Measuring Changes in Corruption over Time*

Preliminary Version: This paper is under active development. Results and conclusions may change as research progresses.

Word Count: 3192

Maya Dalton† and Justin Esarey
Wake Forest University
Department of Politics and International Affairs

August 10, 2021

Abstract

Although the most commonly used quantitative measures of corruption are highly correlated, we find that these measures do not accurately track changes in corruption within a given country over time. Many research designs rely on such within-country changes to identify causal relationships. We illustrate how conclusions can change depending on the measure of corruption used by replicating a recently published, important, and sound paper that uses change in corruption as its key independent variable. We argue that findings based on changes in corruption within countries should be interpreted with caution and that factor scores extracted from multiple measures of corruption be used whenever possible.

*We gratefully acknowledge 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.

†Corresponding author: daltma18@wfu.edu

Can we employ existing measures of corruption to study its causes and effects using international panel data? Every measure of corruption is constructed differently, but all are intended to measure "abuse of public office for private gain" (Treisman, 2007, p. 211). Criticisms directed at weaknesses of a measure's methodology are less concerning if their construct validity is strong. For example, although aspects of the Transparency International's Corruption Perceptions Index's (TI CPI, Transparency International, 2020) methodology are questionable (Lambsdorff, 2006), we can might neglect these flaws if four other indicators using different methodology report similar levels of corruption.[1] On the other hand, if there is little correspondence between different measures of corruption, *at most* one of the measures corresponds to the underlying concept. Determining which (if any) of the measures is correct in this situation would require such a direct understanding of the concept that an indirect measure would hardly be necessary.

The five most commonly used quantitative measures[2] of corruption—TI CPI, the World Bank's Control of Corruption governance indicator (WBGI, The World Bank Group, 2020), the Bayesian Corruption Index (BCI, Standaert, 2015), the Varieties of Democracy overall corruption index (V-Dem, Coppedge et al., 2021), and the International Country Risk Guide's corruption risk indicator (ICRG, The PRS Group, 2020)—are highly correlated with one another, with pairwise correlation coefficients over $\rho = 0.9$ (see Table 3 in Standaert,

[1]We do not mean to single out the TI CPI as flawed; every measure we study has potential problems. The specific nature of those problems is not something we address in this paper. We posit that all of the measures may be imperfect but construct validity can enable us to have confidence in them despite the imperfections.

[2]These five measures have exerted a strong influence on corruption research, having been cited approximately 10,800 times according to Google Scholar. We produced this figure by searching for the number of citing articles for each corruption measure; for example, for the BCI we searched *("BCI" OR "Bayesian Corruption Index") AND intitle:corruption*. Google Scholar citation counts are estimates and therefore inexact. See Appendix C for a more detailed description of each measure.

2015, p. 789). However, "this 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. For this reason, we may conclude that these measures are valid for cross-sectional studies of countries but not necessarily for changes in corruption within a country over time. Within-country changes are important because causal inferences are often predicated on their accurate measurement. For example, a difference-in-differences design compares changes in corruption in a treated country to changes in an untreated country over the same time period to determine the treatment's effect on corruption (Angrist and Krueger, 1999, pp. 1293-1299). Armstrong et al. (2021) uses a variant of this design to determine how changes in corruption impact the representation of women in the cabinet.

In this paper, we report that changes in corruption over time within countries are weakly correlated across different measures whether using annual corruption scores or decennial country averages. We also report that changes in corruption across different measures load similarly on a primary factor extracted via principal component analysis, but this factor explains comparatively little variance in the component scores. We conclude that these corruption measures lack construct validity for measuring change in corruption over time. There are substantively meaningful differences between them not ascribable to transient measurement error. Therefore, we advise caution when interpreting findings that rely on measurements of change in corruption. When studying change in corruption, we believe that the best option at present is to use scores from the first principal component extracted from multiple panel-adjusted measures of corruption. We provide these scores as a part of our replication data set for researchers to use in the future.

Data

Our data set has information about 199 countries from 1980-2020 from three sources: the Quality of Government data set (QOG, Teorell et al., 2019), the V-Dem data set, and the ICRG data set.[3] The key variables in our analysis are the five influential measures of corruption we mentioned in the introduction. We re-scaled all corruption measures to range from 0 (least corrupt) to 100 (most corrupt) for ease of comparison. Some countries do not have all corruption scores available for certain years and some indicators are only available for segments of the time period.[4]

Methods

High correlation between independent measures of corruption is an indicator of their construct validity: if all measures capture the same concept, they should be closely related. Although correlation between corruption measures is strong, this is attributable to differences in corruption *between* countries and not changes in corruption within countries over time (Standaert, 2015). Hence, we need to extract differences in corruption between countries to evaluate how well indicators measure change in corruption *within* a country over time. This can be done with a fixed effects model:

$$y_{it} = \alpha_i + \phi_t + \varepsilon_{it} \tag{1}$$

where y_{it} is a measure of corruption for country i in year t, α_i is the average across time in country i, ϕ_t is the average across countries in year t, and all remaining variance is in ε_{it}. We estimate these models for each corruption measure using simple least-squares dummy variables regression.[5] We then extract the residuals from the model, ε_{it}, to create a new

[3]See Appendix A for the list of countries in our data set.

