European Journal of Political Economy 82 (2024) 102523

Contents lists available at ScienceDirect

European Journal of Political Economy

journal homepage: www.elsevier.com/locate/ejpe

Full Length Article

Governments manipulate official Statistics: Institutions matter

Andre Briviba [a,b], Bruno Frey [b,c,*], Louis Moser [b], Sandro Bieri [b]

[a] University of Fribourg, Bd. de Pérolles 90, 1700, Fribourg, Switzerland
[b] CREMA, Center for Research in Economics, Management and the Arts, Südstrasse 11, Zürich, Switzerland
[c] University of Basel, Peter-Merian-Weg 6, Basel, Switzerland

ARTICLE INFO

JEL classification:
E01
E02
H11
P44

Keywords:
Data manipulation
Official statistics
Political economy
Government
Institutions

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 night-time lighting data – and „official“ GDP on these institutional factors suggests that economic openness and democracy decrease 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.

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 in the economy if the data upon which decisions are taken is incorrect. 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.

* Corresponding author. University of Basel, Peter-Merian-Weg 6, Basel, Switzerland.
E-mail address: bruno.frey@bsfrey.ch (B. Frey).

[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).

https://doi.org/10.1016/j.ejpoleco.2024.102523
Received 12 June 2023; Received in revised form 8 February 2024; Accepted 1 March 2024
Available online 4 April 2024
0176-2680/

International organizations may, likewise, undertake mistaken policies if a country reports manipulated data. For instance, they may hand out 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 night-time lighting captured by satellites to predict the economic output. We explain the deviation of "official" GDP from "true" GDP, predicted by night-time lights, with variables based on our hypotheses. For this purpose, we construct a panel-data set consisting of 195 countries for the years 2013–2019.

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. A similar heterogeneity applies to the alternative measure of international political openness, passport power. Its explanatory role is stronger and statistically significant in the overreporting samples, while it does not play any role in the underreporting samples. Finally, we do find that neither freedom of the press, independence of the statistical office, nor political rights 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. In 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 refers 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. However, as shown by Cronin and McInerney (2023), institutional arrangements such as specific fiscal rules also play a decisive role in the accuracy of fiscal forecasts.

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 et al. 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 et al. 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 et al. 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 et al. 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 et al. 2020; Adam and Tsarsitalidou 2022) or countries with lower economic development (Balashov et al. 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). Our approach is related to Martinez (2022), who exploits one certain threshold for low-income countries (IDA threshold) and derives extensive insights on the GDP composition affected. We contribute to this literature by focusing on additional institutional arrangements and their effects in particular. Moreover, we refer to our strategy of splitting the sample into over- and underreporting countries to investigate more general relationships and emphasize the determinants leading to manipulation.

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 their legitimacy partly 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.

Data revisions are one of the mechanisms to counter manipulated or incorrectly recorded official statistics. Employed for instance, by the WTO, those revisions can result in substantial changes. Kerner et al. (2017) provide evidence for low-income countries, concluding that in contrast to initial GNI statistics, ex-post revised statistics reflect less data management practices. However, these corrections not only apply to GNI statistics but also affect rainfall and conflict data. In this case, the revisions changed the postulated relationship of rainfall shocks causing civil conflict (Liang and Sim 2019).

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 et al. 2021). Two implications can be derived: first, accountability increases the reliability of public data; second, systematic data manipulation occurs when political leaders benefit from false data.

Democratic regimes per se are not exempt from manipulating economic data. Cases such as Argentina and Greece have often been suspected of falsifying official statistics. Coremberg (2014) reproduces Argentina's GDP data and finds that the reported GDP data is too high, putting into doubt that the country has the highest growth rates in South America. Greece, especially 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 and Maria, 2004; Von Hagen and Wolff, 2006; Koen and Van den Noord, 2005).

Electoral cycles are another cause identified to pressure governments to manipulate official data. Martinez (2022) shows that the deviation from GDP estimated by nightlight data for authoritarian regimes is higher just before elections. A similar effect was also found for democracies (Alt et al. 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 et al. 2014; Chan et al., 2019).

2.3. Hypotheses

Previous literature has shown that pressure from the international or domestic sphere can influence the incentives to manipulate official data. While most scholarly work focuses on factors making data manipulation more likely, institutional and economic factors can also make data manipulation less prevalent. Government actors having an interest in falsifying official data are constrained by various institutional and constitutional arrangements (Buchanan and Gordon, 1965). Democracy, for example, can be understood as a constitutional arrangement ensuring that the government acts in the interest of the population. Government actors are constrained in their desire to manipulate data in their own interest 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 better able to evaluate the current state of its national economy and is less likely to be misled by false statistics. These considerations lead to 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 with competencies to collect data could constrain the possibilities of central governments to manipulate data. For example, if government data represents an aggregate of subnational data, it is more difficult to manipulate. However, the same argument applies vice versa: decentralization increases the scope for subnational entities to manipulate data or correlates with missing uniform standards, resulting in aggregation errors. Either way, decentralization increases the number of actors with their own interests and therefore, the potential of manipulation in general.

Most of the empirical evidence points in this direction. A 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 statistical office’s independence and the level of data manipulation. The main factors determining the extent of independence relate to the appointment of the top positions, fiscal autonomy, and regulations imposed by the government.

