Governments Manipulate Official Statistics: Institutions Matter

Bruno S. Frey[*], Louis Moser[†], Sandro Bieri[‡]

October 18, 2022

[*]University of Basel and CREMA – Center for Research in Economics, Management, and the Arts, email: bruno.frey@bsfrey.ch

[†]CREMA – Center for Research in Economics, Management, and the Arts, email: louis-moser@bluewin.ch

[‡]CREMA – Center for Research in Economics, Management, and the Arts, email: sandrobieri@gmx.ch

Governments Manipulate Official Statistics: Institutions Matter

Abstract

Many governments have been reported to systematically manipulate official statistics. However, scholarly research has not extensively dealt with the determinants of data manipulation, beyond the effect of autocracy. We extend the literature by including institutional factors hypothetically affecting data manipulation.

Regressing the gap between GDP – predicted by nighttime lighting data – and „official" GDP on these institutional factors suggests that economic openness decreases manipulation, while decentralization increases manipulation. Political openness decreases manipulation for countries under-reporting GDP and increases manipulation for countries over-reporting GDP. Surprisingly, no effects are found for press freedom and the independence of the statistical office.

Keywords: Data Manipulation; Official Statistics; Political Economy; Government; Institutions

JEL Codes: E01; E02; H11; P44

1 Introduction

Winston Churchill is said to have quipped: *"I only believe in statistics that I have doctored myself"*. Reports in the media support this skeptical view of official statistics. European countries, such as Greece and Italy, have been accused of falsifying the size of their budget deficit and government debt in the context of entering the Euro system. Other countries, such as Argentina, Turkey, and China, have reportedly manipulated many official economic statistics.[1] How substantial these manipulations can be is demonstrated for Turkey. According to The Economist (2022d), "in late June a group of researchers put inflation in Turkey at 160%, double the official rate of 79%. A survey showed that seven out of ten Turks believed that group's figures rather than the government's."

The scientific literature in economics and other disciplines has largely disregarded the extent and determinants of the manipulation of official statistics. Even when authors discuss biased statistics, they rarely deal with governments' incentives for manipulation (e.g. Feldstein 2017).

Knowledge about the true state of an economy is important for various reasons. Policymakers themselves are induced to undertake mistaken interventions into the economy if the data upon which decisions are taken is not correct. This holds even if they know that manipulations have been undertaken, as they are not precisely informed about the current economic situation. Actors on a lower level of the state hierarchy may be less informed about the manipulations and may take the official statistics to reflect the actual conditions of the economy. This again tends to bias governmental interventions.

International organizations may, likewise, undertake mistaken policies if a country reports manipulated data. For instance, they may hand out

[1] A collection of some media articles about this issue can be found in The Economist (2014, 2015, 2017, 2020, 2021a, 2021b, 2022b, 2022c, 2022a, 2022d)

subsidies and aid programs, which are too high than what would be correct in view of the real state of the recipients' economies. Moreover, countries could be included in newly established monetary systems, although they do not meet the requirements.

Knowledge about the gap between the real and the officially communicated state of the economy is also important for research. Many empirical analyses, particularly econometric ones, employ officially published data to analyze and predict economic factors. If these data are systematically biased, a mistaken view of the economy and distorted predictions may result.

Our paper seeks to contribute to a better theoretical and empirical understanding of data falsification. Building on political economic theory we establish five hypotheses, which aim to explain why some governments more often manipulate official data than others. Specifically, we hypothesize that a country's openness, freedom of press, and independence of the statistical office decrease, and decentralization increases a country's incentives for manipulating official data. To test these hypotheses we use nighttime lighting captured by satellites to predict the economic output. We explain the deviation of "official" GDP from "true" GDP, predicted by nighttime lights, with variables based on our hypotheses. For this purpose, we construct a panel-data set consisting of 195 countries for the years 2013 to 2019.

Our results suggest that economic globalization decreases data manipulation, while decentralization increases data manipulation. Both of these findings support our hypotheses. We find a negative relation between political globalization and data manipulation only for countries under-reporting GDP statistics. For countries over-reporting GDP our analysis suggests that political globalization increases data manipulation. Finally, we do not find that the freedom of the press and the independence of the statistical office are associated with data manipulation.

This article contributes to the yet relatively sparse literature on data manipulation by governments. While most of the current literature has focused on democracy, electoral cycles, and dictatorship as causes of differences in data manipulation, we enrich the literature with additional political economic hypotheses, which include different institutional elements. However, due to serious data and methodological limitations, our results should not be interpreted as necessarily being causal. Our contribution, rather, provides first correlational insight into hypotheses that go beyond existing research.

Section 2 outlines both the theoretical and empirical contributions to the literature on systematic data manipulation of official data. Subsequently, empirically testable hypotheses on the determinants of data manipulation are derived. Section 3 outlines the various variables and data sources. The methodology employed for this purpose is presented in section 4. Section 5 presents the empirical results and section 6 is devoted to robustness checks. Section 7 concludes.

2 Literature

2.1 Theoretical Background

The way governments manipulate official data is discussed only scarcely in previous literature. Georgiou (2021) identifies the manipulation of official statistics as corruption, i.e., "acts committed [...] enabling leaders to benefit at the expense of the public good" (p. 86). By *acts committed*, he refers to manipulating statistical procedures by setting certain standards or controlling data sources. For the case of Russia, Kalgin (2016) analyzes data manipulation by two types of behavior. Bureaucrats were either found to be prudent and did not like to report a too high variation in their reported data. Such bureaucrats kept variation especially low when manipulating data, whereas others strategically inflated their numbers for better performance. With performance measurements and in the absence of other incentives for data manipulation, such as targets and rankings, it may be a dominant strategy for individual bureaucrats to keep a low profile by reducing the variation in the data reported. Going a step further, Aragão and Linsi (2022) classify four different types of data manipulation. First, outright manipulation, referring to the situation where the true statistical figure is known, but the government pushes for the publication of different figures. Second, guessing the statistical number, even if it is not known. Third, using other statistical methods to achieve more convenient results. And fourth, intervening through taking indirect means, for example, by manipulating data sources before the statistical analysis is performed. If economic data is manipulated, it is likely through one of these four methods, as the authors show in case studies for Greece, Argentina, and Brazil.

When manipulating official statistics, the most evident goal of governments is to cover up poor performance. The pressure to misreport official data can emerge both from within a country or on an international level. On one hand, governments control information by manipulating data to influence popular opinion (King, Pan, and Roberts 2017; Lorentzen 2014). On the other hand, pressure to manipulate statistics can be induced by international organizations, for example, when there is a need to comply with international standards (Alt, Lassen, and Wehner 2014; Rauch et al. 2011; Dafflon and Rossi 1999).

