When Countries Revise Their Data*

Iasmin Goes†

March 2025

Abstract

Macroeconomic indicators like GDP, trade, unemployment, and foreign direct investment are routinely revised to incorporate better data, refine methodologies, and correct errors. What drives the frequency and magnitude of these revisions? Using GDP data from the World Development Indicators (1994–2021), I show that revisions are more frequent for data collected by democracies and IMF borrowers, as freedom of expression and reliance on foreign credit promote transparency and scrutiny. However, the magnitude of revisions varies: while democracies tend to report larger adjustments, IMF borrowers may limit revisions to protect their reputations. Additionally, the conditions at the time of data collection matter more than those at the time of revision, as later adjustments remain constrained by initial reporting practices. These findings highlight a trade-off between validity and reliability. Revising data is a statistical best practice that improves validity but introduces inconsistencies, which can undermine public trust in official statistics.

*Thanks to Ryan Briggs, Terry Chapman, Andrés Cruz, Kerice Doten-Snitker, Andrew Kerner, Carolina Moehlecke, Daniel Weitzel, and audiences at Colorado State University, Fundação Getulio Vargas, IE University, the 2023 European Workshops in International Studies, the 2022 ACUNS–UN Workshop, and the 2022 IPE Conference of the Austrian, German, and Swiss Political Science Associations for feedback. Thanks to Sabrina Arias and Martina Müller for sharing data.

†Assistant Professor, Colorado State University. Contact: iasmin.goes@colostate.edu.

1 Introduction

In August 2024, the United States Bureau of Labor Statistics revised its preliminary employment data, finding that 818,000 fewer jobs had been created in March 2024 than initially reported. This scheduled annual benchmark revision, while the largest in 15 years, came as no surprise to experts: initial job estimates are always based on incomplete household and business surveys, and as survey funding and response rates decline over time, the uncertainty around initial estimates tends to increase. The magnitude of the revision was marginal: it meant that 159.2 million individuals, not 160 million individuals, were employed in the US in March 2024. However, in the wake of the revision, then-presidential candidate Donald Trump accused then-president Joe Biden of "fraudulently manipulating job statistics," an opinion echoed by several others in the Republican party.[1] Once elected, President Trump took more drastic steps, firing thousands of federal workers, deleting government datasets, and disbanding expert panels like the Federal Economic Statistics Advisory Committee.

The US is not unique in facing controversy over economic statistics. Around the world, data revisions have had far-reaching political and economic consequences. In 2009, Prime Minister George Papandreou came to power in Greece and requested help from Eurostat and the International Monetary Fund (IMF) to revise the country's finances. As these revisions quickly established, previous administrations had overestimated tax revenues, disregarded social security and military expenses, and engaged in creative accounting to hide government liabilities. Between April and October 2009, the planned deficit ratio for 2009 was revised from 3.7 to 12.5 percent of the Gross Domestic Product (GDP) (European Commission, 2010), a number that — according to Eurostat — was still far below the truth (Aragão and Linsi, 2022). Following these revisions, credit rating agencies downgraded Greece, which requested multiple IMF and EU loans to avoid default.

A revision in Ireland caused similar controversy. From 2015 to 2016, the Irish Central

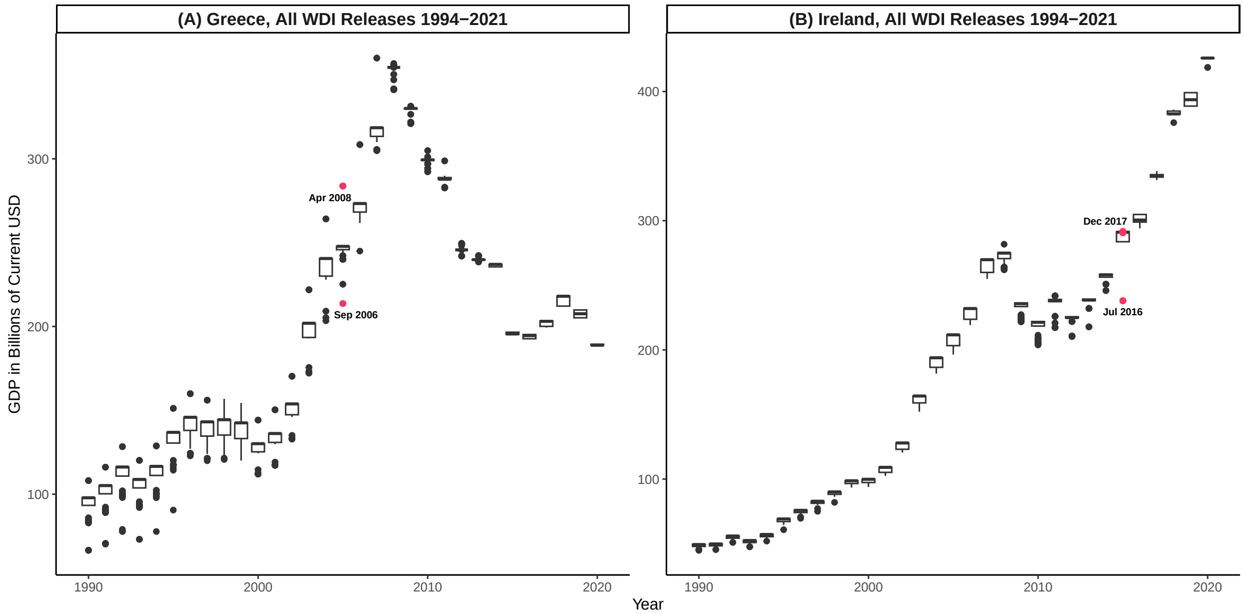
[1] Alicia Wallace. "Trump Routinely Calls Economic Data 'Fake.' Here's Why That's Dangerous." *CNN*. 26 January 2025.

