We multiplied some factor loading matrices by $-1$ to place all first principal component loadings in the same direction.

Table 1: **Factor Loadings and $R^2$ for Principal Components**

| | raw | | panel-adjusted | |
|---|---|---|---|---|
| | PC1 | PC2 | PC1 | PC2 |
| VDEM | 0.473 | -0.362 | 0.580 | -0.767 |
| WBGI | 0.463 | -0.023 | 0.436 | 0.174 |
| BCI | 0.451 | -0.369 | 0.473 | 0.570 |
| ICRG | 0.403 | 0.856 | 0.372 | 0.065 |
| TI CPI | 0.443 | 0.006 | 0.334 | 0.227 |
| $R^2$ | 0.900 | 0.050 | 0.395 | 0.218 |

The factor loadings on the first two principal components (PC1 and PC2) for each corruption measure listed in the rows. The $R^2$ row displays the proportion of variance in corruption measures that is explained by the principal component in the column.

# Conclusion

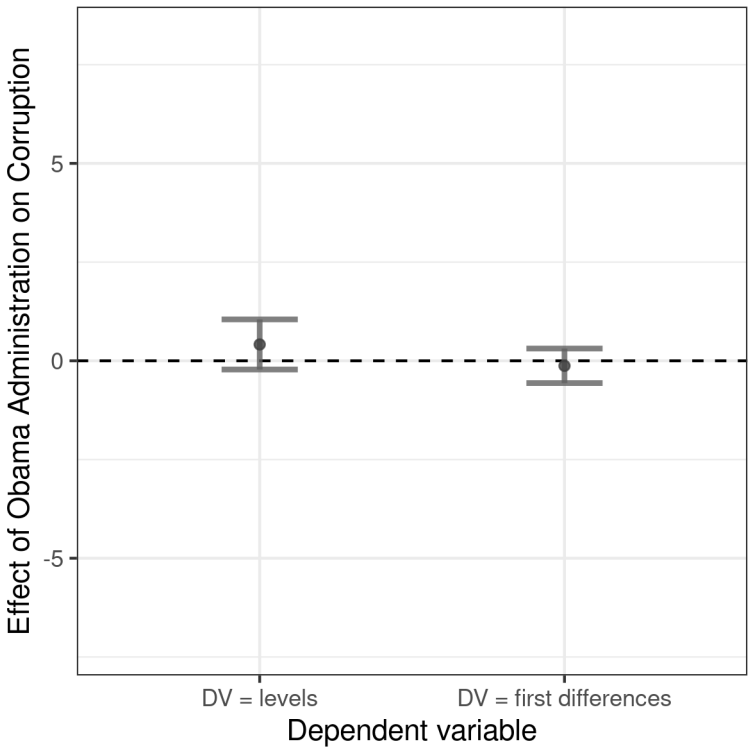
In conclusion, we return to the question we asked at the start: are criticisms of the comparability of country-level corruption perception measures still a concern for researchers, despite methodological improvements implemented after these criticisms were first made? Based on our findings, the most widely used corruption perception indicators still do not agree when measuring changes in corruption within a given country over time. Our measurement model and associated PPCA results show that changes in corruption measures *do* track a common factor corresponding to a shared concept of aggregate corruption in the country, but this factor explains relatively little of the variance in each component measure. Our PPCA model also identifies a second factor that exerts a strong influence on some measures and is substantively unidentifiable. Concordantly, we believe that the best option at present is to be careful when interpreting results that crucially depend on within-country changes in corruption.

As an alternative to using these existing measures as they are, we have shown that it is more desirable to (a) extract a common factor from multiple corruption measures using

probabilistic principal components analysis of panel-adjusted corruption measures and (b) use the resulting scores as a measure of (change in) corruption within countries. Recall our analysis in Figure 2, where we assessed the relationship between country-level corruption perception and Barack Obama's presidency using the TI CPI, WBGI CCE, BCI, ICRG, and V-Dem measures individually. In that original analysis, we found evidence that the Obama administration was more, less, or equally corrupt (compared to other administrations between 1980 and 2020) depending on the measure used. When we instead use our factor score measures to repeat that analysis, we produce the result shown in Figure 5. As we would expect from our own experience living under the Obama administration, both levels and differences of the PPCA-based corruption perception measure indicate that corruption during the Obama presidency was substantively small and statistically indistinguishable from corruption during the other years in the sample.