2.2 Empirical Studies

Estimating data manipulation has been a methodological challenge. Some authors try to detect manipulation with Benford's law (e.g. Bond et al. 2022; Adiguzel, Cansunar, and Corekcioglu 2020; Adam and Tsarsitalidou 2022), which identifies anomalies in the occurrence of certain numbers. Other studies rely on surveys or interviews. Newer possibilities have emerged with the analysis of satellite data, where economic activity is approximated by satellite imagery. This allows to approximate "true" GDP or economic growth (Henderson, Storeygard, and Weil 2012; Ghosh et al. 2010; Hodler and Raschky 2014; Chen and Nordhaus 2011). The same method has also been applied to various other issues in economics, such as poverty (Jean et al. 2016), local wealth (Weidmann and Schutte 2017), health (Ebener et al. 2005), and ecological economics (Sutton and Costanza 2002). It is also used as a control variable in contexts where data on economic activity is not available on a subnational level.

A large share of the most recent empirical contributions focused on the manipulation of Covid statistics. Using Benford's law, it has been shown that reported Covid-19 cases were more likely to show signs of statistical manipulation in less democratic countries (Adiguzel, Cansunar, and Corekcioglu 2020; Adam and Tsarsitalidou 2022) or countries with lower economic development (Balashov, Yan, and Zhu 2021). Authoritarianism as a cause for statistical manipulation has often been mentioned in the literature also outside manipulated Covid-19 statistics. In an important recent study using nightlight satellite data, Martinez (2022) shows that the more authoritarian a state is, the more it tends to inflate its GDP growth rates. This result is supported by Magee and Doces (2015). Two factors are taken to increase the incentives for authoritarian states to manipulate official data: First, authoritarian states and countries with weaker democratic institutions have fewer constraints on the government. Therefore, there is less control over the correct use of data and application of statistical methods (Aragão and Linsi 2022; Magee and Doces 2015). Second, authoritarian states can experience much pressure in times of crisis as these regimes derive a part of their legitimacy from good economic performance (Nathan 2020). Developing countries are generally more prone to manipulate official data. Sandefur and Glassman (2015) provide survey and administrative data evidence on how in African regions with lower economic development, two principal-agent problems lead to insufficient data quality. Statistics on economic progress are exaggerated because of the need to attract more donations.

Other empirical studies do not focus on cross-country evidence but on case studies. Newspaper articles, as well as academic research, have especially focused on China (e.g., Deaton 2013; Wang and Yang 2021). A nightlight analysis shows that the prospect of career promotion increases the probability of county officials inflating local GDP numbers. However, this effect is less pronounced for county officials with more accountability (Chen, Qiao, and Zhu 2021). Two implications follow: first, accountability increases the trust one can put into public data; second, systematic manipulation of data occurs when political leaders benefit from false data.

Democratic regimes are not exempt from manipulating economic data. Argentina and Greece, for example, have often been suspected of falsifying official statistics. Coremberg (2014) reproduce Argentina's GDP data and find that the reported GDP data is too high, putting into doubt that the country has the highest growth rates in South America. Greece, especially in the period around the financial crisis in 2008/09, has also been found to manipulate economic data in order to comply with European Union criteria (Rauch et al. 2011). Many governments in the EU indeed engage in "creative accounting" (Milesi-Ferretti 2004; Von Hagen and Wolff 2006; Koen and Van den Noord 2005).

Electoral cycles are another cause found to pressure governments to manipulate official data. Martinez (2022) shows that for authoritarian regimes the deviation from GDP estimated by nightlight data is higher just before elections. A similar effect was also found for democracies (Alt, Lassen, and Wehner 2014). Further, times of crisis, characterized by low growth rates, appear to provide strong incentives for authoritarian governments (Martinez 2022; Wallace 2016) as well as democracies to manipulate data (Alt, Lassen, and Wehner 2014; Chan et al. 2019).

2.3 Hypotheses

Previous literature has shown that pressure from the international or domestic sphere can influence the incentives for manipulating official data. While most literature focuses on factors making data manipulation more likely, there can also be institutional and economic factors making data manipulation less likely. Government actors having an interest in falsifying official data are constrained by various institutional and constitutional arrangements (see for example Buchanan and Tullock 1965). Democracy, for example, can be understood as a constitutional arrangement ensuring that the government acts in the interest of the population. In this sense, government actors are constrained in their desire to manipulate data, because they have to act in the interest of the public. Additional arrangements constraining or encouraging data manipulation are discussed in the following.

One factor leading to lower levels of data manipulation could be the extent of openness of a country. Political openness, characterized by membership in international organizations, can result in higher compliance with international standards. When joining an international organization, the statistical bureau is assumed to follow the „codes of statistical practice" agreed upon. Membership in the OECD, IMF, or the World Bank can constrain national statistical offices because international organizations monitor the work performed by a national statistical office. When the official data is questionable, it may be controlled by the data agency of the international organization or even collected by it, as was the case with the European Commission (European Commission 2009). This leads to our first hypothesis:

Hypothesis 1: *The more politically open a country is, the less official statistics are manipulated.*

Economic openness can pressure governments to report correct statistics. First, the more economically integrated a country is, the easier it is to reproduce government data. This was shown, for example, in the case of Myanmar (Kubo 2012). Second, economic openness also improves the flow of information and can foster the growth of media (Yang and Shanahan 2003). A well-informed civil society is more able to evaluate the current state of their national economy and is less likely to be misled by false statistics. From this follows the second hypothesis:

Hypothesis 2: *The more economically open a country is, the less official statistics are manipulated.*

Domestic pressures to report correct national data can come from a free press. Large media corporations themselves are often engaged in collecting data and may question unreliable government data. They inform individuals, such as voters, members of interest groups, or even opposition parties, empowering them to take action against the government in power. A free press reduces the possibility of governments misreporting data as it can more easily be discovered and publicized. Evidence of a positive correlation between good statistical performance and freedom of the media has been reported by the World Bank (2021, p. 69). For the case of terrorist incidence, underreporting was found for countries that are more authoritarian and the press of which is less free (Drakos and Gofas 2006). Thus, the third hypothesis reads as follows:

Hypothesis 3: *The greater the freedom of the press is, the less official statistics are manipulated.*

Other literature suggests a relationship between decentralization and data manipulation. However, the direction of this relationship is not clear.

On the one hand, the existence of subnational units could constrain the possibilities of governments to manipulate data. For example, if government data represents an aggregate of subnational data, it is difficult to manipulate. However, most of the empirical evidence points in the opposite direction. The study about performance reviews in Russia mentions the distrust of government officials in data from local departments (Kalgin 2016, p. 119). Further evidence was found in African development data, where local officials manipulated their data to get more funding (Sandefur and Glassman 2015). This shows that local governments can have incentives to misreport data in order to signal higher performance. Koen and Van den Noord (2005) show that the use of alternative statistical procedures and "gimmicks" is less often observed in more centralized countries. This leads to the fourth hypothesis:

Hypothesis 4: *The more decentralized a country is, the more official statistics are manipulated.*

The final hypothesis deals with the relationship between the independence of the statistical office and the level of data manipulation. The World Bank (2021, p. 69) finds that the presence of an independent statistical office is correlated with better statistical performance. An independent national bureau of statistics is exposed to different incentives than the government, similar to an independent central bank. Furthermore, the process of publishing official data will become less arbitrary. This leads to the fifth hypothesis:

Hypothesis 5: *The more autonomous the official statistics bureau is, the less official statistics are manipulated.*

3 Data

In order to test the five hypotheses, we construct a panel-dataset including the relevant variables for 195 countries for the years 2013 to 2019. We limit ourselves to this time period because data coverage is seriously limited before 2013. After 2019, GDP was highly affected by the Covid pandemic; therefore, the years after 2019 are excluded. The unit of observation is the country-year level. Summary statistics for all variables are shown in table A1.