[4]See Figure 2 in Appendix B for the availability of indicators over time.

[5]All analyses are conducted using R 4.1.0 (R Core Team, 2021), in this case with the basic `lm` function.

measure of corruption with between-country differences and worldwide time trends removed; we refer to this as a *panel-adjusted corruption measure*. Finally, we calculate the correlation of all panel-adjusted corruption measures with each other.

If yearly changes in corruption are not correlated across measures, this could be a result of measurement error that averages to zero over time. Therefore, we also assess the construct validity of ten year averaged corruption scores for each panel-adjusted indicator. We intend that decade averages span a short enough time such that the change in the true level of corruption within a country is minimal. The fixed effects model from equation (1) can be applied by substituting decennial averaged corruption scores $\bar{y}_{it}$ as the dependent variable; t now indicates a decade instead of a year. Higher correlation among decade-averaged corruption scores would be consistent with stronger validity in measurements of long-run change in corruption compared to annual changes.

Principal components analysis (PCA) provides a second window into the construct validity of corruption measures (Murphy, 2012, pp. 381-416). If all our measures load onto a single dimension that explains most of their variance, it indicates that each is a measure of corruption and all agree on the level of corruption in a country-year. If all measures do not load onto a single dimension or that dimension does not explain most of the variance, the measures may target different notions of corruption, may be confounded by unobserved variables, or might be highly contaminated by measurement error. We use probabilistic PCA (PPCA) to account for the missingness of some corruption measures for some country-years.[6]

Results

In this section, we report metrics of the construct validity of the CPI, BCI, ICRG, WBGI, and V-Dem corruption measures using (i) annual scores and (ii) decennial averages. For both annual scores and decennial averages, we estimate the correlation (a) among the raw

[6]PPCA is implemented by the `pcaMethods` package (Stacklies et al., 2007).

corruption measures and (b) among the panel-adjusted measures created from the residuals of fixed effects models shown in equation (1). We also show PPCA results for all four cases.

Table 1 reports correlations among corruption measures.[7] While the raw annual measures are highly correlated with one another (as reported by Standaert, 2015), the panel-adjusted measures are only weakly correlated (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 in corruption 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. Similarly, correlation among decade averaged scores is barely higher than that for annual measures (median $\hat{\rho} = 0.265$). We therefore infer that the differences between raw and panel-adjusted annual corruption measures are not simply due to transitory measurement error but represent persistent, substantive differences between the measures.

Table 2 shows factor loadings for all corruption scores on the first two principal components (PC1 and PC2) produced by PPCA; 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 2), about 90% of the variance in corruption measures is accounted for by a single dimension (PC1); all corruption scores load positively on this dimension.[8] 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: *corruption*.

For the panel-adjusted measures of corruption, PPCA identifies a PC1 dimension (shown

[7] For more detailed correlation plots between raw measures and fixed effect residual scores, see Figure 3 in Appendix D and Figure 4 in Appendix E.

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

Table 1: **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. VDEM = the Varieties of Democracy political corruption index; WBGI = the World Bank Governance Indicators Control of Corruption measure; BCI = the Bayesian Corruption Index; ICRG = the International Country Risk Group's corruption risk measure; TI CPI = Transparency International's Corruption Perception index.

in the second two columns of Table 2) with factor loadings extremely similar to the PC1 dimension for raw corruption scores. However, PC1 only explains 39.5% of 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 raw corruption scores. This finding suggests that corruption measures *do* track a common component of change in corruption within a country over time but with much less accuracy 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 variance. For example, PC2 explains over 21% of the variance in corruption for the panel-adjusted scores. PPCA of decade averaged scores for the panel-adjusted corruption measures (shown in the last two columns in Table 2) produces results similar to those for the annual measures.