Not only do the results of Figure 5 have stronger face validity than many of the results in Figure 2, we believe this interpretation is bolstered by the methodological features of our PPCA-based score relative to its individual measures. Specifically, our measurement model is designed to extract common signals of change from year to year among these five measures, ignoring global trends and time-constant differences between countries, while simultaneously remaining robust to various forms of bias and error correlation that served as an important basis of criticism in past work. We provide our PPCA-based corruption perception measures for the international system between 1980 and 2020 as a part of the replication material for this paper; our hope is to make our measure easy for future researchers to study, criticize, and use in substantive work.

Figure 5: **Estimated effect of the Obama administration (2009-2016) on corruption in the United States using first principal component extracted from panel-adjusted measures of corruption using PPCA**




The panel shows the coefficient and 95% confidence interval for the *Obama* variable in the model of equation 1 using the first principal component extracted using PPCA from the panel-adjusted versions of CPI, BCI, ICRG, WBGI, and V-Dem corruption measures. The $x$-axis shows whether levels or differences of this principal component serve as the dependent variable. The dependent variable (PPCA scores) range from -7.26 to 7.97, with a mean of $\approx 0$, a median of 0.0721, a standard deviation of 1.16, and an interquartile range of 1.28.

# References


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.

Apaza, Carmen R. 2009. "Measuring Governance and Corruption through the Worldwide Governance Indicators: Critiques, Responses, and Ongoing Scholarly Discussion." *PS: Political Science & Politics* 42(1):139–143.

Arellano, Manuel and Stephen Bond. 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *The Review of Economic Studies* 58(2):277–297.

Bello y Villarino, José-Miguel. 2021. "Measuring Corruption: A Critical Analysis of the Existing Datasets and Their Suitability for Diachronic Transnational Research." *Social Indicators Research* 157(2):709–747.

Blundell, Richard and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics* 87(1):115–143.

Bro, Rasmus and Age K. Smilde. 2014. "Principal component analysis." *Analytical Methods* 6:2812–2831.

Brooks, Graham, David Walsh, Chris Lewis and Hakkyong Kim. 2013. Measuring Corruption. In *Preventing Corruption: Investigation, Enforcement, and Governance*. Palgrave Macmillan UK chapter 2.

Charron, Nicholas. 2016. "Do corruption measures have a perception problem? Assessing the relationship between experiences and perceptions of corruption among citizens and experts." *European Political Science Review* 8(1):147–171.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen et al. 2021. *V-Dem Codebook v11.1*. Varieties of Democracy (V-Dem) Project. URL: https://www.v-dem.net/en/data/data/v-dem-dataset-v111/.

Dincer, Oguzchan C. and Burak Gunalp. 2012. "Corruption and Income Inequality in the United States." *Contemporary Economic Policy* 30(2):283–292.

Donchev, Dilyan and Gergely Ujhelyi. 2014. "What Do Corruption Indices Measure?" *Economics & Politics* 26(2):309–331.

Esarey, Justin and Maya Dalton. 2023. "The Changing Relationship between Gender and Corruption." Working Paper. URL: http://www.justinesarey.com/Corruption_and_Gender_Over_Time__Dalton_and_Esarey.pdf.

Florini, Ann. 1998. "The End of Secrecy." *Foreign Policy* 111:50–63.

Galtung, Fredrik. 2006. Measuring the Unmeasurable: Boundaries and Functions of (Macro) Corruption Indices. In *Measuring Corruption*, ed. Charles J. G. Sampford, Arthur Shacklock, Carmel Connors and Fredrik Galtung. Burlington, VT: Ashgate Publishing, Ltd. pp. 101–130.

Gründler, Klaus and Niklas Potrafke. 2019. "Corruption and economic growth: New empirical evidence." *European Journal of Political Economy* 60:101810. URL: https://dx.doi.org/10.1016/j.ejpoleco.2019.08.001.

Hawken, Angela and Gerardo L. Munck. 2008. Measuring Corruption: A Critical Assessment and a Proposal. In *Tackling Corruption, Transforming Lives*, ed. Anuradha K. Rajivan and Ramesh Gampat. Macmillan India.