3.1 Dependent Variable

Nightlight Data: We use annual VIIRS Nighttime Lights (VNL) V2, made available by the Earth Observation Group (Elvidge et al. 2021), as our estimate for *true* economic activity. Nightlight data has been shown to be a viable proxy for economic activity (Sutton, Elvidge, Ghosh, et al. 2007; Chen and Nordhaus 2011; Henderson, Storeygard, and Weil 2012), and has the benefit of not being an *official* statistic. Rather, it is independently and identically measured for all countries through remote sensing.

The annual VNL data are aggregated on a country level by taking the mean of light values in each country's boundaries for each year from 2013 to 2019. The country boundaries are taken from Natural Earth (2009-2022). Because the distribution of nighttime lights per country and year is right skewed, the data is logarithmically transformed.

GDP: As our measure of *official* GDP, we take GDP reporting from the World Bank (2022). This variable measures GDP on a country-year level in current US-$. The dollar values of GDP are converted from domestic currencies for every year using single year official exchange rates. Alternative conversion factors are used if the official exchange rate does not reflect the rate effectively applied to actual foreign exchange transactions. The World

Bank takes reports from national accounts and sometimes makes adjustments to match international standards. Again, we perform a logarithmic transformation of the data.

Other Measures of Data Quality: For robustness we use additional measures of data quality. The Open Data Barometer (ODB) and the Open Data Inventory (ODIN) are used as dependent variables in robustness checks (World Wide Web Foundation 2013-2017; Open Data Watch 2015-2020). ODB data is available for the years 2013 to 2016, and ODIN data for the years 2015 to 2020, which limits the robustness checks to these years respectively. Both, ODB and ODIN measure the transparency, availability, and quality of official data, all factors which make data misreporting less likely. They are, however, not direct measures of data misreporting.

3.2 Independent Variables

Independence of Statistical Agency: Data on the independence of the statistical agency is taken from the Ibrahim Index of African Governance (IIAG) (Mo Ibrahim Foundation 2020). The IIAG data is, however, limited to African countries. The measure used, indicates the degree of independence of the statistical office from the government in quarter-steps on a scale from 0 to 100.

Economic Openness: The measure of economic openness used, is the de facto trade globalization from the KOF Globalization Index (KOF Swiss Economic Institute 2021). The variable comprises the percentage of trade in goods and services of a country's GDP and a country's trade partner diversity.

Political Openness: Political openness is measured by using de jure political globalization from the KOF Globalization Index (KOF Swiss Economic Institute 2021). This measure comprises a country's membership in international organizations and treaties and the treaty partner diversity.

Freedom of Press: The measure of press freedom used is the World Press Freedom Index from Reporters Without Borders (2002-2022). This index provides a score between 0 and 100 for 180 countries around the world.

Decentralization: As a measure for decentralization, the Regional Offices Relative Power variable from the V-Dem Institute (2022) is used.

3.3 Control Variables

Democracy: The Electoral Democracy Index from V-Dem is used to measure democracy (V-Dem Institute 2022). This index captures the extent to which the ideal of electoral democracy is in its fullest sense achieved on a scale from 0 to 1.

Contribution to GDP by Sector: The contributions of the different economic sectors to GDP are controlled for, as different sectors might emit different amounts of light (e.g. Bhandari and Roychowdhury 2011). A sector contributing greatly to GDP, but emitting little light, leads to a difference in economic activity measured in nighttime lights from official GDP. Our measure of the contribution of different sectors to GDP is taken from the World Bank (1990-2021). With this data, four variables can be created – each measuring the share of the respective sector in total output. The four sectors classified by the World Bank are agriculture, industry, manufacturing, and services.

Informal Economy: A big informal economy can also lead to a difference in economic activity as measured in nighttime lights from official GDP if the informal economy emits light, which is generally not included in the official GDP measure (e.g. Ghosh et al. 2009). Our variable of the informal economy is from the World Bank (1990-2018). It is a dynamic general equilibrium (DGE) model estimate of the share of informal output of total GDP.

4 Methodology

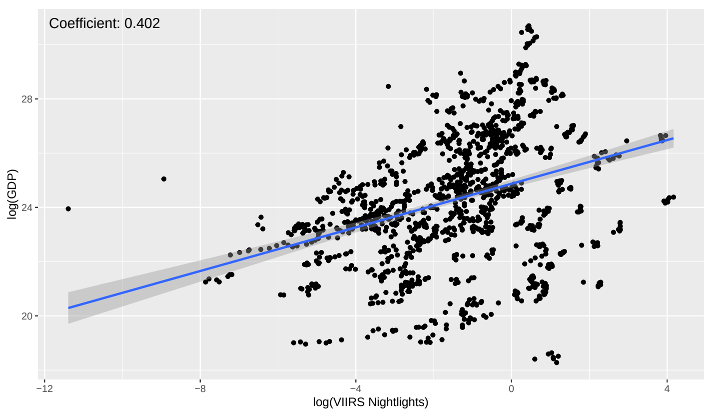
To test our hypotheses we proceed in two steps. First, we estimate how much each country's official GDP statistic deviates from the one predicted by nighttime lighting. Second, we explain these deviations with the variables suggested by our hypotheses.

In the first step, official GDP is regressed on VIIRS nighttime lights via OLS:

$$gdp_{c,t} = vnl_{c,t} + y_{c,t} \tag{1}$$

where $gdp_{c,t}$ is the logarithm of the official GDP measure for a country c in year t, $vnl_{c,t}$ is the logarithm of the economic activity measured by VIIRS nighttime lights and $y_{c,t}$ is the error term, i.e. the variance of GDP not explained by nighttime lights. The residuals $y_{c,t}$ measure the possible scale of manipulation of GDP data and are used as the dependent variable in the second step of the analysis. Figure 1 shows the results of the first regression.

Figure 1: Regressing GDP on Nightlights



This figure shows the data points, regression line and 95%-Confidence Interval of the regression in equation (1).