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 spanning from 2013 to 2019. We limit ourselves to this period because data coverage is heavily 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 Night-time 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 et al., 2007; Chen and Nordhaus 2011; Henderson et al. 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 night-time 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 2013 to 2016, and ODIN data for 2015 to 2020, which limits the robustness checks to these years, respectively. ODB and ODIN both measure the transparency, availability, and quality of official data, all factors that 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 and not necessarily of reliable quality. 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 (Gygli et al., 2019). The variable comprises the percentage of trade in goods and services of a country's GDP and trade partner diversity.

Political Openness: Political openness is measured using de jure political globalization from the KOF Globalization Index (Gygli et al., 2019). This measure comprises a country's membership in international organizations and treaties and the treaty partner diversity. Further we introduce passport power as an alternative measure of international political openness. It reflects the number of countries citizens can travel to visa-free. To proxy internal political openness, we exploit the political rights score by the Freedom House. This index consists of the electoral process, political pluralism and participation, and the functioning of government.

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 (no freedom of press) and 100 (full freedom of press) for 180 countries around the world.

Decentralization: The Regional Offices Relative Power variable from the Coppedge et al (2024) is used as a measure of decentralization.

3.3. Control variables

Democracy: The Electoral Democracy Index from V-Dem is used to measure democracy (Coppedge et al., 2024). This index captures the extent to which the ideal of electoral democracy is, in its fullest sense, achieved on a scale from 0 (no democracy) to 1 (full democracy).

Contribution to GDP by Sector: The contributions of the different economic sectors to GDP are controlled for, as different sectors are likely to emit different amounts of light (e.g., Bhandari and Roychowdhury 2011). A sector contributing greatly to GDP but emitting little light results in a difference in economic activity measured in night-time 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: An extensive informal economy can also result in a deviation between economic activity measured by night-time lights and official GDP. This difference depends on the informal economy's light emissions, is generally not included in the official GDP measure, but can constitute a significant share of GDP (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 night-time lighting. Second, we differentiate between positive and negative residuals and explain these deviations with the variables suggested by our hypotheses. In the first step, official GDP is regressed on VIIRS night-time lights via OLS[2]:

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

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 night-time lights and $y_{c,t}$ is the error term, i.e., the variance of GDP not explained by night-time 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. Fig. 1 shows the results of the first regression.

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

[2] In addition to the original first stage regression ($gdp_{c,t} = \alpha + vnl_{c,t} + y_{c,t}$), we also include the measure of informal economy into the first regression to minimize mostly unidirectional influences of underreporting ($gdp_{c,t} = \alpha + vnl_{c,t} + informaleconomy_{c,t} + y_{c,t}$). This generates new groups of positive and negative residuals without the asymmetric influence of the informal economy. The majority of results stay robust to this alternation. Notably, decentralization and press freedom deviate the most. In particular, decentralization now also shows heterogenous effect for over- and underreporting samples, while freedom of press becomes congruent and more informative. The results of these extended regressions are placed in the Appendix (Table A3).

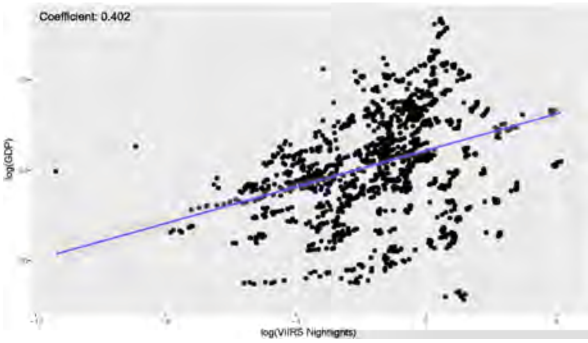

Fig. 1. Regressing GDP on nightlights.

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

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

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. **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 **K** are included in the main regression. They capture other factors that may explain why a country's official GDP differs from the GDP measured by night light. The vector **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 night-time 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.[3] We present results for subsets of countries with either possible over-reporting or under-reporting, respectively. Fig. 2 shows the spatial distribution of the residuals.

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

By using night-time 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, night-time 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.

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 night-time 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 Fig. 3 show the results. The independence of the statistical office is excluded from this analysis, as data is only available for African countries.

This figure shows correlation plots for the results of regressing the positive residuals on each explanatory variable using a subset of the data for 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

[3] A critical aspect of our study's robustness is whether countries switch between these over and under-reporting categories over time. Such fluctuations could question the validity of our approach and results. Within the seven-year span of our study, there are 18 countries that switch between the categories, which would, at first glance, be indeed a concern. However, if we exclude the smallest residuals around 0 (-0.1 to 0.1 with a range from -3.5 to 3) only one unique switch is reported (Estonia). As such, our variation of this phenomenon constitutes, at most, a neglectable issue. This holds as well for the additional outcome measure of ln(sum/area).

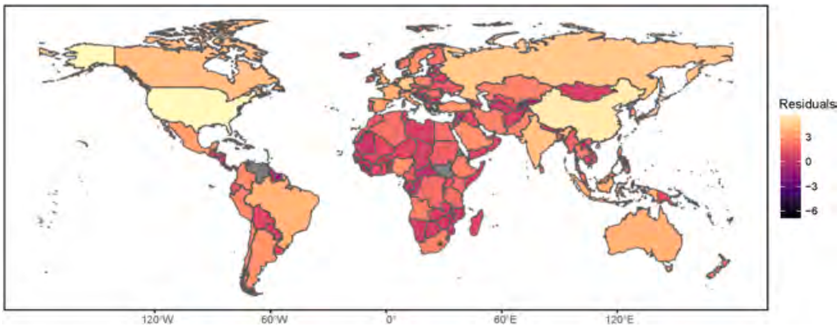

Fig. 2. Spatial distribution of the dependent variable.