Statistics Office reported a GDP growth of 26.3 percent. Citing statistical confidentiality rules, authorities initially refused to release detailed data, instead devising a new statistic — the Modified Gross National Income (GNI*) — to remove "globalization-related" distortions. Ireland's "leprechaun economics" became a source of ridicule, and the country was blacklisted as a tax haven. Years later, economists discovered that these distortions had been driven by Apple's decision to onshore intellectual property assets to Ireland in 2015 (Polyak, 2023).

Beyond Europe, the Ghana Statistical Service released new GDP estimates in 2010, with support from the Danish International Development Agency and the IMF. After upgrading from the 1968 to the 1993 System of National Accounts (SNA), updating the base year, and disaggregating data by economic sector, it concluded that the country's GDP was 60.3 percent larger than previously thought (Jerven and Ebo Duncan, 2012). Kenya, Nigeria, and others similarly reported GDP increases after incorporating new information from informal activities (African Development Bank, 2013). In the following years, the World Bank upgraded all three countries from low income to lower middle income economy. This shift was associated with less generous lending terms, as countries with per capita incomes above a certain threshold lose access to concessional lending (Kerner, Jerven and Beatty, 2017).

As these examples show, measuring the economy is an iterative process: preliminary estimates are routinely revised to reflect methodological improvements, incorporate new data, capture routine recalculations, and correct mistakes (Carson, Khawaja and Morrison, 2004). Macroeconomic indicators like GDP, trade, foreign direct investment, and unemployment are pieced together from many different sources, and not even the most advanced nations can get the "correct" numbers upfront. As a result, new data often contradict previous data. Consider Figure 1, which uses data from the World Development Indicators (WDI). Between the September 2006 and April 2008 WDI releases, Greece's GDP for 2005 increased by nearly 33 percent, from 213.7 billion to 283.7 billion. Between the July 2016 and the December 2017 WDI, Ireland's 2015 GDP figures increased by 22.6 percent. It might seem counterintuitive, but revisions are a statistical best practice, a natural and necessary com-

Figure 1: Current GDP of Greece and Ireland, 1990–2020



These boxplots present the distribution of current GDP estimates for (A) Greece and (B) Ireland from 1990 to 2020, using data drawn from the 104 WDI releases from April 1994 to December 2021. The estimate reported for Greece in 2005 is 70 billion dollars (32.78 percent) larger in the April 2008 WDI than in the April 2007 WDI. The estimate reported for Ireland in 2015 is 53.8 billion dollars (22.6 percent) larger in the December 2017 WDI than in the July 2016 WDI. Section 4 discusses the data in detail.

ponent of the data production process. Revisions are widespread; the IMF, the World Bank, AFRISTAT, Eurostat, and other international organizations provide regular assistance to countries from Albania to Zambia to improve their data collection and revision process.

Revisions might be a statistical best practice but are also a political liability. Acknowledging past errors can undermine the government's reputation, underscore its low commitment to transparency, trigger market instability, and erode public trust, particularly if the magnitude of revisions is large. The general public often misunderstands the purpose of revisions, and opposition parties — like the Republican party in the US — capitalize on this misunderstanding. As the cases of Ghana, Kenya, and Nigeria show, even *upward* revisions can be costly, weakening the government's bargaining position in international financial negotiations. Countries that revise their data risk severe political, economic, and reputational consequences. Given this risk, what explains the likelihood and magnitude of revisions?

I structure this study along the two dimensions of data quality: validity and reliability (McMann et al., 2022). Valid data accurately capture the underlying theoretical concept, whereas reliable data provide consistent information across repeated measurements. For example, valid GDP estimates quantify a country's economic output with precision, correct methodology, and minimal error in data collection, whereas reliable GDP estimates provide consistent results across different data sources and releases. I begin by examining why countries fail to disseminate valid data in the first place: due to high political interference, low statistical capacity, and poor data management. Moving on to reliability, I investigate what drives revisions to previously disseminated data — a less studied phenomenon[2] for which no systematic explanations exist.