In a second step, the residuals from regression (1), i.e. the deviation of the official GDP from the predicted GDP by nighttime lights, are regressed on the variables of interest. We estimate the following OLS regression:

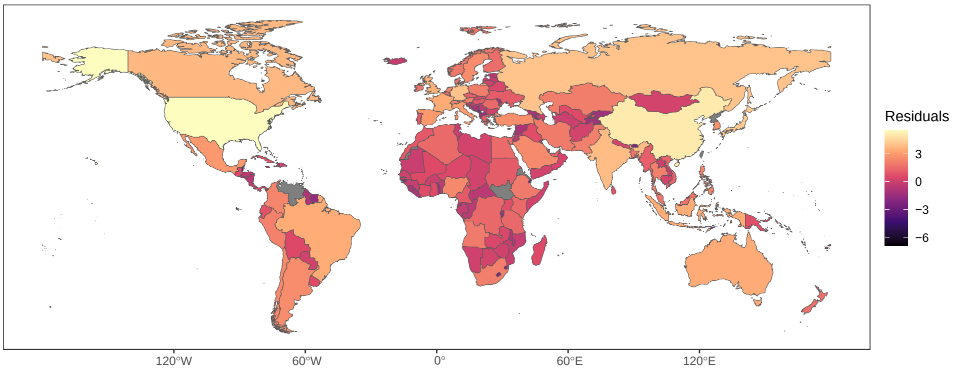
$$y_{c,t} = \alpha + \beta_1 iso_{c,t} + \beta_2 po_{c,t} + \beta_3 eo_{c,t} + \beta_4 pf_{c,t} + \beta_5 dec_{c,t} + \mathbf{K}\gamma + \theta_t + \epsilon_{c,t} \tag{2}$$

where $y_{c,t}$ is our aforementioned measure for possible data manipulation for country c for year t and α is a constant. $iso_{c,t}$ measures the independence of the statistical office, $po_{c,t}$ measures the political openness of a country, $eo_{c,t}$ is a measure of the economic openness of a country, $pf_{c,t}$ measures freedom of the press, and $dec_{c,t}$ measures of the extent of decentralization in a country. $\mathbf{K}$ is a vector of control variables, further discussed below. Time fixed effects θ_t are included to control for year-specific unobservables. $\epsilon_{c,t}$ represents the error term. We refrain from using country fixed effects because the variation over these few years in our variables is minimal.

As $y_{c,t}$ does not directly measure data manipulation, but only the possible scale of manipulation of GDP data, control variables $\mathbf{K}$ are included in the main regression. They capture other factors, which may explain why a country's official GDP is different from the GDP measured by nightlight. The vector $\mathbf{K}$ includes the shares of different sectors contributing to GDP, a measure of the extent of the informal economy, and the degree of democracy.

The method of analyzing the residuals of official GDP from predicted GDP by nighttime lights allows us to analyze both over- and under-reporting of official GDP numbers. Positive residuals suggest that there is over-reporting, whereas negative residuals suggest under-reporting of official GDP statistics. We present results for subsets of countries with either

Figure 2: Spatial Distribution of the Dependent Variable



This figure shows the spatial distribution of the residuals for the year 2017, when regressing GDP on nighttime lights, as in equation (1).

possible over-reporting or under-reporting, respectively. Figure 2 shows the spatial distribution of the residuals.

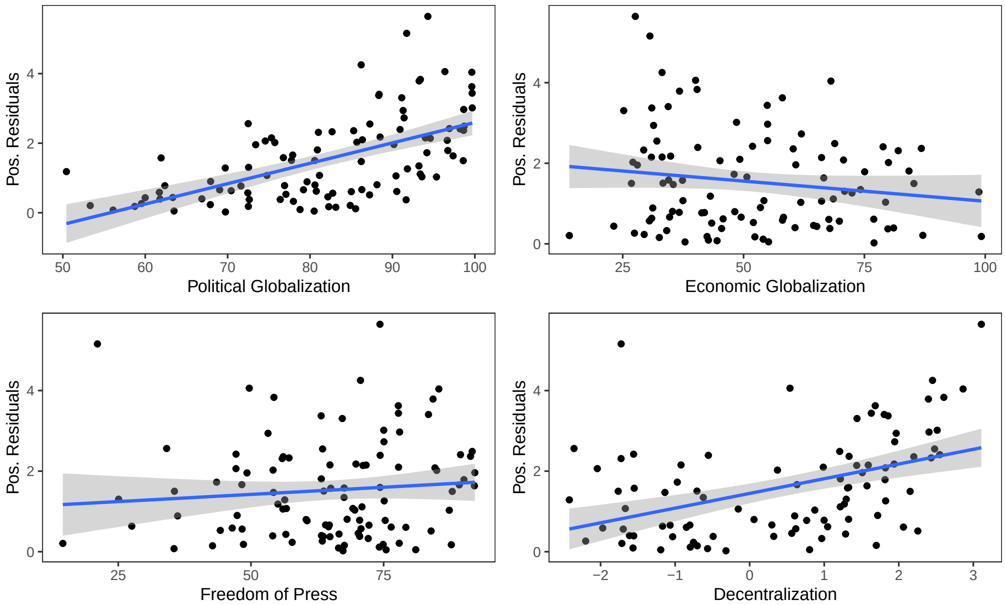
By using nighttime lights we have a clear advantage over other methods of detecting data manipulation. While Benford's law can best account for methods of manipulation where numbers are directly altered, nighttime light analysis can account for other methods of manipulation. This includes those four methods of manipulation introduced by Aragão and Linsi (2022). For an analysis of more than one country, it is useful to have a comparable quantitative measure rather than surveys and interviews, which have more validity for case studies.

5 Results

5.1 Simple Correlations

We start by showing the results of simple bivariate regressions to display the relationship between our explanatory variables and the positive residuals from regressing official GDP data on nighttime lights. These regressions are based on a cross-sectional subset of all countries in our data set for the year 2019. The correlation plots in figure 3 show the results. The independence of the statistical office is excluded from this analysis, as data is only available for African countries.

Figure 3: Correlation Plots for Over-Reporting of GDP



This figure shows correlation plots for the results of regressing the positive residuals on each of the explanatory variables using a subset of the data for the year 2019.

The correlation plots show correlations in the expected direction for hypotheses 2 and 4. Higher economic openness is related to lower over-reporting and powerful subnational units are associated with an increase in the deviation of official numbers from the genuine numbers. The first plot shows that political globalization is positively associated with possible over-reporting, which runs counter to hypothesis 1. Finally, the third plot shows no clear relationship between freedom of the press and a possible over-reporting of GDP statistics, where our hypothesis 3 would predict a negative association. These results are, however, purely correlational and are likely to be confounded by missing variables. This is the case because we expect the different mechanisms to be at work at the same time, meaning that every one of these explanatory variables is an omitted variable. Further, certain other variables will also explain a possible deviation of official GDP statistics from the value predicted by nighttime lights. To counteract possible confounding, all explanatory variables will be included simultaneously and control variables are added in the following regression analysis.

5.2 Regression Analysis

Table 1 presents the results for possible *over-reporting* of official GDP statistics. Our analysis suggests that economic globalization has a negative impact on over-reporting and decentralization has a positive impact on over-reporting, both as hypothesized. Interestingly, political globalization seems to be positively associated with over-reporting, running counter to the hypothesized direction. These results are based on a subset of countries with positive residuals from regression (1). Prior to the analysis, all variables have been standardized for easier interpretation and comparison of the coefficients. Columns (1) and (2) present results for the full set of countries within this subset, whereas columns (3) and (4) include a variable about the independence of the statistical office, which is limited to African countries. We judge our estimates by a 5%-level of statistical significance if not indicated differently.