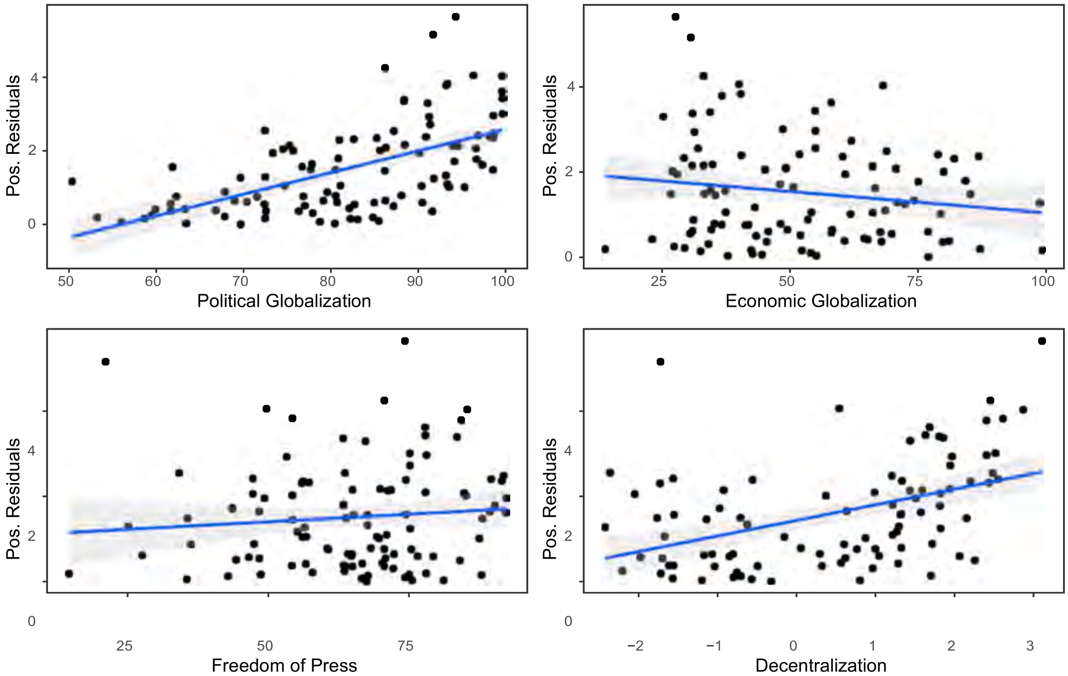

Fig. 3. Correlation plots for over-reporting of GDP

likely to be confounded by missing variables. To mitigate the undeniable influence of other mechanisms and omitted variables, which also might affect the deviation of official GDP statistics and predicted night-time light values, we include all explanatory variables simultaneously. In addition, control variables are accounted for 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 were standardized for a more straightforward 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 95%-level of statistical significance if not indicated differently.

Column (1) shows results without year-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. On average, a one standard deviation increase in political globalization is associated with a 0.441 standard deviation 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

Table 1
OLS regressions for positive residuals (overreporting).

	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
R2	0.606	0.616	0.501	0.506
Adjusted R2	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 night-time 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.

globalization on data manipulation. In contrast, economic globalization is significantly negatively related to over-reporting, which aligns 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 (Martinez 2022). 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. This also applies to the variable independence of the statistical office. All other variables are similar to before but change in magnitude. The association for political globalization is now smaller, suggesting that 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.

Table 2
OLS regressions for negative residuals (underreporting).

	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
R2	0.627	0.630	0.781	0.790
Adjusted R2	0.594	0.583	0.719	0.700

Notes: *Notes*: This table reports OLS estimates for countries with an "official" GDP lower than predicted GDP by night-time 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 are in parentheses. Constant and control variables are included in all calculations.

This suggests that in Africa, decentralization has a larger negative effect, and democracy has a bigger positive effect on publishing proper data than is the case in the rest of the world.

Controls: Including the control variables increases the explained variance substantially, while the significance of the main variables is not affected. Informal economy has a statistically significant but small divergence-reducing effect. Share of agriculture, industry, and services have a similar magnitude, but all point in the other direction.

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 GDP data, i.e., more possible manipulation. The results are similar to 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. It suggests that a one standard deviation increase in political globalization decreases under-reporting of GDP by 0.29 standard deviations, which is in line with our hypothesis 1. A possible interpretation of this result may be that stronger political globalization reduces the possibilities to underreport their GDP. In contrast, an isolated country does not face such pressure to appear poorer than they are. This outcome may be attributed to possible financial obligations of other countries when the country in question appears to be poorer than it really is. We acknowledge that this is just one possible explanation among many others and requires further investigation.

Similarly, economic globalization is negatively related to under-reporting but only on a 90% 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 indicate 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.

Controls: Including the control variables increases the explained variance substantially. Share of service indicates a small divergence increasing effect.

Focusing on African countries in columns (3) and (4), only political globalization seems to affect 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

As a robustness check for our outcome of interest, we alternate the measure of night light intensity. Since the mean value of night light intensity by country has some drawbacks, such as missing geographical influence, we extend our analysis to the natural logarithm of the sum of night light intensity by the country's area. Thus representing the proportionality of the night light intensity relative to its

Table 3
OLS regressions for residuals – sum/area.