Heywood, Paul M. 2014. Measuring Corruption. In *Routledge Handbook of Political Corruption*. Routledge Handbooks chapter 10, pp. 137–153.

Heywood, Paul M. and Jonathan Rose. 2014. "'Close but no cigar': the measurement of corruption." *Journal of Public Policy* 34(3). Publisher: Cambridge University Press.

June, Raymond, Afroza Chowdhury, Nathaniel Heller and Jonathan Werve. 2015. "A User's Guide to Measuring Corruption." United Nations Development Program. URL: https://www.undp.org/publications/users-guide-measuring-corruption.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi. 2007. "The Worldwide Governance Indicators Project: Answering the Critics." World Bank Policy Research Working Paper. URL: http://documents.worldbank.org/curated/en/979231468178138073/The-worldwide-governance-indicators-project-answering-the-critics.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi. 2010. *The Worldwide Governance Indicators: Methodology and Analytical Issues*. SSRN. URL: http://ssrn.com/paper=1682130.

Kellstedt, Paul M. and Guy D. Whitten. 2018. *The Fundamentals of Political Science Research*. Third ed. Cambridge University Press.

Knack, Stephen. 2007. "Measuring Corruption: A Critique of Indicators in Eastern Europe and Central Asia." *Journal of Public Policy* 27(03):255–291.

Ko, Kilkon and Ananya Samajdar. 2010. "Evaluation of international corruption indexes: Should we believe them or not?" *The Social Science Journal* 47(3):508–540.

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

Lindberg, Staffan I., Maria C. Lo Bue and Kunal Sen. 2022. "Clientelism, corruption and the rule of law." *World Development* 158:105989.

Little, Andrew and Anne Meng. 2023. "Subjective and Objective Measurement of Democratic Backsliding." Social Science Research Network, January 18. Available at SSRN: https://ssrn.com/abstract=4327307 or http://dx.doi.org/10.2139/ssrn.4327307.

McMann, Kelly, Daniel Pemstein, Brigitte Seim, Jan Teorell and Staffan I Lindberg. 2016. "Strategies of Validation: Assessing the Varieties of Democracy Corruption Data." *V-Dem Working Paper Series* 2016(23). URL: https://www.v-dem.net/media/publications/v-dem_working_paper_2016_23.pdf accessed 3/19/2023.

Mondo, Bianca Vaz. 2016. "Measuring political corruption from audit results: a new panel of Brazilian municipalities." *European Journal on Criminal Policy and Research* 22:477–498.

Mungiu-Pippidi, Alina and Mihály Fazekas. 2020. How to Define and Measure Corruption. In *A research agenda for studies of corruption*, ed. Alina Mungiu-Pippidi and Paul M Heywood. Edward Elgar Publishing pp. 7–26.

Oman, Charles and Christiane Arndt. 2006. *Uses and Abuses of Governance Indicators*. Development Centre Studies OECD Publishing. URL: https://books.google.com/books?id=dvoKJhaZvm8C.

Qiu, Debin. 2015. *aTSA: Alternative Time Series Analysis*. R package version 3.1.2.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
**URL:** *https://www.R-project.org/*

Rainey, Carlisle. 2014. "Arguing for a Negligible Effect." *American Journal of Political Science* 58(4):1083–1091.

Sampford, Charles, Arthur Shacklock, Carmel Connors and Fredrik Galtung, eds. 2006. *Measuring Corruption*. Ashgate Publishing Company.

Stacklies, Wolfram, Henning Redestig, Matthias Scholz, Dirk Walther and Joachim Selbig. 2007. "pcaMethods – a Bioconductor package providing PCA methods for incomplete data." *Bioinformatics* 23:1164–1167.

Standaert, Samuel. 2015. "Divining the level of corruption: A Bayesian state-space approach." *Journal of Comparative Economics* 43(3):782–803. URL: https://doi.org/10.1016/j.jce.2014.05.007.

Tabish, S.Z.S. and Kumar Neeraj Jha. 2012. "The impact of anti-corruption strategies on corruption free performance in public construction projects." *Construction Management & Economics* 30(1):21–35.