My starting assumption is that governments with higher state capacity should revise their data more frequently: they have better-trained statisticians, regular household surveys, and comprehensive administrative data systems that allow them to consistently incorporate new information. Yet conditional on the *capacity* to revise, I argue that revisions are more likely when experts act as *accountability agents*, pressuring governments to collect and disseminate comprehensive official estimates that undergo posterior scrutiny. These agents of accountability can be national or international.

At the national level, democracies tend to have independent statistical offices that publish their data sources and methodology. This, coupled with academic freedom and free speech, empowers experts to scrutinize official records, identify discrepancies, and demand corrections. Through media coverage, public debate, and academic research, these experts compel democratic governments to acknowledge errors and correct official estimates. My argument is not that democracies *want* to revise their data; rather, journalists, academics, and other experts *compel* democracies to do so. At the international level, IMF staff play a similar role in enforcing transparency. The IMF closely scrutinizes borrowing economies to ensure that the disbursed funds are meeting pre-established targets. This requires gov-

[2]For an important exception, see Fariss et al. (2022), who quantify the reliability of GDP, GDP per capita, and population measurements.

ernments to improve data collection and disclose economic indicators in ways that conform to international standards. Even if governments comply reluctantly, data collected under the oversight of IMF staff should be more transparent, increasing the likelihood of future monitoring and updating even after the agreement ends. Together, free speech and foreign credit generate demand for transparent data from the outset, forcing governments to bear the political cost of subsequent data revisions. As a result, countries facing such pressures are more likely to update (or "vintage") their data. I find support for my expectations in an analysis of GDP data published by the WDI between 1994 and 2021.

I conclude by discussing three implications. First, democracies produce data with higher validity but lower reliability: democratically-elected leaders are less likely to lie about growth rates or COVID-19 deaths (Martínez, 2022; Adiguzel, Cansunar and Corekcioglu, 2020), yet their numbers tend to change frequently over time. Even revisions resulting from good-faith improvements can create a perception of mismanagement and unreliability — an issue that is politically costlier for democracies, which are held to higher transparency standards than autocracies. In conducting regular revisions, democratic governments give their opposition fodder that can lead to an asymmetric credibility loss.

Second, the context of data *collection* is more important than the context of data *revision*. Aside from technical and methodological refinements, revisions tend to reflect structural factors present at the time of initial data collection rather than ex post influences. Once information is collected, the potential for improvement is limited: information lost at the source cannot be recovered retroactively. For all their technical improvements, revisions cannot undo the limitations imposed by the initial data collection process. It is important to collect transparent information from the outset; future revisions will be infeasible unless the initial data collection process is explicitly designed with future revisions in mind. This implication is particularly poignant for countries dismantling their bureaucratic and statistical infrastructure, like the United States under President Trump: this dismantlement will have irreversible consequences.

Lastly, my results speak to replications by Goes (2023), Johnson et al. (2013), and Croushore and Stark (2003), who show that published studies would come to significantly different conclusions depending on the chosen data version — after all, new versions often modify previous data. There is widespread heterogeneity in data quality: researchers can make more consistent inferences about some countries and periods than others. To my knowledge, this study is the first to examine the systematic predictors of such heterogeneity.

2 Recording and Revising Economic Data

2.1 Recording Valid Data

Macroeconomic indicators have long faced criticism for oversimplifying abstract concepts like wealth, inequality, or unemployment (Mügge, 2022). For example, GDP excludes unpaid household services, which are disproportionately performed by women (DeRock, 2021). But even when experts agree on how to define these concepts, the resulting measurements often lack validity — they fail to capture the "correct" information — due to political interference, low statistical capacity, and inadequate data management.

In terms of political interference, autocracies are less likely to report policy-relevant data (Hollyer, Rosendorff and Vreeland, 2011); when they do, they overstate growth rates (Martínez, 2022; Magee and Doces, 2015), particularly in politically sensitive times (Wallace, 2014), and underreport COVID-19 deaths (Adiguzel, Cansunar and Corekcioglu, 2020). Some directors of national statistical offices (NSOs) are political appointees who lack autonomy and can be dismissed at any moment, generating incentives to misreport data. In federations like Nigeria, states inflate population figures to receive higher fiscal transfers from the federal government (Devarajan, 2013). Aid-dependent countries systematically understate their finances to appear poor and attract more aid (Kerner, Jerven and Beatty, 2017). Even industrialized democracies overstate how much climate aid they provide — particularly when domestic constituencies value environmental objectives (Michaelowa and