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

	annual raw		annual panel-adjusted		decade average panel-adjusted	
	PC1	PC2	PC1	PC2	PC1	PC2
VDEM	0.473	-0.362	0.580	-0.767	0.439	0.619
WBGI	0.463	-0.023	0.436	0.174	0.539	0.005
BCI	0.451	-0.369	0.473	0.570	0.421	-0.774
ICRG	0.403	0.856	0.372	0.065	0.427	0.134
TI CPI	0.443	0.006	0.334	0.227	0.397	-0.014
R^2	0.900	0.050	0.395	0.218	0.498	0.184

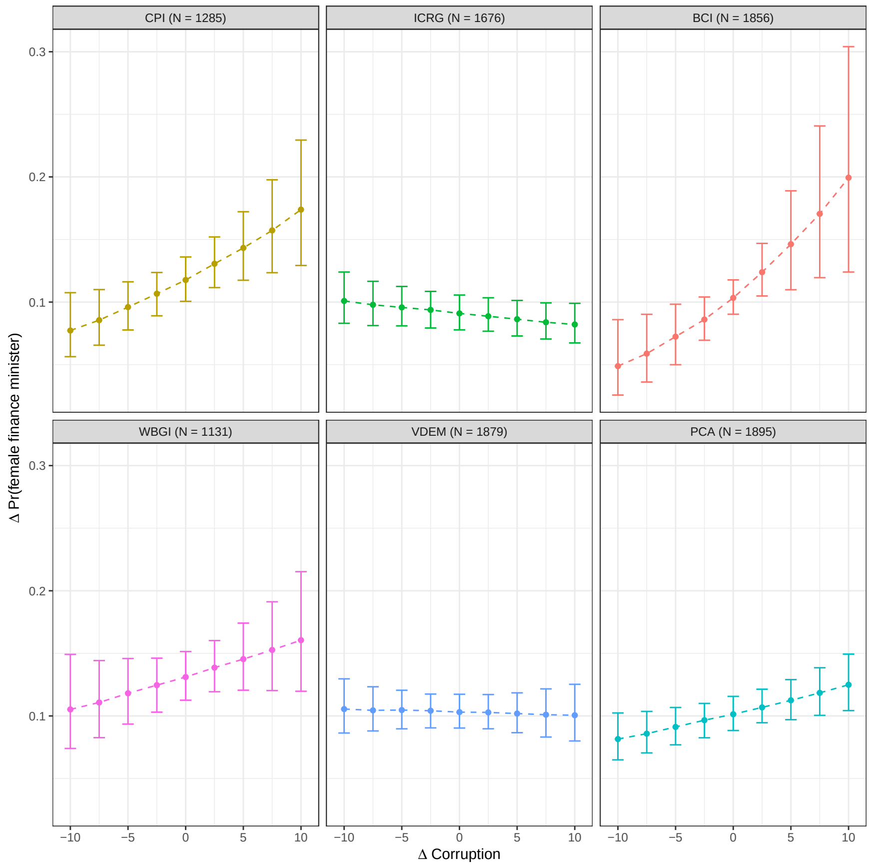
The factor loadings on the first two principal components for all corruption scores. The R^2 row displays the proportion of variance in corruption scores that is explained by each principal component.

Application

The potential scientific impact of differences among corruption measures is profitably illustrated via example. We reproduced the findings of a recent, important, and scientifically sound paper that uses change in corruption as a key independent variable: Armstrong et al. (2021). Specifically, we repeated their analysis using their measure of corruption (the TI CPI), the other four corruption measures, and the estimated scores from the first principal component dimension produced by PPCA of panel-adjusted measures. Armstrong et al.'s central argument is that "heads of government sometimes have incentives to use women as anti-corruption signals" (p. 1). They test whether "conditional on the presence of free and fair elections, women are more likely to occupy the finance ministry when countries experience increasing levels of corruption" (p. 3) using data from 150 countries between 2001-2017.

Using Armstrong et al.'s data and code, we were able to reproduce the published result

Figure 1: **Effect of Change in Corruption on** Pr(**Female Finance Minister**) **among Countries with Free and Fair Elections**



Points show the sample average estimated probability (Hanmer and Kalkan, 2013) that the finance minister is a woman among countries classified as having free and fair elections at values for change in corruption between -10 to 10. Bars indicate 95% confidence intervals. The dependent variable is the probability of the finance minister being female; the independent variable is the within-country change in corruption between six years prior to the observation and one year prior. Corruption measure and sample size are listed in the panel header. All corruption measures are scaled from 0-100 (with 100 indicating the highest corruption). The full logit models producing these results are based on Table 1, Model 2 in Armstrong et al. (2021) and are available in Appendix Table 3.

from Table 1, Model 2: positive changes in the TI CPI score in a country over the prior six years are associated with a higher probability that the finance minister is female.[9] Figure 1