Teorell, Jan, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Natalia Alvarado Pachon and Richard Svensson. 2019. "The Quality of Government Dataset, version Jan19." URL: http://www.qog.pol.gu.se.

The PRS Group. 2020. "International Country Risk Guide: Corruption Risk Measure." URL: https://www.prsgroup.com/explore-our-products/international-country-risk-guide/.

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

Transparency International. 2012. "Corruption Perceptions Index 2012: Technical Methodology Note." URL: https://tinyurl.com/56a4avwp accessed 7/31/2023.

Transparency International. 2016. "Technical Methodology Note." URL: https://bit.ly/2Iu9IrU accessed 6/20/2019.

Transparency International. 2020. "Corruption Perceptions Index Overview." URL: https://www.transparency.org/en/cpi/2020/index/nzl.

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. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev.polisci.10.081205.095418.

Warren, Mark E. 2004. "What Does Corruption Mean in a Democracy?" *American Journal of Political Science* 48(2):328–343.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Second ed. MIT Press.

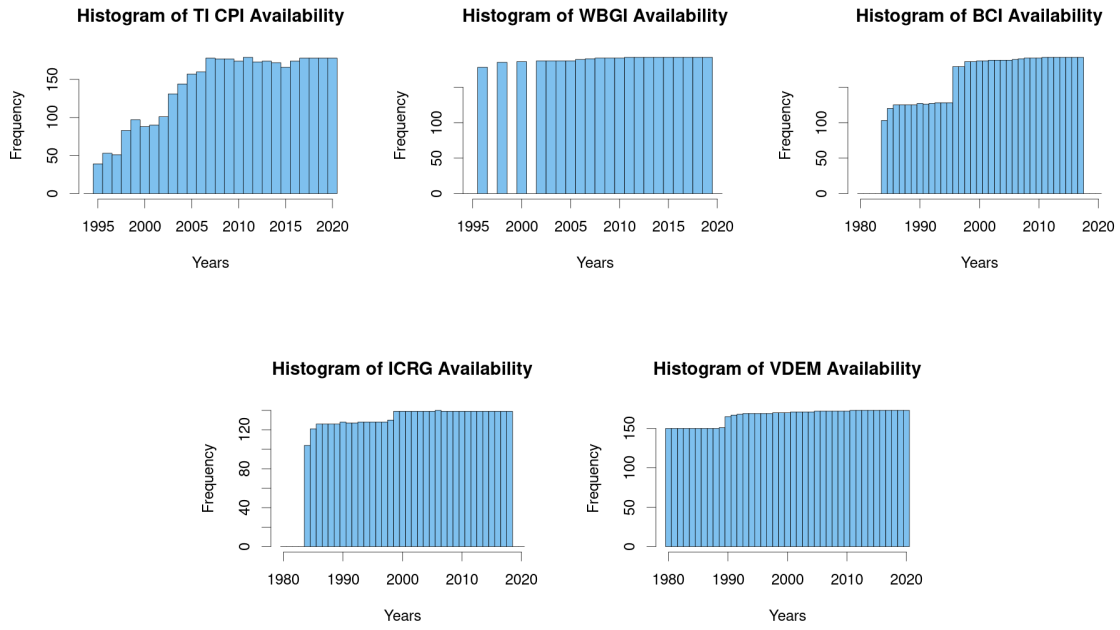
# Appendices

## A List of Countries with Available Data from 1980-2020

Afghanistan, Albania, Algeria, Andorra, Angola, Antigua and Barbuda, Azerbaijan, Argentina, Australia, Austria, Bahamas, Bahrain, Bangladesh, Armenia, Barbados, Belgium, Bhutan, Bolivia, Bosnia and Herzegovina, Botswana, Brazil, Belize, Solomon Islands, Brunei, Bulgaria, Myanmar, Burundi, Belarus, Cambodia, Cameroon, Canada, Cape Verde, Central African Republic, Sri Lanka, Chad, Chile, China, Taiwan, Colombia, Comoros, Congo, Congo, Democratic Republic, Costa Rica, Croatia, Cuba, Cyprus (1975-), Czechoslovakia, Czech Republic, Benin, Denmark, Dominica, Dominican Republic, Ecuador, El Salvador, Equatorial Guinea, Ethiopia (-1992), Ethiopia (1993-), Eritrea, Estonia, Fiji, Finland, France (1963-), Djibouti, Gabon, Georgia, Gambia, Germany, Germany, East, Germany, West, Ghana, Kiribati, Greece, Grenada, Guatemala, Guinea, Guyana, Haiti, Honduras, Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Cote d'Ivoire, Jamaica, Japan, Kazakhstan, Jordan, Kenya, Korea, North, Korea, South, Kuwait, Kyrgyzstan, Laos, Lebanon, Lesotho, Latvia, Liberia, Libya, Liechtenstein, Lithuania, Luxembourg, Madagascar, Malawi, Malaysia (1966-), Maldives, Mali, Malta, Mauritania, Mauritius, Mexico, Mongolia, Moldova, Montenegro, Morocco, Mozambique, Oman, Namibia, Nauru, Nepal, Netherlands, Vanuatu, New