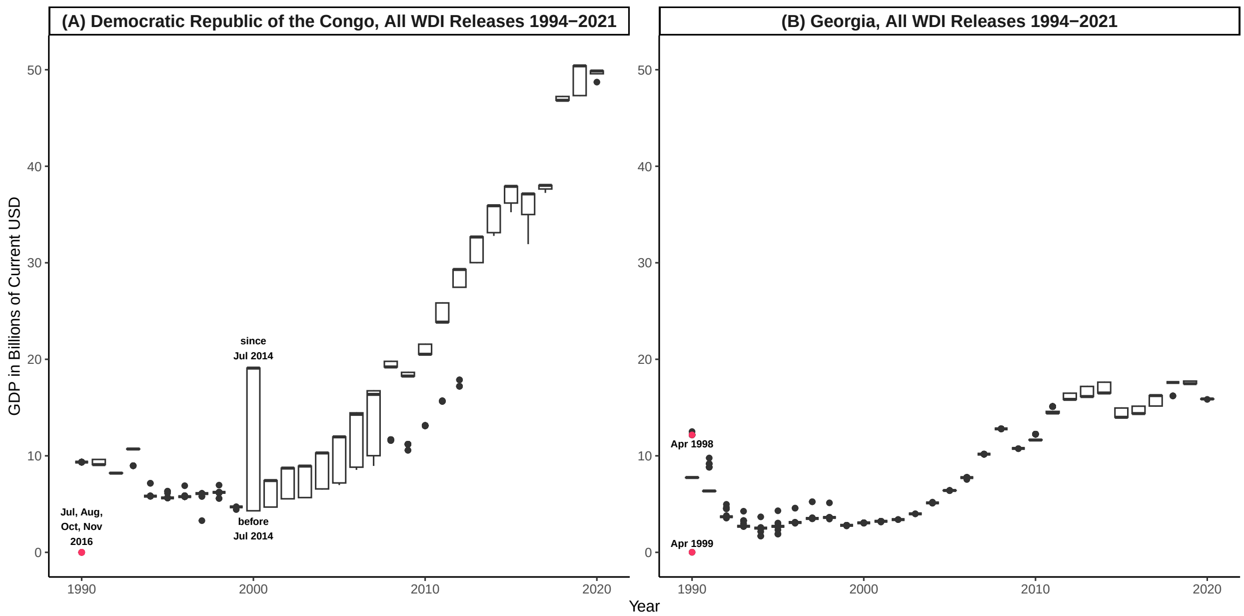
Michaelowa, 2011) — and misrepresent public finance statistics to abide by the rules of the European Union, as Greece did (Alt, Lassen and Wehner, 2014). Conversely, political competition and frequent turnover generate uncertainty about future outcomes, motivating incumbents to pass Freedom of Information (FOI) laws that increase transparency (Berliner, 2014).

Estimates may also lack validity due to low statistical capacity. NSOs might be underfunded, understaffed, use outdated methods, or experience frequent turnover, which limits their ability to collect, standardize, and disseminate high-quality data. Since data collection is expensive, NSOs often rely on outdated information about businesses and households. Population figures tend to be extrapolated from the last census; in countries like Lebanon (which last conducted a census in 1932), these extrapolations grow progressively inaccurate over time (Devarajan, 2013). Some countries might halt data collection altogether due to natural disasters and civil war.

A 2005 survey by the United Nations Economic Commission for Africa (2005) found that some NSOs in the continent had as few as three national accountants. A 2023 survey of 14 NSOs, conducted by the Inter-American Development Bank, found that only half of the employees working with statistical analysis displayed basic competence in probability, descriptive statistics, survey sampling, and arithmetic (Mejía Guerra et al., 2023, 14). Many lacked the expertise to report data consistent with the SNA, a global standardization framework, and struggled with difficult-to-measure concepts like imputed rent, thus underestimating household final consumption expenditure — an important component of GDP (Olinto Ramos, Pastor and Rivas, 2008). Moreover, it is difficult to quantify the size of the informal economy, which accounts for up to 44 percent of the GDP in the developing world (Coyle, 2014, 110).

Idiosyncratic data management errors pose a final threat to validity. As Figure 2 shows, four different WDI releases reported the GDP of the Democratic Republic of the Congo in 1990 as *zero*; two other releases reported this value as missing. Georgia's 1990 GDP —

Figure 2: Current GDP of the Democratic Republic of the Congo and Georgia, 1990–2020