[9] When using the TI CPI measure from the authors' replication data, our reproduction of their results was exact. When using the TI CPI measure from our merged data set, there were two observations for Hong

shows the sample average[10] probability that the finance minister is female when corruption changes between -10 points and 10 points over the five years prior to the observation. We find that the relationship is *not* consistent across all measures of change in corruption over time. For two measures (ICRG and V-Dem), the relationship is in the opposite direction from the published result. For the WBGI measure, increasing corruption has a positive but statistically insignificant effect on the probability of a female finance minister (two-tailed $p = 0.139$). The corruption measure in which we have the greatest confidence—factor scores derived from the first principal component of PPCA on panel-adjusted corruption scores—has an effect that is roughly *half* the magnitude of the published result despite the independent variable being rescaled to the same 0-100 range as the authors' original measure. This change is scientifically meaningful because the magnitude of a finding is directly related to its substantive importance and significance (Gross, 2015).

Conclusion

Based on our findings, we advise researchers to use caution when change in corruption over time within countries is a key independent or dependent variable. The suite of commonly used corruption indicators do not agree when measuring changes in corruption within a given country over time, even when comparing long-run average changes. Although sharing some commonalities, our replication of Armstrong et al. (2021) illustrates that they are different enough to produce incompatible conclusions when used as independent or dependent variables.

Our analysis indicates that changes in corruption measures *do* share a common factor, but this factor explains relatively little of the variance in each component measure. Concor-

Kong in the replication data that were not present in our data set that account for small, substantively unimportant differences in our findings and theirs. Table 3 in Appendix F shows the full set of results for both our reproduction of their original logit model and models using alternative measures of corruption.

[10]See Hanmer and Kalkan (2013) for details on this procedure.

dantly, we believe that the best option at present is to (a) extract this common factor 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. We have provided these scores for the international system between 1980 and 2020 as a part of the replication material for this paper to make them easy for future researchers to use.

References

Angrist, Joshua D. and Alan B. Krueger. 1999. Empirical strategies in labor economics. In *Handbook of Labor Economics*. Vol. 3 Elsevier pp. 1277–1366.

Armstrong, Brenna, Tiffany D. Barnes, Diana Z. O'Brien and Michelle Taylor-Robinson. 2021. "Corruption, Accountability, and Women's Access to Power." *The Journal of Politics* OnlineFirst:1–10. DOI: https://doi.org/10.1086/715989.

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/.

Gross, Justin H. 2015. "Testing What Matters (If You Must Test at All): A Context-Driven Approach to Substantive and Statistical Significance." *American Journal of Political Science* 59(3):775–788.

Hanmer, Michael J. and Kerem Ozan Kalkan. 2013. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models." *American Journal of Political Science* 57(1):263–277. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5907.2012.00602.x.

Lambsdorff, Johann Graf. 2006. Measuring Corruption-The Validity and Precision of Subjective Indicators (CPI). In *Measuring Corruption*, ed. Charles J. G. Sampford, Arthur

Shacklock, Carmel Connors and Fredrik Galtung. Burlington, VT: Ashgate Publishing, Ltd. pp. 81–99.

Murphy, Kevin P. 2012. *Machine Learning: A Probabilistic Perspective*. MIT Press. URL: https://www.cs.ubc.ca/~murphyk/MLbook/.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: https://www.R-project.org/.

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.

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/.

The World Bank Group. 2020. "Worldwide Governance Indicators." URL: https://datacatalog.worldbank.org/dataset/worldwide-governance-indicators.

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.