	Dependent variable:			
	Residuals			
	Pos. res	Pos. res	Neg. res	Neg. res
	(1)	(2)	(3)	(4)
Press Freedom	−0.073	0.076	0.345	−0.010
	(0.090)	(0.172)	(0.110)	(0.144)
Ind. Stat. Office		−0.101		−0.434
		(0.124)		(0.111)
Pol. Globalization	1.330	0.934	−0.776	−1.530
	(0.140)	(0.207)	(0.054)	(0.150)
Econ. Globalization	−0.297	−0.487	−0.106	0.023
	(0.048)	(0.128)	(0.065)	(0.115)
Decentralization	0.396	0.578	0.194	0.398
	(0.065)	(0.146)	(0.073)	(0.132)
Democracy	−0.460	−0.332	−0.280	0.142
	(0.126)	(0.147)	(0.093)	(0.122)
Year F.E.	Yes	Yes	Yes	Yes
Observations	345	61	176	100
R2	0.676	0.825	0.545	0.692
Adjusted R2	0.662	0.766	0.505	0.637

Notes: This table reports OLS estimates for countries with an "official" GDP higher than predicted GDP by the sum of night-time lights over area in columns (1) and (2). Columns (3) and (4) present the results for underreporting countries. All variables have been standardized prior to analysis. Columns (1) and (3) use all countries; columns (2) and (4) use African countries. Robust standard errors in parentheses. Constant and control variables are included in all calculations.

area. A compelling indicator for our findings would be the consistency of the coefficients' direction with previously computed mean night light values.

The results are presented in Table 3. The columns (1) and (2) show the results for the overreporting sample, while regressions (3) and (4) show the underreporting sample results. In comparison to Table 1, the coefficients' signs algin in all cases. In the underreporting main sample, press freedom now becomes statistically significant, whereas all other variables align with the results of Table 2. For the subsample of African countries, only decentralization increases in magnitude.

Overall, the results hold within over- and under-reporting countries for the alternative outcome specification. Across the main subsample, we again observe some heterogeneity in effects. Political openness increases the divergence of GDP and nightlight in the overreporting countries, while it decreases it in the underreporting countries. Decentralization, across all specifications, indicates consistently more overreporting. Contrary, democracy suggests less manipulation in either direction. Economic openness is also consistently a reducing driver but doesn't hold its statistical significance in the subsample of African countries.

Controls: Informal economy has a statistically significant but small divergence-reducing effect. The four sectors have small and heterogeneous effects for positive and negative residuals.

To improve our measure of political openness, we differentiate between international and internal political openness. While international treaties and memberships in international organizations probably influence the adaptation of official GDP statistics, the internal political openness of a country could also play a crucial role. The more open a political system is for citizens to participate in, the less divergence of night light and GDP is to be expected.

We employ the political rights score by the Freedom House, which is calculated based on the electoral process, political pluralism and participation, and the functioning of government. As a further proxy for international political openness, we exploit passport power. This variable reflects the number of countries citizens can travel to visa-free.

In Table 4, the variables passport power and political rights are included in all specifications. Here, we restrict the results to the main sample without the independent statistical office to enable a direct comparison of the two outcome measures. For mean night light intensity, even though passport power has a substantial magnitude, the other coefficients remain statistically significant and of similar magnitude. In the case of summed night light over area, also no variable changes its sign.

For the pooled sample of all residuals, including the political rights score does not affect the coefficients at all (see Appendix A2). The largest change is found for the democracy variable, which decreases in its third decimal. We conclude that while internal political rights seem a plausible determinant for the divergence of GDP and night light intensity, we cannot find a relevant statistical link. For passport power, we find heterogeneous effects. In the overreporting samples, its explanatory role is stronger and statistically significant, while it does not play any role in the underreporting samples. This is consistent with the results of political openness and points in the direction of heterogeneous effects for over- and underreporting countries.

Table 5 presents robustness checks of the results using alternative data quality measures. We employ ODIN and ODB for robustness checks. They are, however, no direct measures of data manipulation, but rather indicate overall data quality and accessibility of official

Table 4
OLS Regressions for Residuals – Passport Power & Political rights.

	Dependent variable:			
	Residuals			
	Mean		Sum/area	
	Pos. res	Neg. res	Pos. res	Neg. res
	(1)	(2)	(3)	(4)
Press Freedom	−0.046	0.068	−0.209	0.358
	(0.043)	(0.062)	(0.094)	(0.111)
Pol. Globalization	0.400	−0.292	1.180	−0.790
	(0.057)	(0.020)	(0.141)	(0.056)
Econ. Globalization	−0.189	−0.073	−0.336	−0.119
	(0.022)	(0.035)	(0.047)	(0.064)
Decentralization	0.068	0.126	0.290	0.185
	(0.026)	(0.039)	(0.068)	(0.076)
Democracy	−0.154	−0.175	−0.562	−0.140
	(0.072)	(0.081)	(0.179)	(0.137)
Passport Power	0.155	0.050	0.459	0.063
	(0.037)	(0.048)	(0.086)	(0.118)
Political Rights	−0.051	0.010	0.066	−0.168
	(0.077)	(0.076)	(0.174)	(0.127)
Year F.E.	Yes	Yes	Yes	Yes
Observations	397	124	345	176
R2	0.634	0.633	0.703	0.550
Adjusted R2	0.619	0.578	0.688	0.504

Notes: This table reports OLS estimates for countries with an "official" GDP higher than predicted GDP by night-time 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 in all calculations.