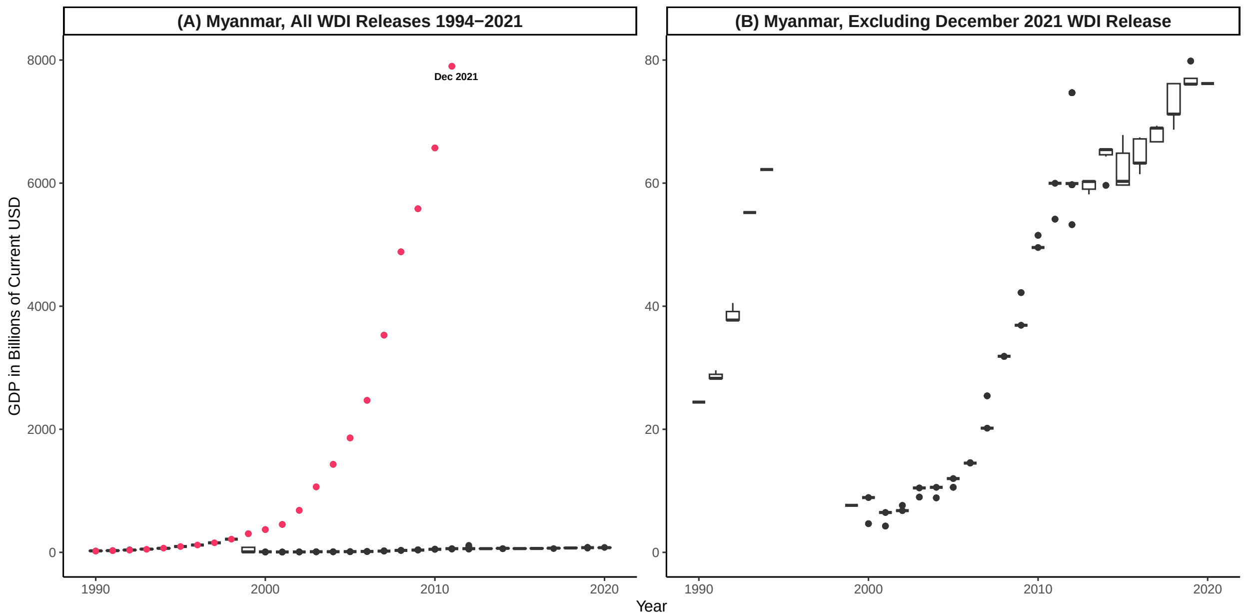


These boxplots present the distribution of current GDP estimates from 1990 to 2020 for (A) the Democratic Republic of the Congo and (B) Georgia, using data drawn from the 104 WDI releases from April 1994 to December 2021. Four WDI releases reported a GDP of zero for the Democratic Republic of the Congo in 1990. In addition, all 32 releases before July 2014 reported a GDP of 4.3 billion for 2000, a figure revised to 19.1 billion in July 2014. Georgia's GDP in 1990 was reported as 12.1707 *million* in some vintages and 12.1707 *billion* in others. Section 4 discusses the data in detail.

reported to be around 12.1707 *billion* until April 1998 — momentarily "lost" three digits in the April 1999 and April 2000 vintages, shrinking to 12.1707 *million*.[3] The December 2021 update contains a similar error for Myanmar, illustrated in Figure 3. In nearly all available vintages, Myanmar's GDP in 2011 ranged from 54 to 59 billion current US dollars. However, the December 2021 release reported a figure over 100 times as high: 7.899 *trillion*. Other trivial errors include: in 2006, "the country names for Burundi and Cameroon are in reverse order;" in 2008, "columns are incorrectly labeled as 1990; data are for 1995;" and in 2010, "an error for Zimbabwe's data" meant that several indicators, including GDP, "should be presented as not available for all years in the WDI database" (see World Bank 2023 for errata). The most plausible explanation for these singular discrepancies is a data

[3] Though Georgia gained formal independence from the Soviet Union in December 1991, its WDI coverage begins in 1990.

Figure 3: Current GDP of Myanmar, 1990–2020



These boxplots present the distribution of current GDP estimates from 1990 to 2020 for Myanmar, using data drawn from the 104 WDI releases from April 1994 to December 2021. The December 2021 WDI release (in pink) is included in (A), but not in (B). As the different y-axes show, the December 2021 release was an outlier, reporting exceptionally high values for the entire time series. Section 4 discusses the data in detail.

management error — perhaps the worst kind of error, as it is impossible to predict.[4]

2.2 Correcting the Record

The timely release of macroeconomic data increases the quality of governance by allowing for better, more informed policymaking (Islam, 2006). Accordingly, international organizations advise countries to publish preliminary annual data swiftly, then revise estimates following a regular, publicized schedule that clearly distinguishes between preliminary and updated data.[5] About 39.8 percent of all economies surveyed by the IMF in 2020 first released

[4]In April 2022, I raised some of these issues to the WDI team via e-mail; three months later, a member of the World Bank's Development Data Group responded that they were related to the timing of the IT team's periodical maintenance and suggested I clear my browser cache. Doing so did not rectify these issues.

[5]As of 2024, 95 percent of IMF members subscribe either to the Special Data Dissemination Standard (SDDS) or the enhanced General Data Dissemination System (e-GDDS). SDDS subscribers must disseminate preliminary national accounts data with a lag of no more than one quarter after the end of the reference period. Subscribers of the less demanding e-GDDS do not have a hard cutoff. These preliminary estimates should not be confused with flash or advance estimates, which many industrialized nations release within 30

annual GDP data within 90 days of the reference period, whereas 54.4 percent did so within 91 to 365 days, with no available information for the remaining 5.8 percent (Baer, Guerreiro and Silungwe, 2022, 17). Many of these economies later revised their preliminary data, as recommended.