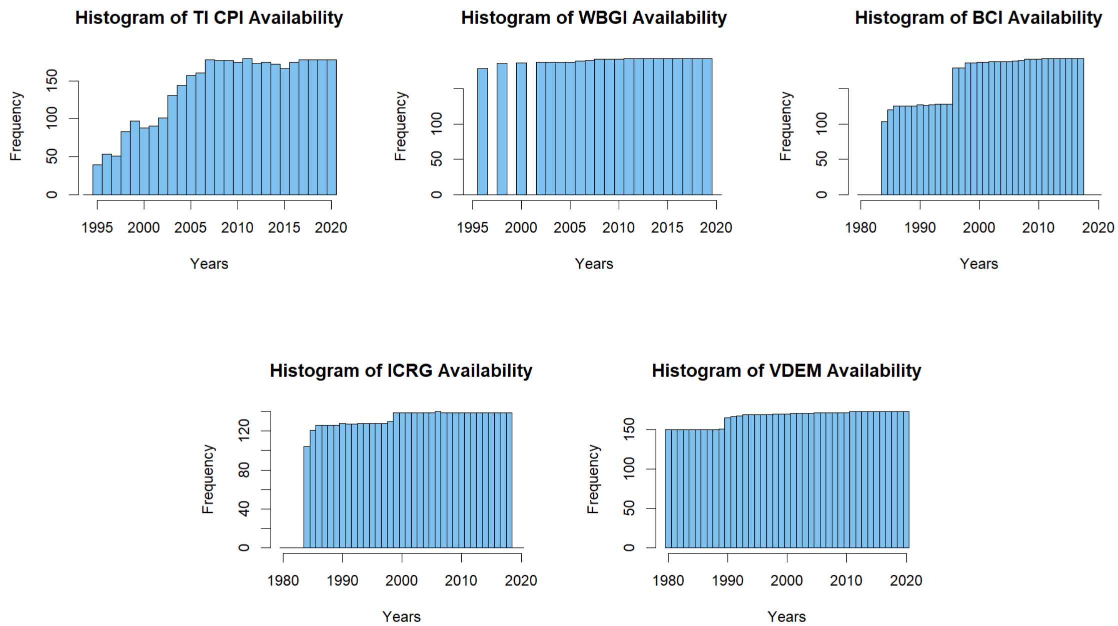
Appendices

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 2: **Availability of Corruption Measures by Year**



The availability of each of the five corruption measures during different time periods, ranging from 1980 to 2020.

C Descriptions of Frequently-Used Corruption Measures

The descriptions in this section are identical to descriptions in the online appendix of Esarey and Dalton (2021); these papers were written at the same time using (some of) the same variables.

1. *Transparency International's Corruption Perceptions Index (CPI)*[11]

   The CPI is an extremely influential indicator of corruption widely used by scholars and policymakers.[12] It is constructed from averaging at least three (but as many as thirteen) different corruption scores taken from perception-based surveys and assessments of corruption in a given country. The CPI targets corruption in the public sector within a country and compiles relevant data from multiple, independent sources. The CPI standardizes the corruption scores from these sources to the same scale, then averages the scores. Finally, the standard error and confidence interval for each country's CPI value is calculated to account for any variation in the sources. The CPI ranges from 0 (most corrupt) to 100 (least corrupt), and is available from 1995-2020.

2. *World Bank Group's Worldwide Governance Indicators (WBGI)*[13]

   The WBGI is created from 30 data sources from a variety of surveys, organizations, and governments. It utilizes a 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". The WBGI ranges from -3 (least control over corruption - highly corrupt) to 3 (most control over corruption - least corrupt), and is available for 1996, 1998, 2000, and 2002-2020.

3. *Bayesian Corruption Index (BCI)*[14]

   The BCI is an index of perceived overall corruption (abuse of public power for private gain) within a country. It is constructed from 17 different surveys from countries' inhabitants,

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

[12]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."

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

[14]Information about the BCI has been paraphrased from Standaert (2015).

business executives, and governments. The BCI expands upon the number of sources used by the WBGI and CPI and is available over a larger time span than either of these two measures, but the measurement models used by the BCI and WBGI are broadly similar. Unlike the WBGI, the BCI's measurement model accounts for variation over time to avoid discrepancies in corruption measurements and prevent selection bias. The BCI ranges between 0 (least corrupt) to 100 (most corrupt) in countries and is available from 1984 onward.

4. *Political Risk Service's International Country Risk Guide (ICRG)*[15]

   The ICRG provides political, economic, and financial risk ratings to inform businesses about potential risks to their firms when operating within certain countries. It aims to assess how much corruption within the political system can threaten foreign investment and political stability. It measures corruption through the prevalence of bribery, patronage, nepotism, extortion, suspicious ties between business and politics, and related phenomena. The measurement ranges between 0 (low political risk from corruption) and 6 (high political risk from corruption), is a source included in the CPI, and available for 1994 onward.

5. *Varieties of Democracies (V-Dem)*[16]