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 contradicts our hypothesis 3 and is 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 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

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 night-time lighting captured by satellites to predict the economic output. We explain the deviation of "official" GDP from "true" GDP, predicted by night-time lights, with variables based on our hypotheses. The analysis is bolstered by an alternative outcome measure controlling for geography and including further measures to differentiate political openness. Our paper seeks to contribute to a better theoretical and empirical understanding of governments' data falsification.

The emphasis lies on the comparison across the over-and underreporting samples. We observe that the majority of countries fall into the overreporting category and generally have a higher explained variance. With the larger explanatory relevance of coefficients, it implies more evidence in the overreporting countries. Considering the alternative outcome measure, we find great consistency within and across over- and underreporting countries. Noteworthy is the press freedom's change in magnitude and significance, which requires a more fine-grained analysis in order to provide a satisfactory interpretation.

Following, we contextualize the overall findings for each variable. They are discussed in order of the magnitude of their effect size and therefore, their explanatory relevance.

For the *political openness* variable, we find heterogeneous effects. 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 financially 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. A similar heterogeneity occurs for the alternative measure of international political openness, *passport power*. Its explanatory role is stronger and statistically significant in the overreporting samples, while it does not play any role in the underreporting samples. This is consistent with the results of political openness and points in the direction of heterogeneous effects for over- and underreporting countries. For our measure of internal political openness, *political rights*, the results are inconclusive, of low magnitude, and no

Table 5
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.	Yes	Yes	Yes	Yes
Observations	215	215	201	201
R2	0.495	0.557	0.594	0.603
Adjusted R2	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.

statistical significance.

Economic globalization tends to consistently 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).

The empirical estimates for the control variable of *democracy* confirm that more democratic regimes tend to manipulate GDP to a lesser extent, as established by Martinez (2022) or Magee and Doces (2015). This serves as a sanity check for our main results.

As hypothesized, *decentralization* increases the government's scope to present manipulated official statistics. Decentralization increases the scope for subnational entities to manipulate data. This outcome may also be attributable to missing standards, resulting in aggregation errors. Our results that decentralization increases manipulation are in line with Koen and Van den Noord (2005). However, it remains to be explored which specific decentralized competencies drive these results and whether, for instance, accountability can mitigate it.

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 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 applied does not sufficiently capture the relationship between the activity of the press and how official data is produced. Fourth, the hypothesis insufficiently considers 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 (global) social media, is an important factor in enabling political participation (Karakaya and Glazier 2019). Foreign newspapers might also have a substantial impact. Though they do not influence the local freedom of press index, they still play a role in unveiling local government data manipulation.

Our findings provide limited support for the World Bank (2021) report that an independent statistical office can mitigate data manipulation. While the statistical significance of our results is modest, the direction of the coefficients in the underreporting sample consistently indicates a trend toward reducing deviations, aligning with our hypothesis. Importantly, this result relates only to the subsample of African countries, and no statistical link was found in the overreporting sample. 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.

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. In addition, our explanatory variables partly also rely on officially reported statistics. Those are, however, much more detailed and harder to manipulate to be in accordance with the respective GDP manipulation. Most of the variables are indices and comprise different sorts of information, making a systematic bias less likely (though not unlikely). Reverse or simultaneous causality might be a problem in our empirical design. If official data has been manipulated, a government has an incentive to curtail press freedom further. Rampant data manipulation could also lead to other countries cutting ties with the manipulating country or 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.

8. Conclusion

This paper analyzes the manipulation of official data and how different institutional arrangements might affect 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. Various approaches have been employed to cope with this phenomenon. They rely on corrections made by international organizations such as the WTO, which are more trustworthy than the nationally reported data. Kerner et al. (2017) provide evidence for low-income countries, concluding that in contrast to initial GDP statistics, ex-post revised statistics reflect stronger data management practices. Data manipulation biases econometric results of, e.g., policy evaluations. Thus, at least some national policymakers and international organizations are wrongly informed about the state of the economy.

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 from 2013 to 2019. Night-time lighting data is employed to predict "real" GDP. Subsequently, the deviations from predicted to "official" GDP are split into over- and underreporting country samples. Those deviations are explained with determinants suggested by our theoretical hypotheses. In addition, we employ a further outcome measure sum/area and also include the variable informal economy in the first stage regression to minimize the mostly unidirectional influences of underreporting (Table A3). Nightlight data is a good proxy for measuring "real" economic activity (e.g., Henderson et al. 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.

Summarizing our main insights, the study reveals that economic openness and democracy generally reduce data manipulation.

Decentralization increases it for the overreporting sample and has no robust effect in underreporting countries. Political openness shows heterogeneous effects for over- or underreporting samples; it increases manipulation in overreporting countries but decreases it in underreporting ones. This aligns with the results of passport power, which arguably is a comparable measure of international relations. Against our expectations, the effect of press freedom turns out to be unclear. Taking the informal economy into account in the first stage regression, the results of press freedom align with Table 4. A decreasing manipulation effect for overreporting countries and an increasing effect for underreporting countries is observed. These results are robust to an alternative outcome measure and to further control variables such as the internal political openness measuring political rights. However, the specific mechanisms remain to be explored in detail.

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., 2023) or online reviews (e.g., Mayzlin et al. 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 night-time satellite imagery (see e.g., Lehnert et al., 2023), 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 of disciplining the government.