Croushore and Stark (2003) identify two types of data revisions: information-based and structural. Information-based revisions occur when countries "discover" new data that allow for more precise measurement. Every five to ten years, the International Comparison Program (ICP) surveys how much the same basket of goods costs in different currencies and constructs purchasing power parity (PPP) exchange rates. These exchange rates, in turn, are used to convert SNA data from nominal (current) to PPP terms, which are comparable across borders. Until 1996, ICP price surveys only covered the developed world, making less accurate extrapolations for the developing world (Deaton and Aten, 2017). ICP rounds in 2005, 2011, and 2017 reduced uncertainty by incorporating new information from large developing countries, leading to substantial data revisions.[6] In a more specific case, recall Greece's "discovery" of social security and military expenses, which led to a ballooning deficit in 2009. In a study of quarterly data, Croushore and Stark (2003) show that information-based revisions are most common up to one year after the initial data release. In subsequent years, revisions usually reduce noise, or measurement error.

In contrast, structural data revisions occur when there are changes in the definition of concepts, the base year, or the aggregation method. For example, the 1993 SNA introduced the concept of imputed rent, which significantly altered the definition and calculation of GDP (Olinto Ramos, Pastor and Rivas, 2008). In 2014, EU countries revised their GDP definition to include drug trafficking and prostitution; as a result, the Italian and British economies increased by four percent each (Coyle, 2014, 110). Nicaragua revised its national accounts in 2003, changing the base year from 1980 to 1994 and implementing the 1993 SNA; as a

days of each quarter's end using incomplete data.

[6] However, price surveys in China were only conducted in urban areas, introducing yet another potential measurement issue (Bolt and van Zanden, 2024).

result, the country's current GDP for the year 2000 increased by 70 percent (Olinto Ramos, Pastor and Rivas, 2008, 9). As these examples show, the more time has elapsed since the original data release, the more technical the nature of revisions.

To illustrate the prevalence of such revisions, I briefly discuss the IMF-produced Report on the Observance of Standards and Codes (ROSC) Data Module,[7] which reviews countries' data capabilities as needed, at each country's request. Most ROSC Data Modules were conducted between 1999 and 2010, and 98 percent of them have been published (Pardo, 2011). As of 2024, 87 countries have requested at least one ROSC Data Module, including high-income democracies like France and Norway, but also autocracies at various income levels, such as Belarus and Tajikistan. This indicates that many governments are interested in voluntarily improving their data collection, dissemination, and revision practices.

ROSC Data Modules praise countries that regularly revisit their data: "The revision of national accounts follows regular and publicized procedures ... The magnitude of the revisions is always investigated. When revised figures are published, significant revisions are commented and explained in the text" (Kazakhstan, 2003). Those that do not revisit their data are advised to do so: "Studies and analyses of revisions should be conducted routinely and used to inform statistical processes and data users" (Sweden, 2001). Others are encouraged to publish multiple data versions: "Even preliminary data, with the understanding that these are subject to revisions, would be useful" (Uruguay, 2001). Two otherwise very different countries, Sweden and Oman, received identical criticism in their 2001 and 2005 reports, respectively: "Data are considered final when first published," with revisions not carried out routinely, only on an ad-hoc basis. Chad's 2007 ROSC mirrors this criticism: "No revision studies are conducted for national accounts and BOP [Balance of Payments] statistics, although they would usefully inform the statistical processes."

These reports assume that preliminary data are imperfect but can be improved: over time, revisions get closer to the "correct" information, increasing the validity of the data.

[7] All ROSC Data Modules are available at https://www.imf.org/en/Publications/rosc?sortBy=Topic&sortVal=Data%20Dissemination

Of course, this assumption could be wrong: revisions could reflect a *decrease* in validity. Some governments could be revising their data in bad faith, introducing mistakes to accurate preliminary data. While this revisionist scenario is admittedly difficult to observe, evidence shows that today's ill-intentioned authorities are more sophisticated than Stalin (Guriev and Treisman, 2019); they do not cook the books ex post. Instead of retroactively replacing "correct" information with fictitious information, governments withhold statistics (like Venezuela since 2015), postpone the initial release (Zimbabwe in 2019), or release doctored numbers to begin with (Argentina in 2007–2008). Previously published statistics might be erased from public records (as with the United States in 2025), but do not tend to be replaced with falsified information.

2.3 Validity vs. Reliability

There is a trade-off between validity and reliability. If governments prioritize validity ("correctly" measuring their national accounts), they will revise their data each time new, more accurate information becomes available, even if this comes at the expense of reliability. Conversely, if governments prioritize reliability (minimizing changes over time to maintain consistency), they may sacrifice validity, failing to incorporate new information so their statistics remain constant over time. To be clear, revisions are not the only threat to reliability. Different data sources can also report inconsistent data. The two most common data sources in political science and economics — the WDI and the Penn World Table, respectively (Goes, 2023; Johnson et al., 2013) — provide estimates that can differ by over 25 percent (Ram and Ural, 2014). Exporters and importers record the same bilateral trade flows differently (Linsi, Burgoon and Mügge, 2023), and a comparison of export data from two sources — the IMF and the UN Commodity Trade Statistics — concludes that oftentimes the data are not even correlated (Amin Gutiérrez de Piñeres, 2006, 35). Climate aid (Michaelowa and Michaelowa, 2011; Weikmans and Roberts, 2019) and FDI (Kerner, 2014) suffer from similar discrepancies. Even within one country, different agencies might compete for data collection, presenting contradictory results (Pellechio and Cady, 2006). Yet inconsistency across different versions *of the same source* is even more consequential. Revisions reflect the need to correct *the same data* even when holding definitions, extrapolation methods, ICP benchmarks, and price level adjustments constant.