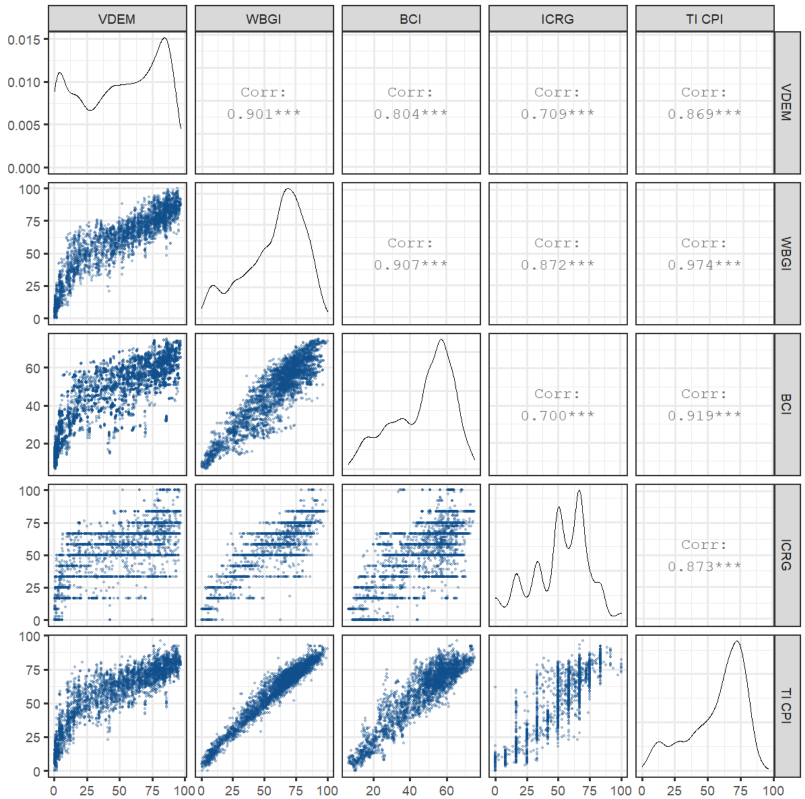
   The V-Dem project as a whole constructs 470 democracy measures created from subjective, expert-led assessments that score how well governments are performing relating to democratic ideals. One of their products is a measure of overall corruption in a country-year. This composite measure is created from averaging four other sub-indicators of corruption: (i) the public sector corruption index, (ii) the executive corruption index, (iii) a measure of legislative corruption, and (iv) a measure of judicial corruption. These four measures are in turn created from expert assessments of corruption in the corresponding sector of government. The resulting composite measure of overall corruption ranges from 0 to 1, with 0 indicating low corruption, and is available from 1980 to 2020.

[15]Information about the ICRG has been paraphrased from The PRS Group (2020)

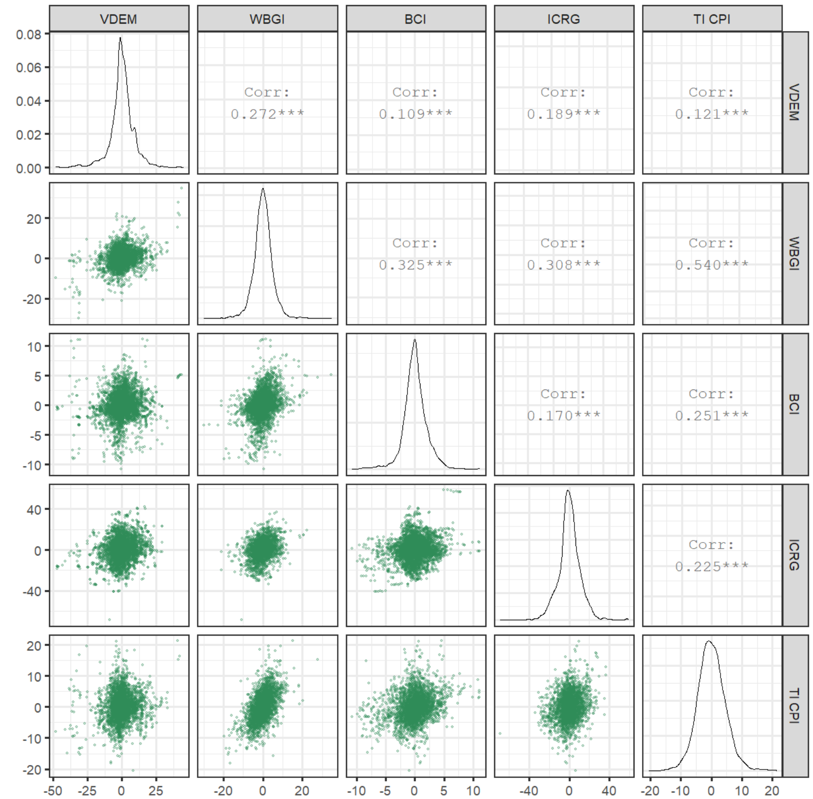
[16]Information about the V-Dem has been paraphrased from Coppedge et al. (2021)