Overall, this paper sheds light on the critical role that institutions play in ensuring the accuracy and reliability of official statistics. Our conclusion: Institutions matter and will also be required to further leverage the potential of nightlight data in the future.

CRediT authorship contribution statement

Andre Briviba: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing, Supervision. **Bruno Frey:** Conceptualization, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing, Resources. **Louis Moser:** Data curation, Formal analysis, Software, Visualization. **Sandro Bieri:** Data curation, Investigation, Visualization.

Declaration of competing interest

I have no conflicts of interest to disclose and have not received financial support from any source. We did not collect data on human subjects.

Data availability

Data will be made available on request.

Acknowledgements

We are grateful for insightful discussions with the participants of the EPCS Conference in Hannover (2023) and for the constructive correspondence with the reviewers. We also want to thank Karlis Briviba, Reiner Eichenberger, René Frey, Patrick Hoferer, Vanessa Kasties, Kai Konrad, Fritz Schneider, Ronnie Schöb, and Margit Osterloh for their encouragement and helpful comments.

A Appendix.

Table A1
Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
GDP (in bln)	1312	420.618	1730.881	0.087	21,433.225
Nighttime Lights Mean	1357	1.706	6.378	0.00001	64.399
Freedom of Press	1033	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	1294	72.179	18.893	14.972	99.693
Econ. Globalization	1259	56.126	18.403	11.771	99.203
Democracy	1203	0.529	0.255	0.016	0.923
Informal Economy	888	28.962	11.301	8.000	64.600
Share Agriculture	1262	10.937	10.614	0.014	60.611
Share Manufacturing	1207	11.852	6.381	0.000	38.733
Share Industry	1267	25.886	11.283	4.556	73.099
Share Services	1263	55.263	11.466	16.775	94.256

Table A2
Pooled Residuals

	Dependent variable:			
	Residuals			
	Mean		Sum/area	
Press Freedom	−0.039	−0.140	−0.409	−0.080
	(0.042)	(0.099)	(0.081)	(0.184)
Ind. Stat. Office		0.022		0.356
		(0.039)		(0.099)
Pol. Globalization	0.067	−0.042	1.589	1.819
	(0.041)	(0.098)	(0.077)	(0.134)
Econ. Globalization	−0.158	−0.143	−0.182	−0.263
	(0.021)	(0.049)	(0.046)	(0.109)
Decentralization	0.124	0.183	0.374	0.268
	(0.026)	(0.049)	(0.055)	(0.124)
Democracy	−0.121	−0.087	−0.177	−0.394
	(0.045)	(0.079)	(0.104)	(0.153)
Year F.E.	Yes	Yes	Yes	Yes
Observations	521	161	521	161
R2	0.504	0.415	0.805	0.722
Adjusted R2	0.490	0.354	0.800	0.693

Notes: This table reports OLS estimates using the pooled residuals as a dependent variable (non-absolute values). All variables have been standardized prior to analysis. Robust standard errors in parentheses. Constant and controls are included but not reported.

Table A3
Residuals, controlled for informal economy

	Dependent variable:					
	Residuals					
	Mean			Sum/area		
	Pos. res	Neg. res	Pooled	Pos. res	Neg. res	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)
Press Freedom	-0.097	0.062	-0.092*	-0.319***	0.305***	-0.345***
	(0.063)	(0.049)	(0.050)	(0.113)	(0.112)	(0.098)
Pol. Globalization	0.323***	-0.493***	-0.149***	0.612***	-0.919***	1.485***
	(0.062)	(0.042)	(0.051)	(0.140)	(0.078)	(0.072)
Econ. Globalization	-0.194***	0.050	-0.156***	-0.103*	-0.106*	-0.218***
	(0.027)	(0.039)	(0.026)	(0.045)	(0.064)	(0.056)
Decentralization	0.153***	0.025	0.173***	0.323***	0.043	0.245***
	(0.037)	(0.039)	(0.033)	(0.084)	(0.066)	(0.067)
Democracy	-0.188**	-0.108*	-0.153**	-0.331**	-0.018*	-0.264**
	(0.076)	(0.065)	(0.061)	(0.154)	(0.109)	(0.126)
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	351	170	521	334	187	521
R2	0.430	0.587	0.278	0.256	0.478	0.596
Adjusted R2	0.408	0.553	0.260	0.225	0.439	0.586

Notes: This table reports OLS estimates for both outcome measures. The dependent variable is the residuals from the regression including informal economy. All variables have been standardized prior to analysis. Columns (1), (2), and (3) use the mean night light intensity; columns (4), (5), and (6) use the sum/area night light intensity. Robust standard errors in parentheses. Constant and control variables are included in all calculations.

Noteworthy deviations from the initial regressions (without informal economy in the first stage regression) are marked with a grey background color.

References

Adam, Antonis, Tsarsitalidou, Sofia, 2022. Data misreporting during the COVID19 crisis: the role of political institutions. Econ. Lett. 213, 110348.

Adiguzel, Fatih Serkant, Cansunar, Asli, Corekcioglu, Gozde, 2020. Truth or dare? Detecting systematic manipulation of COVID-19 statistics. Journal of Political Institutions and Political Economy 1 (4), 543–557.

Alt, James, Dreyer Lassen, David, Wehner, Joachim, 2014. It isn't just about Greece: domestic politics, transparency and fiscal gimmickry in Europe. Br. J. Polit. Sci. 44 (4), 707–716.