To some extent, ROSC Data Modules consider it tolerable to pursue validity at the expense of reliability: "Based on annual figures, the discrepancy in recent years has generally been within a very acceptable range of less than 1% of GDP" (Estonia, 2001). However, the magnitude of revisions matters. Referring to a revision of Greece's planned deficit ratio (from 3.7 to 12.5 percent of GDP), a report by the European Commission (2010, 3) states: "Revisions of this magnitude in the estimated past government deficit ratios have been extremely rare in other EU Member States, but have taken place for Greece on several occasions. These most recent revisions are an illustration of the lack of quality of the Greek fiscal statistics (and of macroeconomic statistics in general) and show that the progress in the compilation of fiscal statistics in Greece, and the intense scrutiny of the Greek fiscal data ..., have not sufficed to bring the quality of Greek fiscal data to the level reached by other EU Member States."

This is the paradox of frequent revisions: *if small*, they enhance a country's credibility, signaling a desire to improve already valid data in line with international data management standards put forward by organizations like the IMF and the EU. *If large*, revisions signal deep institutional issues, including a lack of "independence, integrity and accountability of the national statistical authorities" (according to the aforementioned European Commission report). Overall, statistical officers face a delicate balance: they must publish valid data swiftly and increase the validity of these data through routine revisions, but large revisions jeopardize the NSO's credibility, signalling not only low reliability but also low validity. Given this delicate balance, it is not self-evident that countries will revise their data unless pressured to do so. In the next section, I develop a theory that treats data revisions as a governmental response to accountability pressures.

3 How Accountability Pressures Drive Data Revisions

My starting assumption is that countries with greater state capacity are better able to collect, disseminate, and revise macroeconomic data in a timely manner. These countries can train and retain qualified personnel, conduct regular surveys, digitize administrative records, audit past estimates, and facilitate cooperation between agencies (for example, between the NSO, tax agency, and finance ministry). However, state capacity is a necessary but not sufficient condition. Several high-income Middle Eastern nations that could afford to release and revise their data choose not to, for political or institutional reasons (Williams, 2009). Beyond the *ability* to revise, countries must be *compelled* to do so. I argue that experts serve precisely this function: they are *accountability agents* that pressure for regular revisions, no matter how high the political cost of doing so.

I begin by discussing the role of domestic accountability agents. As outlined in previous sections, Williams (2009), Hollyer, Rosendorff and Vreeland (2011), Magee and Doces (2015), and Martínez (2022) find that autocracies and unconstrained executives are more likely to withhold or embellish data. I argue that regime type also plays a key role in the revision process: evidence produced by autocracies is less likely to be revised. Since autocracies are less committed to transparency, they are less likely to disclose their data sources and methodology. Without checks and balances, the central government might tamper with the NSO's work, preventing the collection of important evidence. When freedom of speech is limited, journalists and academics may be hesitant to question official statistics or pressure for corrections. Information collected by autocrats is a black box: the opaque nature of the data-generation process and the centralized control of information leave little room for subsequent public scrutiny.

Admittedly, autocracies like Azerbaijan, Belarus, Oman, and Tajikistan recognize the instrumental value of data transparency, having requested multiple ROSC Data Modules in the past. Yet their data revisions are strategic and selective — for example, to attract development aid and foreign investment, gain or maintain access to capital markets, monitor internal challenges, and allocate resources to secure elite support (Hollyer, Rosendorff and Vreeland, 2018). Instead of promoting systematic revisions, "informational autocrats" gain legitimacy by disseminating selective information about economic successes while concealing information about their economic failures (Guriev and Treisman, 2019). Even countries that experience a democratic transition are constrained by the quality of the initial data collection. Anticipating future challenges to their power, autocrats might deliberately withhold data or adopt obscure methodology; if so, there is not much their democratically elected successors can do to correct the record.

In contrast, leaders facing stiff political competition and regular turnover are more likely to institutionalize transparency through FOI laws, securing future access to government information (Berliner, 2014). According to Brambor et al. (2020), democracies are better at collecting and processing data not because of political competition, but because of expanded suffrage: states must collect fine-grained information to enable broad popular participation. I propose that the mechanism behind revisions is not related to competition or suffrage but to expertise. Democracies are more likely to publicly release raw data, codebooks, and data collection protocols; retain experienced data managers instead of replacing them with political appointees; establish external advisory committees composed of experts who provide technical guidance and independent oversight; and promote press and academic freedom, allowing journalists and researchers to scrutinize official statistics, identify inconsistencies, and challenge misreporting. All this increases the probability of revisions.