Column (1) shows results without country fixed effects. Although the coefficient for freedom of press shows a negative association with over-reporting, the coefficient is not statistically significant. We can, thus, not reject the null-hypothesis, which states that there is no effect. The coefficient for political globalization suggests a positive effect on over-reporting. A one standard deviation increase in political globalization is on average associated with a 0.441 standard deviations increase in the deviation of "official" GDP from "real" GDP. Although we can reject the null-hypothesis the result runs counter to our hypothesis 1, suggesting a negative effect of political globalization on data manipulation. In contrast, economic globalization is significantly negatively related to over-reporting, which is in line with hypothesis 2. A one standard deviation increase in economic globalization suggests a decrease in over-reporting of 0.167 standard deviations. We can, thus, reject the null-hypothesis in favor of our hypothesis. Further, decentralization is significantly and positively associated with over-reporting of GDP numbers, rejecting the null-hypothesis in favor of hypothesis 4. A one standard deviation increase in decentralization suggests a 0.084 standard deviations increase in over-reporting. Finally, the level of democracy is negatively and statistically significantly associated with over-reporting, confirming previous research. Column (2) adds year fixed effects. The coefficients only change marginally.

Column (3) introduces independence of the statistical office, which limits the data to African countries. The coefficient of the freedom of press again is not significant but points in the opposite direction than before. Our variable for the independence of the statistical office is neither statistically significant. All other variables are similar as before but change in magnitude. The association for political globalization is now smaller, suggesting than a one standard deviation increase in political globalization increases over-reporting only by 0.27-0.28 standard deviations in African countries. The coefficient for economic globalization is almost equal, whereas the coefficients for decentralization and democracy are slightly bigger. This suggests that in Africa decentralization has a bigger negative effect and democracy a bigger positive effect on publishing proper data than is the case in the rest of the world.

Table 2 shows the results for possible *under-reporting* of GDP data. The dependent variable is converted into absolute values, meaning that a higher value of the dependent variable represents more under-reporting of

Table 1: OLS Regressions for Positive Residuals

	Dependent variable:			
	Residuals			
	(1)	(2)	(3)	(4)
Press Freedom	−0.014	−0.019	0.043	0.037
	(0.041)	(0.042)	(0.093)	(0.099)
Ind. Stat. Office			0.038	0.035
			(0.036)	(0.039)
Pol. Globalization	0.441	0.447	0.269	0.280
	(0.055)	(0.056)	(0.093)	(0.098)
Econ. Globalization	−0.167	−0.167	−0.161	−0.173
	(0.021)	(0.021)	(0.047)	(0.049)
Decentralization	0.084	0.084	0.110	0.105
	(0.027)	(0.026)	(0.048)	(0.049)
Democracy	−0.166	−0.172	−0.195	−0.191
	(0.056)	(0.056)	(0.074)	(0.079)
Year F.E.	No	Yes	No	Yes
Observations	397	397	110	110
R^2	0.606	0.616	0.501	0.506
Adjusted R^2	0.596	0.602	0.445	0.427

Notes: This table reports OLS estimates for countries with an "official" GDP higher than predicted GDP by nighttime lights. The dependent variable is the residuals from regression (1). All variables have been standardized prior to analysis. Columns (1) and (2) use all countries; columns (3) and (4) use African countries. Robust standard errors in parentheses. Constant and control variables are included but not reported.

GDP data, i.e. more possible manipulation. The results are similar as for over-reporting, with the difference that political globalization is now negatively associated with under-reporting, as hypothesized. The table is constructed in the same way as Table 1. Columns (1) and (2) again do not show a significant association between freedom of press and the manipulation of statistics. The coefficient for political globalization points in the opposite direction now and suggests that a one standard deviation increase in political globalization decreases under-reporting of GDP by 0.29

Table 2: OLS Regressions for Negative Residuals

	Dependent variable:			
	Residuals			
	(1)	(2)	(3)	(4)
Press Freedom	0.050	0.054	−0.120	−0.142*
	(0.059)	(0.060)	(0.092)	(0.086)
Ind. Stat. Office			−0.093	−0.113*
			(0.058)	(0.067)
Pol. Globalization	−0.285	−0.289	−0.568	−0.550
	(0.020)	(0.020)	(0.068)	(0.098)
Econ. Globalization	−0.064	−0.069	−0.069	−0.050
	(0.034)	(0.035)	(0.065)	(0.104)
Decentralization	0.143	0.144	0.126	0.145
	(0.036)	(0.037)	(0.099)	(0.097)
Democracy	−0.147	−0.152	0.069	0.090
	(0.042)	(0.045)	(0.090)	(0.092)
Year F.E.	No	Yes	No	Yes
Observations	124	124	51	51
R^2	0.627	0.630	0.781	0.790
Adjusted R^2	0.594	0.583	0.719	0.700

Notes: This table reports OLS estimates for countries with an "official" GDP lower than predicted GDP by nighttime lights. The dependent variable is the residuals from regression (1). All variables have been standardized prior to analysis. Columns (1) and (2) use all countries; columns (3) and (4) use African countries. Robust standard errors in parentheses. Constant and control variables are included but not reported.

standard deviations, which is in line with our hypothesis 1. Similarly, economic globalization is negatively related to under-reporting but only on a 10% significance level. A one standard deviation increase in economic globalization suggests a decrease of 0.06-0.07 standard deviations in GDP under-reporting. This suggested effect is much smaller than for countries over-reporting their GDP numbers. The results suggest a positive effect of decentralization on data manipulation. A one standard deviation increase in decentralization increases under-reporting of GDP by 0.14 standard deviations. A one standard deviation increase in democracy suggests a decrease in under-reporting by 0.15 standard deviations.

Focusing on African countries in columns (3) and (4), only political globalization seems to have an effect on data manipulation. The coefficient suggests that a one standard deviation increase in political globalization decreases under-reporting of GDP by 0.55 standard deviations. For press freedom, independence of the statistical office, economic globalization, decentralization, and democracy there are no significant associations when focusing on African countries.

6 Robustness Checks

Table 3 presents robustness checks of the results using alternative measures of data quality. We use ODIN and ODB for robustness checks. They are, however, not direct measures of data manipulation, but rather indicate overall data quality and accessibility of official data. A higher value in either one of these measures represents higher official data quality, which we interpret as making data manipulation less likely.