Aragão, Roberto, Linsi, Lukas, 2022. Many shades of wrong: what gov- ernments do when they manipulate statistics. Rev. Int. Polit. Econ. 29 (1), 88–113.

Balashov, Vadim S., Yan, Yuxing, Zhu, Xiaodi, 2021. Using the New- comb–Benford law to study the association between a country's COVID- 19 reporting accuracy and its development. Sci. Rep. 11 (1), 22914.

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

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

Buchanan, James M., Gordon, Tullock, 1965. The Calculus of Consent: Logical Foundations of Constitutional Democracy, vol. 100. University of Michigan Press, Michigan.

Chan, Ho Fai, Frey, Bruno S., Ahmed, Skali, Torgler, Benno, 2019. Po- litical entrenchment and GDP misreporting. CESifo Working Paper Series 7653. CESifo.

Chen, Xi, Nordhaus, William D., 2011. Using luminosity data as a proxy for economic statistics. Proc. Natl. Acad. Sci. USA 108 (21), 8589–8594.

Chen, Shuo, Qiao, Xue, Zhu, Zhitao, 2021. Chasing or cheating? Theory and evidence on China's GDP manipulation. J. Econ. Behav. Organ. 189, 657–671.

Coremberg, Ariel, 2014. Measuring Argentina's GDP growth. World Econ. 15 (1), 1–32.

Cronin, D., McInerney, N., 2023. Official fiscal forecasts in EU member states under the European Semester and Fiscal Compact – an empirical assessment. Eur. J. Polit. Econ. 76, 102227.

Dafflon, Bernard, Rossi, Sergio, 1999. Public accounting fudges to- wards EMU: a first empirical survey and some public choice consid- erations. Publ. Choice 101 (1), 59–84.

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

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

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

Ebener, Steeve, Murray, Christopher, Tandon, Ajay, Elvidge, Christopher C., 2005. From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light im- agery. Int. J. Health Geogr. 4 (1), 1–17.

Elvidge, Christopher D., Zhizhin, Mikhail, Ghosh, Tilottama, Hsu, Feng-Chi, Taneja, Jay, 2021. Annual time series of global VIIRS night-time lights derived from monthly averages: 2012 to 2019. Rem. Sens. 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 Communi- ties.

Feldstein, Martin, 2017. Underestimating the real growth of GDP, per- sonal income, and productivity. J. Econ. Perspect. 31 (2), 145–164.

Geers, Sabine, 2020. News consumption across media platforms and con- tent: a typology of young news users. Publ. Opin. Q. 84 (S1), 332–354.

Georgiou, Andreas V., 2021. The manipulation of official statistics as cor- ruption and ways of understanding it. Stat. J. IAOS 37 (1), 85–105.

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. *Open Geogr. J.* 3 (1): 147–160.

Ghosh, Tilottama, Anderson, Sharolyn, Rebecca, L., Powell, Paul C Sutton, Elvidge, Christopher D., 2009. Estimation of Mexico's informal economy and remittances using night-time imagery. Rem. Sens. 1 (3), 418–444.

Henderson, J Vernon, Adam, Storeygard, Weil, David N., 2012. Measur- ing economic growth from outer space. Am. Econ. Rev. 102 (2), 994–1028.

Hodler, Roland, Paul, A. Raschky, 2014. Regional favoritism. Q. J. Econ. 129 (2), 995–1033.

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

Jean, Neal, Burke, Marshall, Xie, Michael, Davis, W Matthew, Lo- bell, David B., Ermon, Stefano, 2016. Combining satellite imagery and ma- chine 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. Publ. Manag. Rev. 18 (1), 110–138.

Karakaya, Suveyda, Glazier, Rebecca A., 2019. Media, information, and political participation: the importance of online news sources in the absence of a free press. J. Inf. Technol. Polit. 16 (3), 290–306.

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

Kerner, A., Jerven, M., Beatty, A., 2017. Does it pay to be poor? Testing for systematically underreported GNI estimates. The Review of International Organizations 12 (1), 1–38.

King, Gary, Pan, Jennifer, Roberts, Margaret E., 2017. How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. Am. Polit. Sci. Rev. 111 (3), 484–501.

Koen, Vincent, Van den Noord, Paul, 2005. Fiscal gimmickry in Europe: one-off measures and creative accounting. In: OECD Economics Depart- Ment Working Papers, vol. 417. OECD Publishing, Paris.

Gygli, Savina, Haelg, Florian, Potrafke, Niklas, Sturm, Jan-Egbert, 2019. The KOF Globalisation Index – Revisited. Rev. Int. Org. 14 (3), 543–574. https://doi.org/10.1007/s11558-019-09344-2.

Kubo, Koji, 2012. Trade policies and trade misreporting in Myanmar. ASEAN Econ. Bull. 29 (2), 146–159.

Lehnert, P., Niederberger, M., Backes-Gellner, U., Bettinger, E., 2023. Proxying economic activity with daytime satellite imagery: Filling data gaps across time and space. PNAS Nexus 2 (4), gad099.

Liang, W., Sim, N., 2019. Did rainfall shocks cause civil conflict in Sub-Saharan Africa? The implications of data revisions. Eur. J. Polit. Econ. 60, 101808.

Lorentzen, Peter, 2014. China's strategic censorship. Am. J. Polit. Sci. 58 (2), 402–414.

Magee, Christopher S.P., Doces, John A., 2015. Reconsidering regime type and growth: lies, dictatorships, and statistics. Int. Stud. Q. 59 (2), 223–237.