Zealand, Nicaragua, Niger, Nigeria, Norway, Micronesia, Marshall Islands, Palau, Pakistan (1971-), Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Guinea-Bissau, Timor-Leste, Qatar, Romania, Russia, Rwanda, St Kitts and Nevis, St Lucia, St Vincent and the Grenadines, Sao Tome and Principe, Saudi Arabia, Senegal, Serbia, Seychelles, Sierra Leone, Singapore, Slovakia, Vietnam, Slovenia, Somalia, South Africa, Zimbabwe, Spain, South Sudan, Sudan (2012-), Sudan (-2011), Suriname, Eswatini (former Swaziland), Sweden, Switzerland, Syria, Tajikistan, Thailand, Togo, Tonga, Trinidad and Tobago, United Arab Emirates, Tunisia, Turkey, Turkmenistan, Tuvalu, Uganda, Ukraine, North Macedonia, USSR, Egypt, United Kingdom, Tanzania, United States, Burkina Faso, Uruguay, Uzbekistan, Venezuela, Samoa, Yemen, Serbia and Montenegro, Zambia

# B Corruption Measure Availability by Year

Figure 6: **Availability of Corruption Measures by Year**




The availability of each of the five corruption measures during between the years 1980 to 2020, inclusive.

# C Annual Correlation Table

Table 2: **Correlation Among Corruption Measures**

| Measure | Data Type | VDEM | WBGI | BCI | ICRG | TI CPI |
|---|---|---|---|---|---|---|
| VDEM | raw score, annual | 1.000 | 0.901 | 0.804 | 0.709 | 0.869 |
| | panel-adjusted, annual | 1.000 | 0.272 | 0.109 | 0.189 | 0.121 |
| | panel-adjusted, decade avg. | 1.000 | 0.381 | 0.144 | 0.272 | 0.202 |
| WBGI | raw score, annual | 0.901 | 1.000 | 0.907 | 0.872 | 0.974 |
| | panel-adjusted, annual | 0.272 | 1.000 | 0.325 | 0.308 | 0.540 |
| | panel-adjusted, decade avg. | 0.381 | 1.000 | 0.392 | 0.417 | 0.573 |
| BCI | raw score, annual | 0.804 | 0.907 | 1.000 | 0.700 | 0.919 |
| | panel-adjusted, annual | 0.109 | 0.325 | 1.000 | 0.170 | 0.251 |
| | panel-adjusted, decade avg. | 0.144 | 0.392 | 1.000 | 0.228 | 0.246 |
| ICRG | raw score, annual | 0.709 | 0.872 | 0.700 | 1.000 | 0.873 |
| | panel-adjusted, annual | 0.189 | 0.308 | 0.170 | 1.000 | 0.225 |
| | panel-adjusted, decade avg. | 0.272 | 0.417 | 0.228 | 1.000 | 0.259 |
| TI CPI | raw score, annual | 0.869 | 0.974 | 0.919 | 0.873 | 1.000 |
| | panel-adjusted, annual | 0.121 | 0.540 | 0.251 | 0.225 | 1.000 |
| | panel-adjusted, decade avg. | 0.202 | 0.573 | 0.246 | 0.259 | 1.000 |
| | # of country-years | 6802 | 3977 | 5639 | 4662 | 3650 |

Correlation among corruption measures for raw annual scores, panel-adjusted annual scores, and panel-adjusted decennial average scores.