D Annual Correlation Plot

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



(a) Raw Measures

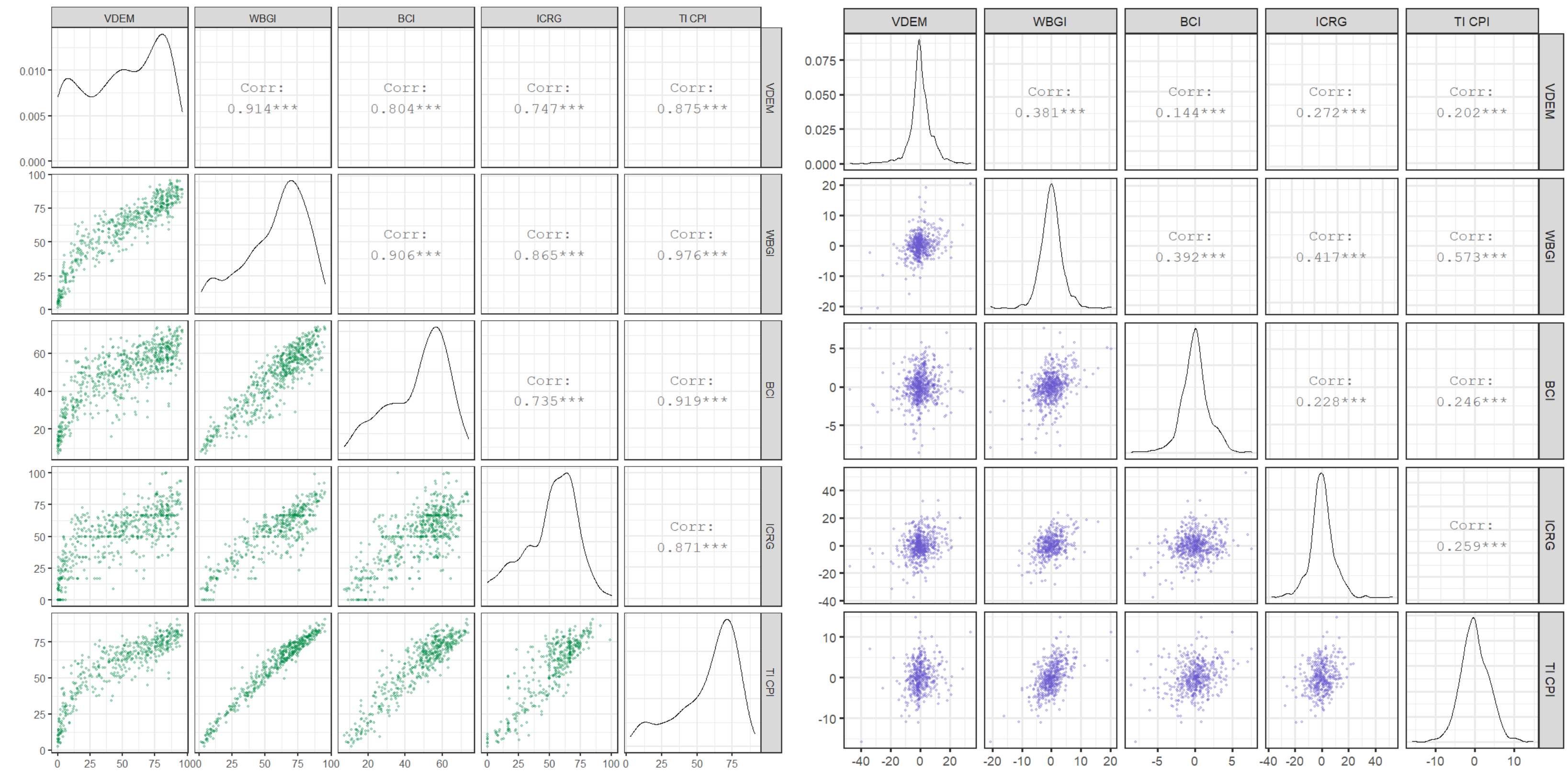


(b) Panel-Adjusted Measures

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: ($* = p < 0.1, ** = p < 0.05, *** = p < 0.01$).

E Decade Averaged Correlation Plot

Figure 4: **The correlation among raw and panel-adjusted measures of corruption averaged over decades**



(a) Raw Measures

(b) Panel Adjusted Measures

Figure 4a depicts the correlation between pairs of raw measures of corruption (named at the top and right panels) with the corruption scores averaged over decades. Figure 4b depicts the correlation between panel-adjusted measures. The stars indicate statistical significance: $(* = p < 0.1, ** = p < 0.05, *** = p < 0.01)$.