Martinez, Luis R., 2022. How much should we trust the dictator's GDP growth estimates? J. Polit. Econ. 130 (10).

Mayzlin, Dina, Dover, Yaniv, Chevalier, Judith, 2014. Promotional reviews: an empirical investigation of online review manipulation. Am. Econ. Rev. 104 (8), 2421–2455.

Milesi-Ferretti, Maria, Gian, 2004. Good, bad or ugly? On the effects of fiscal rules with creative accounting. J. Publ. Econ. 88 (1–2), 377–394.

Nathan, Andrew J., 2020. The puzzle of authoritarian legitimacy. J. Democr. 31 (1), 158–168.

Nguyen, Q.M.P., Do, H.X., Molchanov, A., Nguyen, L., Nguyen, N.H., 2023. Political similarities in credit ratings. Int. Rev. Fin. Anal. 86, 102515. https://doi.org/10.1016/j.irfa.2023.102515.

Open Data Watch, 2015-2020. Open Data Inventory. https://odin.opendatawatch.com/Report/rankings. (Accessed 15 May 2022).
Rauch, Bernhard, Göttsche, Max, Engel, Stefan, Brähler, Gernot, 2011. Fact and fiction in EU-governmental economic data. German Eco- nomic Review 12 (3), 243–255.
Reporters Without Borders, 2002-2022. World Press Freedom Index rsf.org/en/index. (Accessed 15 May 2022).
Sandefur, Justin, Glassman, Amanda, 2015. The political economy of bad data: evidence from african survey and administrative statis- tics. J. Dev. Stud. 51 (2), 116–132.
Sutton, Paul C., Costanza, Robert, 2002. Global estimates of market and non-market values derived from night-time satellite imagery, land cover, and ecosystem service valuation. Ecol. Econ. 41 (3), 509–527.
Sutton, Paul C., Elvidge, Christopher D., Ghosh, Tilottama, et al., 2007. Es- timation of gross domestic product at sub-national scales using night- time satellite imagery. Int. J. Ecol. Econ. Stat. 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.
The Economist, 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.
The Economist, 2017. May 25 Welcome back. Argentina's new, honest inflation statistics. economist. com/the- americas/2017/05/25/argentinas-new-honest-inflation-statistics.
The Economist, 2020. October 15 Can China's reported growth be trusted?. economist. com/finance - and- economics/2020/10/15/can- chinas-reported-growth-be-trusted.
The Economist, 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.
The Economist, 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.
The Economist, 2022a. A study of lights at night suggests dictators lie about eco- nomic growth 21–22. September 29. https://www.economist.com/graphic-detail/2022/09/29/a-study-of-lights-at-night-suggests-dictators-lie-about-economic-growth.
The Economist, 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.
The Economist, 2022d. Turkey grapples with triple-digit inflation 21–22. July 14. https://www.economist.com/europe/2022/07/14/turkey-grapples-with-triple-digit-inflation.
Von Hagen, Jürgen, Wolff, Guntram B, 2006. What do deficits tell us about debt? Empirical evidence on creative accounting with fiscal rules in the EU. J. Bank. Finance 30 (12), 3259–3279.
Wallace, Jeremy L., 2016. Juking the stats? Authoritarian information problems in China. Br. J. Polit. Sci. 46 (1), 11–29.
Wang, Shaoda, Yang, David Y., 2021. Policy Experimentation in China: the Political Economy of Policy Learning. National Bureau of Economic Research. NBER Working Paper Se- ries 29402.
Weidmann, Nils B., Schutte, Sebastian, 2017. Using night light emis- sions for the prediction of local wealth. J. Peace Res. 54 (2), 125–140.
Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Fabio Angiolillo, Michael Bernhard, Cecilia Borella, Agnes Cornell, M. Steven Fish, Linnea Fox, Lisa Gastaldi, Haakon Gjerløw, Adam Glynn, Ana Good God, Sandra Grahn, Allen Hicken, Katrin Kinzelbach, Joshua Krusell, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Juraj Medzihorsky, Natalia Natsika, Anja Neundorf, Pamela Paxton, Daniel Pemstein, Josefine Pernes, Oskar Rydéen, Johannes von Römer, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Aksel Sundström, Eitan Tzelgov, Yi-ting Wang, Tore Wig, Steven Wilson and Daniel Ziblatt. 2024. "V-Dem [Country-Year/Country-Date] Dataset v14" Varieties of Democracy (V-Dem) Project. doi:10.23696/mcwt-fr58.
World Bank 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 Bank, 1990-2018. Informal Economy Database. https://www.worldbank.org/en/research/brief/informal-economy-database. (Accessed 10 May 2022).
World Bank, 2021. World Development Report 2021: Data for Better lives.World Development Report. World Bank, Washington.
World Bank, 2022. GDP (Current US$). https://data.worldbank.org/indicator/NY.GDP.MKTP.CD. (Accessed 10 May 2022).
World Wide Web Foundation, 2013-2017. Open Data Barometer Global Report (All Editions). https://opendatabarometer.org/4thedition/data/. (Accessed 15 May 2022).
Yang, Fang, Shanahan, James, 2003. Economic openness and media penetration. Commun. Res. 30 (5), 557–573.
Zheng, Tianxiang, Wu, Feiran, Law, Rob, Qiu, Qihang, Wu, Rong, 2021. Identifying unreliable online hospitality reviews with biased user- given ratings: a deep learning forecasting approach. Int. J. Hospit. Manag. 92, 102658.