Columns (1) and (2) present the results for the ODIN score as a dependent variable and columns (3) and (4) for the ODB as a dependent variable. The straight columns additionally include year fixed effects. The regression coefficients read as follows: A one point increase in the freedom of press index is associated with a 0.245 points decrease in the ODIN score. This result for press freedom is counter to our hypothesis 3 and also not in line with our main results. Political globalization and economic globalization are significantly and positively associated with the ODIN score. This is partially in line with our main results. Decentralization is significantly and positively associated with the ODIN score, which is against our hypothesis 4 and our main results. Democracy is positively associated with the ODIN

Table 3: Robustness Checks

	Dependent variable:			
	ODIN		ODB	
	(1)	(2)	(3)	(4)
Press Freedom	−0.235	−0.235	0.101	0.101
	(0.092)	(0.096)	(0.086)	(0.085)
Pol. Globalization	0.550	0.522	0.714	0.717
	(0.089)	(0.089)	(0.091)	(0.092)
Econ. Globalization	0.271	0.237	0.094	0.089
	(0.053)	(0.050)	(0.049)	(0.049)
Decentralization	0.190	0.163	0.182	0.178
	(0.069)	(0.065)	(0.058)	(0.058)
Democracy	0.390	0.377	0.240	0.243
	(0.109)	(0.107)	(0.099)	(0.098)
Year F.E.	No	Yes	No	Yes
Observations	215	215	201	201
R^2	0.495	0.557	0.594	0.603
Adjusted R^2	0.482	0.544	0.583	0.588

Notes: This table reports OLS estimates for robustness using the ODIN (columns 1 and 2) and the ODB (columns 3 and 4) as a dependent variable. All variables have been standardized prior to analysis. Robust standard errors in parentheses. Constant included but not reported.

score.

The results for the ODB are similar, freedom of press is, however, not significantly associated with the ODB. This is in line with our main results.

7 Discussion and Conclusion

This paper analyzes the manipulation of official data and how different institutional arrangements might alter the extent of this manipulation. While this topic has been mainly taken up by the popular press, academic research has not substantially analyzed the various institutional factors beyond democracy and autocracy that increase or decrease the incentives for data manipulation. This is arguably due to the difficulty of determining the “true” state of an economy. As a result of data manipulation, at least some national policy-makers and international organizations are wrongly informed about the state of the economy. Using manipulated official data can also lead to biased results in econometric analyses.

We hypothesize that a country’s economic and political openness, an independent statistical office, and press freedom curb the falsification of official data. In contrast, decentralization is expected to increase the government’s scope for data manipulation. We test the hypothesized associations by constructing a panel-data set consisting of 195 countries for the years 2013 to 2019. Nighttime lighting data is employed to predict "real" GDP. Subsequently the deviations from predicted to "official" GDP are explained with determinants suggested by our theoretical hypotheses. Nightlight data is a good proxy for measuring "real" economic activity (e.g. Henderson, Storeygard, and Weil 2012), and is very unlikely to be manipulated by governments. If this measure becomes more prominent, however, governments might try to alter lighting in their countries.

Our empirical research suggests that economic globalization tends to reduce data falsification by the government. Economic integration makes it easier to reproduce data and increases information exchange between countries, both making it less likely that official data is manipulated (Kubo 2012; Yang and Shanahan 2003).

For political openness variable the results are mixed. While for countries over-reporting GDP, our results suggest that political openness increases manipulation, for countries under-reporting GDP, our results suggest the opposite. The former result might be explainable because many decisions in international organizations are based on rankings or evaluations (Kelley and Simmons 2021), so that member-countries have incentives to over-report GDP or statistics in order to improve reputation and lower credit risk. The latter result is in line with our hypothesis. A reason might be that a country politically more exposed to other nations finds it more difficult to under-report its GDP because the other nations may fear they have to support the country. The stronger the political integration with other countries, the more critical these countries are when a lower GDP than plausible is officially reported.

As hypothesized, decentralization increases the government's scope to present manipulated official statistics. Decentralization increases the scope for local governments to manipulate data (e.g. Koen and Van den Noord 2005).

Surprisingly, we do not find any significant association between press freedom and the extent of falsification of official data. This empirical finding goes against the hypothesized notion that a freer press allows for an easier detection and publication of governments' efforts to manipulate data. Four plausible reasons can explain this result. First, the press simply has no effect in restraining a government's effort to manipulate data. To contest the official statistics, the press would need plausible/sensible alternative data, which it might be unable to obtain. Second, the data on press freedom may not be suited to capture the reporting on data manipulation, but be better suited for more narrow political issues. Third, the estimation approach used does not well capture the relationship between the activity of the press and how official data is produced. Fourth, the hypothesis insufficiently takes into account a possibly dwindling importance of the press relative to social media. An increasing number of young people rely much more on discussions in blogs and other digital channels than on newspapers (Geers 2020). In contexts characterized by limited freedom of press, the consumption of online news, including social media, is an important factor in enabling political participation (Karakaya and Glazier 2019).

Our empirical research also does not find a significant relationship between the independence of the statistical office and data manipulation. This result relates only to African countries, as this variable is only available for African countries. While in Africa other determinants might be more important in explaining data manipulation, the independence of the statistical office may well be able to reduce governments' incentives to manipulate data in a non-African context. To our knowledge, data on the independence of the statistical office for non-African countries is not available.

Our study is subject to various limitations and should provide a first inquiry into new hypotheses in the study of government data manipulation. The OLS estimations are not able to capture causal effects. The most common caveats to establishing causality are briefly discussed in the following. Measurement error in the explanatory variables should not be a major problem in our analysis. Most of the variables are indices and comprise different sort of information, making a systematic bias less likely. Reverse or simultaneous causality might be a problem in our empirical design. If official data has been manipulated, a government has an incentive to further curtail press freedom. Rampant data manipulation could also lead to other countries cutting ties with the manipulating country or the country being excluded from international organizations. This would decrease political globalization and to a lesser extent possibly also economic globalization of a country. For the variables of the independence of the statistical office and decentralization, it seems unlikely that data manipulation has a direct short-term impact on them. Finally, omitted variables might bias our estimations.

The empirical estimates for the control variable of democracy, confirm that more democratic regimes tend to manipulate GDP to a lesser extent, as found by (Martinez 2022) or (Magee and Doces 2015). This serves as a sanity check for our main results. Our results that decentralization increases manipulation are in line with (Koen and Van den Noord 2005). However, our estimates do not support previous research by (World Bank 2021), which suggests that an independent statistical office reduces data manipulation. Our further results extend the existing literature.

Falsification of statistics is not limited to governments. Also in the private sector, the manipulation of data is prevalent – for instance in the form of falsified credit ratings (e.g. Nguyen et al. 2022) or online reviews (e.g. Mayzlin, Dover, and Chevalier 2014; Zheng et al. 2021). Forthcoming studies should also include institutions altering the incentives for data manipulation in the private sector.

Future research should try to establish a causal relationship between different types of institutions and data manipulation. This relation may also behave differently for different sets of countries and economic systems. With the emergence of more precise estimating procedures of economic activity, for example by using daytime instead of nighttime satellite imagery (see e.g. Lehnert et al. 2022), the detection of manipulated GDP should become easier and more accurate. Another challenging aspect is to inquire about the role of social media in constraining data manipulation and if it indeed replaces the press in its role to discipline the government.

Bibliography

Adam, Antonis, and Sofia Tsarsitalidou. 2022. "Data misreporting during the COVID19 crisis: The role of political institutions." *Economics Letters* 213:110348.