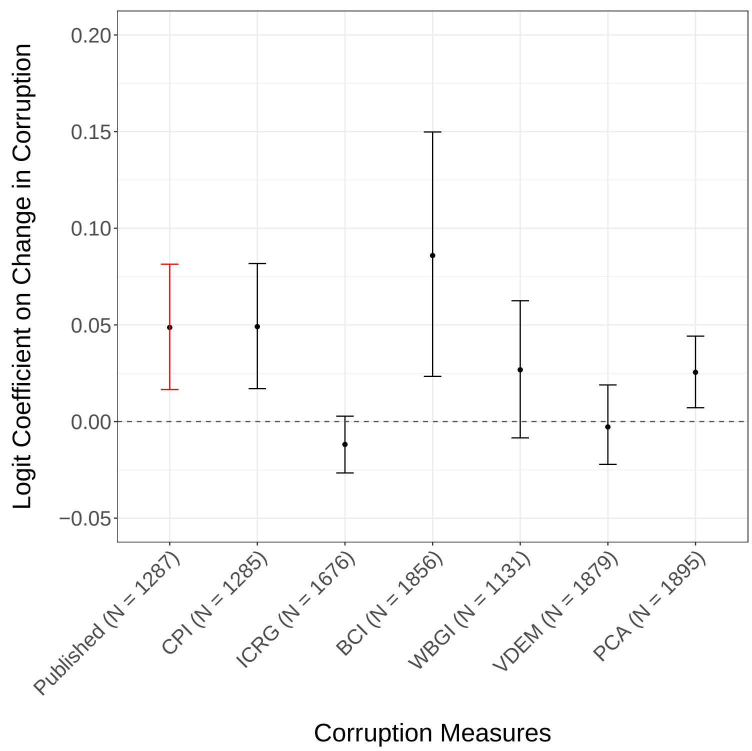
F Full Coefficients Table for Replication Analysis

Table 3: Results from Replication and Extension of Table 1, Model 2 in Armstrong et al. (2021)

	Dependent variable:						
	Female Minister						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Original Measure	0.049***						
	(0.017)						
Δ CPI		0.049***					
		(0.016)					
Δ BCI			0.086***				
			(0.032)				
Δ WBGI				0.027			
				(0.018)			
Δ VDEM					−0.003		
					(0.011)		
Δ ICRG						−0.012	
						(0.007)	
PCA Analysis							0.025***
							(0.009)
date	0.054**	0.054**	0.045***	0.013	0.050***	0.040**	0.048***
	(0.023)	(0.023)	(0.016)	(0.026)	(0.015)	(0.017)	(0.015)
gdppercapita_lag	−0.128	−0.132	−0.174**	−0.114	−0.160**	−0.111	−0.174**
	(0.088)	(0.088)	(0.076)	(0.086)	(0.076)	(0.085)	(0.076)
loggdp_lag	−0.155***	−0.155***	−0.136***	−0.196***	−0.167***	−0.069	−0.139***
	(0.060)	(0.060)	(0.050)	(0.057)	(0.052)	(0.057)	(0.049)
unified	−0.301	−0.297	−0.214	−0.421**	−0.181	−0.113	−0.238
	(0.197)	(0.197)	(0.167)	(0.198)	(0.167)	(0.188)	(0.167)
pwc	0.042***	0.042***	0.049***	0.053***	0.050***	0.040***	0.048***
	(0.008)	(0.008)	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)
presidential	0.423**	0.413*	0.383**	0.334	0.334*	0.473**	0.408**
	(0.213)	(0.214)	(0.187)	(0.213)	(0.185)	(0.210)	(0.185)
Constant	0.642	0.715	0.590	2.372*	1.135	−1.516	0.602
	(1.449)	(1.452)	(1.074)	(1.297)	(1.100)	(1.292)	(1.056)
Observations	1,287	1,285	1,856	1,131	1,879	1,676	1,895
Log Likelihood	−416.129	−414.440	−554.332	−397.638	−560.352	−468.813	−560.482
Akaike Inf. Crit.	848.258	844.881	1,124.665	811.277	1,136.703	953.626	1,136.965
Note:					*p<0.1; **p<0.05; ***p<0.01		

G Logit Coefficients for Effect of Change in Corruption on Pr(Female Finance Minister)

Figure 5: **Logit Coefficients for Replication of Table 1, Model 2 of Armstrong et al. (2021)**



Dots indicate the point estimate for the logit coefficient on Change in Corruption between $(t-1)$ and $(t-6)$ in Model 2 of Table 1 in Armstrong et al. (2021). Bars indicate 95% confidence intervals. The dependent variable is whether a woman is in charge of the finance ministry for the studied country-year.

References

Armstrong, Brenna, Tiffany D. Barnes, Diana Z. O'Brien and Michelle Taylor-Robinson. 2021. "Corruption, Accountability, and Women's Access to Power." *The Journal of Politics* OnlineFirst:1–10. DOI: https://doi.org/10.1086/715989.

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/.

Esarey, Justin and Maya Dalton. 2021. "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.

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

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.

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

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.