Adiguzel, Fatih Serkant, Asli Cansunar, and Gozde Corekcioglu. 2020. "Truth or Dare? Detecting Systematic Manipulation of COVID-19 Statistics." *Journal of Political Institutions and Political Economy* 1 (4): 543–557.

Alt, James, David Dreyer Lassen, and Joachim Wehner. 2014. "It isn't just about Greece: domestic politics, transparency and fiscal gimmickry in Europe." *British Journal of Political Science* 44 (4): 707–716.

Aragão, Roberto, and Lukas Linsi. 2022. "Many shades of wrong: what governments do when they manipulate statistics." *Review of International Political Economy* 29 (1): 88–113.

Balashov, Vadim S., Yuxing Yan, and Xiaodi Zhu. 2021. "Using the Newcomb–Benford law to study the association between a country's COVID-19 reporting accuracy and its development." *Scientific Reports* 11 (1): 22914.

Bhandari, Laveesh, and Koel Roychowdhury. 2011. "Night lights and economic activity in India: A study using DMSP-OLS night time images." *Proceedings of the Asia-Pacific advanced network* 32:218–236.

Bond, Kanisha D, Courtenay R Conrad, Dylan Moses, and Joel W Simmons. 2022. "Detecting anomalies in data on government violence." *Political Science Research and Methods* 10 (3): 634–641.

Buchanan, James M, and Gordon Tullock. 1965. *The calculus of consent: Logical foundations of constitutional democracy.* Vol. 100. Michigan: University of Michigan Press.

Chan, Ho Fai, Bruno S. Frey, Ahmed Skali, and Benno Torgler. 2019. *Political Entrenchment and GDP Misreporting.* CESifo Working Paper Series 7653. CESifo.

Chen, Shuo, Xue Qiao, and Zhitao Zhu. 2021. "Chasing or cheating? Theory and evidence on China's GDP manipulation." *Journal of Economic Behavior & Organization* 189:657–671.

Chen, Xi, and William D Nordhaus. 2011. "Using luminosity data as a proxy for economic statistics." *Proceedings of the National Academy of Sciences* 108 (21): 8589–8594.

Coremberg, Ariel. 2014. "Measuring Argentina's GDP Growth." *World Economics* 15 (1): 1–32.

Dafflon, Bernard, and Sergio Rossi. 1999. "Public accounting fudges towards EMU: A first empirical survey and some public choice considerations." *Public Choice* 101 (1): 59–84.

Deaton, Angus. 2013. *The Great Escape: Health, Wealth, and the Origins of Inequality.* Princeton University Press.

Drakos, Konstantinos, and Andreas Gofas. 2006. "The devil you know but are afraid to face: Underreporting bias and its distorting effects on the study of terrorism." *Journal of Conflict Resolution* 50 (5): 714–735.

Ebener, Steeve, Christopher Murray, Ajay Tandon, and Christopher C Elvidge. 2005. "From wealth to health: Modelling the distribution of income per capita at the sub-national level using night-time light imagery." *International Journal of Health Geographics* 4 (1): 1–17.

Elvidge, Christopher D., Mikhail Zhizhin, Tilottama Ghosh, Feng-Chi Hsu, and Jay Taneja. 2021. "Annual Time Series of Global VIIRS Nighttime Lights Derived from Monthly Averages: 2012 to 2019." *Remote Sensing* 13 (5). Https://eogdata.mines.edu/products/vnl/.

European Commission. 2009. *Report on Greek government deficit and debt statistics.* Office for Official Publications of the European Communities.

Feldstein, Martin. 2017. "Underestimating the real growth of GDP, personal income, and productivity." *Journal of Economic Perspectives* 31 (2): 145–164.

Geers, Sabine. 2020. "News consumption across media platforms and content: A typology of young news users." *Public Opinion Quarterly* 84 (S1): 332–354.

Georgiou, Andreas V. 2021. "The manipulation of official statistics as corruption and ways of understanding it." *Statistical Journal of the IAOS* 37 (1): 85–105.

Ghosh, Tilottama, Sharolyn Anderson, Rebecca L Powell, Paul C Sutton, and Christopher D Elvidge. 2009. "Estimation of Mexico's informal economy and remittances using nighttime imagery." *Remote Sensing* 1 (3): 418–444.

Ghosh, Tilottama, Rebecca L. Powell, Christopher D. Elvidge, Kimberly E. Baugh, Paul C. Sutton, and Sharolyn Anderson. 2010. "Shedding Light on the Global Distribution of Economic Activity." *The Open Geography Journal* 3 (1): 147–160.

Henderson, J. Vernon, Adam Storeygard, and David N Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994–1028.

Hodler, Roland, and Paul A. Raschky. 2014. "Regional Favoritism." *The Quarterly Journal of Economics* 129 (2): 995–1033.

Jean, Neal, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. 2016. "Combining satellite imagery and machine learning to predict poverty." *Science* 353 (6301): 790–794.

Kalgin, Alexander. 2016. "Implementation of Performance Management in Regional Government in Russia: Evidence of Data Manipulation." *Public Management Review* 18 (1): 110–138.

Karakaya, Suveyda, and Rebecca A Glazier. 2019. "Media, information, and political participation: The importance of online news sources in the absence of a free press." *Journal of Information Technology & Politics* 16 (3): 290–306.

Kelley, Judith G, and Beth A Simmons. 2021. "Governance by other means: Rankings as regulatory systems." *International Theory* 13 (1): 169–178.

King, Gary, Jennifer Pan, and Margaret E Roberts. 2017. "How the Chinese government fabricates social media posts for strategic distraction, not engaged argument." *American Political Science Review* 111 (3): 484–501.

Koen, Vincent, and Paul Van den Noord. 2005. *Fiscal gimmickry in Europe: One-off measures and creative accounting.* OECD Economics Department Working Papers 417. Paris: OECD Publishing.

KOF Swiss Economic Institute. 2021. "KOF Globalisation Index," https://kof.ethz.ch/en/forecasts-and-indicators/indicators/kof-globalisation-index.html (accessed May 10, 2022).

Kubo, Koji. 2012. "Trade Policies and Trade Misreporting in Myanmar." *ASEAN Economic Bulletin* 29 (2): 146–159.

Lehnert, Patrick, Michael Niederberger, Uschi Backes-Gellner, and Eric Bettinger. 2022. *Proxying Economic Activity with Daytime Satellite Imagery: Filling Data Gaps across Time and Space.* IZA Discussion Paper Series 15555. IZA Institute of Labor Economics.

Lorentzen, Peter. 2014. "China's strategic censorship." *American Journal of political science* 58 (2): 402–414.

Magee, Christopher S. P., and John A. Doces. 2015. "Reconsidering Regime Type and Growth: Lies, Dictatorships, and Statistics." *International Studies Quarterly* 59 (2): 223–237.

Martinez, Luis R. 2022. "How Much Should We Trust the Dictator's GDP Growth Estimates?" *Journal of Political Economy* 130 (10).

Mayzlin, Dina, Yaniv Dover, and Judith Chevalier. 2014. "Promotional reviews: An empirical investigation of online review manipulation." *American Economic Review* 104 (8): 2421–2455.

Milesi-Ferretti, Gian Maria. 2004. "Good, bad or ugly? On the effects of fiscal rules with creative accounting." *Journal of Public Economics* 88 (1-2): 377–394.

Mo Ibrahim Foundation. 2020. "Ibrahim Index of African Governance (IIAG)," https://iiag.online/downloads.html (accessed May 15, 2022).

Nathan, Andrew J. 2020. "The puzzle of authoritarian legitimacy." *Journal of Democracy* 31 (1): 158–168.

Natural Earth. 2009-2022. "Medium Scale (1:50m) Cultural Vectors – Countries," https://www.naturalearthdata.com/downloads/50m-cultural-vectors/ (accessed May 20, 2022).

Nguyen, Quan MP, Hung Xuan Do, Alexander Molchanov, Lily Nguyen, and Nhut H Nguyen. 2022. "Politically Motivated Credit Ratings." *Available at SSRN 4188482.*

Open Data Watch. 2015-2020. "Open Data Inventory," https://odin.opendatawatch.com/Report/rankings(accessed May 15, 2022).

Rauch, Bernhard, Max Göttsche, Stefan Engel, and Gernot Brähler. 2011. "Fact and fiction in EU-governmental economic data." *German Economic Review* 12 (3): 243–255.

Reporters Without Borders. 2002-2022. "World Press Freedom Index," https://rsf.org/en/index (accessed May 15, 2022).

Sandefur, Justin, and Amanda Glassman. 2015. "The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics." *The Journal of Development Studies* 51 (2): 116–132.

Sutton, Paul C, and Robert Costanza. 2002. "Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation." *Ecological Economics* 41 (3): 509–527.

Sutton, Paul C, Christopher D Elvidge, Tilottama Ghosh, et al. 2007. "Estimation of gross domestic product at sub-national scales using nighttime satellite imagery." *International Journal of Ecological Economics and Statistics* 8 (S07): 5–21.

The Economist. 2014. "Official Statistics. Don't lie to me, Argentina." June 20. https://www.economist.com/leaders/2014/06/20/dont-lie-to-me-argentina.

———. 2015. "The Chinese economy. Whether to believe China's GDP figures." July 15. https://www.economist.com/free-exchange/2015/07/15/whether-to-believe-chinas-gdp-figures.

———. 2017. "Welcome back. Argentina's new, honest inflation statistics." May 25. https://www.economist.com/the-americas/2017/05/25/argentinas-new-honest-inflation-statistics.

———. 2020. "Can China's reported growth be trusted?" October 15. https://www.economist.com/finance-and-economics/2020/10/15/can-chinas-reported-growth-be-trusted.

———. 2021a. "How data detectives spotted fake numbers in a widely cited paper." August 28. https://www.economist.com/graphic-detail/2021/08/28/how-data-detectives-spotted-fake-numbers-in-a-widely-cited-paper.

———. 2021b. "Why the head of the IMF should resign." September 25. https://www.economist.com/leaders/2021/09/25/why-the-head-of-the-imf-should-resign, 14.

———. 2022a. "A study of lights at night suggests dictators lie about economic growth." September 29. https://www.economist.com/graphic-detail/2022/09/29/a-study-of-lights-at-night-suggests-dictators-lie-about-economic-growth, 21–22.

———. 2022b. "Even China's official economic figures look bleak." May 19. https://www.economist.com/finance-and-economics/2022/05/19/even-chinas-official-economic-figures-look-bleak.

The Economist. 2022c. "India's once-vaunted statistical infrastructure is crumbling." May 19. https://www.economist.com/finance-and-economics/2022/05/19/indias-once-vaunted-statistical-infrastructure-is-crumbling.

———. 2022d. "Turkey grapples with triple-digit inflation." July 14. https://www.economist.com/europe/2022/07/14/turkey-grapples-with-triple-digit-inflation, 21–22.

V-Dem Institute. 2022. "Country-Year: V-Dem Core Dataset," https://www.v-dem.net/vdemds.html (accessed May 15, 2022).

Von Hagen, Jürgen, and Guntram B Wolff. 2006. "What do deficits tell us about debt? Empirical evidence on creative accounting with fiscal rules in the EU." *Journal of Banking & Finance* 30 (12): 3259–3279.

Wallace, Jeremy L. 2016. "Juking the stats? Authoritarian information problems in China." *British Journal of Political Science* 46 (1): 11–29.

Wang, Shaoda, and David Y Yang. 2021. *Policy Experimentation in China: The Political Economy of Policy Learning.* NBER Working Paper Series 29402. National Bureau of Economic Research.

Weidmann, Nils B, and Sebastian Schutte. 2017. "Using night light emissions for the prediction of local wealth." *Journal of Peace Research* 54 (2): 125–140.

World Bank. 2021. *World Development Report 2021: data for better lives.* World development report. Washington: World Bank.

———. 2022. "GDP (current US$)," https://data.worldbank.org/indicator/NY.GDP.MKTP.CD (accessed May 10, 2022).

World Bank. 1990-2018. “Informal Economy Database,” https://www.worldbank.org/en/research/brief/informal-economy-database (accessed May 10, 2022).

———. 1990-2021. “World Development Indicators,” https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.MKTP.CD,NV.AGR.TOTL.ZS,NV.IND.TOTL.ZS,NV.IND.MANF.ZS,NV.SRV.TETC.ZS,NV.SRV.TOTL.ZS# (accessed May 10, 2022).

World Wide Web Foundation. 2013-2017. “Open Data Barometer Global Report (All Editions),” https://opendatabarometer.org/4thedition/data/(accessed May 15, 2022).

Yang, Fang, and James Shanahan. 2003. “Economic openness and media penetration.” *Communication Research* 30 (5): 557–573.

Zheng, Tianxiang, Feiran Wu, Rob Law, Qihang Qiu, and Rong Wu. 2021. “Identifying unreliable online hospitality reviews with biased user-given ratings: A deep learning forecasting approach.” *International Journal of Hospitality Management* 92:102658.

A Appendix

Table A1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
GDP (in bln)	1,312	420.618	1,730.881	0.087	21,433.225
Nighttime Lights Mean	1,357	1.706	6.378	0.00001	64.399
Freedom of Press	1,033	65.522	16.411	11.130	93.540
Decentralization	952	0.249	1.441	−2.506	3.108
Ind. Stat. Office	378	33.003	26.770	0.000	100.000
Pol. Globalization	1,294	72.179	18.893	14.972	99.693
Econ. Globalization	1,259	56.126	18.403	11.771	99.203
Democracy	1,203	0.529	0.255	0.016	0.923
Informal Economy	888	28.962	11.301	8.000	64.600
Share Agriculture	1,262	10.937	10.614	0.014	60.611
Share Manufacturing	1,207	11.852	6.381	0.000	38.733
Share Industry	1,267	25.886	11.283	4.556	73.099
Share Services	1,263	55.263	11.466	16.775	94.256