My argument is not that democracies have a higher propensity for releasing "correct" data (though Martínez 2022 and others show that this tends to be the case). Rather, I argue that democracies release data that can be more easily "corrected." This does not happen due to an intrinsic commitment to transparency. Given the high political cost of data corrections, no government — democratic or authoritarian — would like to admit its mistakes. Still, democracies have little choice: their institutions generate incentives to release transparent data ex ante as well as ongoing pressure to revise these data ex post. Through media coverage, public debate, and academic research, experts continuously pressure democratic governments to acknowledge errors and correct official estimates.

Hypothesis 1: Revisions are more likely for data collected by democracies.

Moving to the influence of foreign accountability agents, there is evidence that IMF borrowers are more likely to disseminate data (Hollyer, Rosendorff and Vreeland, 2011). I argue that data disseminated by IMF borrowers are also more prone to revisions. Like other international organizations, the IMF closely scrutinizes its debtors to ensure the disbursed funds meet pre-established targets. IMF loans are attached to conditions that determine whether program benchmarks are being met; if so, subsequent loan tranches can be disbursed. As shown by Kentikelenis and Stubbs (2023), these conditions can include: "develop a monitoring system to verify the quality of the accounting data ... in terms of data consistency and accuracy" (Brazil, 1998); "creation of a fiscal monitoring unit at the Ministry of Finance to prepare, update, report, and analyze fiscal data" (Jordan, 1999); "publish a revision policy and a timetable for compiling and disseminating final national accounts data" (Mozambique, 2004); and "adopt the SNA 93 and publish the 2001 preliminary national accounts on that basis by April 30, 2003" (Senegal, 2003). Such scrutiny creates external pressure to revise and refine macroeconomic data, even for states with limited statistical capacity.

In addition, loan agreements are often attached to technical assistance and capacity-building programs. For instance, Eurostat and the IMF assisted Greece in collecting data, as did Denmark and the IMF in Ghana. Burundi was able to resume data collection after the end of its civil war with support from AFRISTAT and the African Development Bank. These examples have one aspect in common: when Greece, Ghana, and Burundi received technical assistance to fill their statistical gaps, they were all under an IMF loan. As before, the reasoning is not that borrowers report "correct" data; they might still understate their finances to appear poor (Kerner, Jerven and Beatty, 2017). Still, foreign credit in general, and IMF loans in particular, can overcome domestic limitations by providing the resources and motivation for more accurate data collection and regular data reassessment.

Voluntary multilateral initiatives could also make a difference: compliance with the IMF's Special Data Dissemination Standard (SDDS) is associated with more information disclosure (Vadlamannati, Cooray and Brazys, 2018), partly due to technical assistance. Though SDDS compliance or ROSC Data Modules may lead to revisions, I expect the IMF to have more leverage when direct money is on the line.

Hypothesis 2: Revisions are more likely for data collected by IMF borrowers.

Overall, accountability pressures — foreign or domestic — generate incentives to release transparent data that can be revised ex post, increasing the odds of revisions. When collecting data, states can be compelled to adopt strategies that enable future revisions. This initial decision, influenced by experts, academics, journalists, and foreign creditors, is difficult to reverse. Consequently, the posterior adjustment of published data is primarily technical, not political. This expectation is consistent with Croushore and Stark (2003): after the initial year of data release, revisions rarely add new information, instead fine-tuning the original measurements to reduce noise. Posterior data revisions reflect SNA changes, updated base years, corrections to idiosyncratic errors, and other adjustments motivated by international standards rather than domestic politics.

Hypotheses 1 and 2 refer to the *likelihood* of revisions. A separate question is the *magnitude* of revisions — that is, *how much* the estimates for the same country and year change from one data release to another. Here, I expect the role of domestic and international accountability agents to diverge. At the domestic level, since democracies are more transparent and produce information that is constantly scrutinized, this information should be susceptible to revisions of a larger magnitude. Once errors are exposed, democracies cannot easily suppress or moderate the size of revisions, even if these revisions are unfavorable.

At the international level, though, IMF oversight might increase the likelihood of revisions without necessarily leading to larger revisions. IMF borrowers are bound to favor incremental adjustments over drastic corrections that would jeopardize the borrower's credibility by signaling economic mismanagement. While IMF conditions ask borrowers to collect, disseminate, and revise their data, these conditions do not specify *how large* revisions must be. Absent a direct mandate for large-scale revisions, IMF borrowers have at least some control over the revision process; they can make small revisions to satisfy reporting requirements without drastically altering past figures, thus minimizing potential political or economic fallout. As a result, while both democracies and IMF borrowers should be more likely to revise their data, only democracies should report large-magnitude revisions on a systematic basis. Ultimately, the key distinction lies in continuous versus temporary accountability pressures. Democracies face continuous scrutiny, whereas IMF borrowers only face temporary scrutiny tied to the duration of a loan agreement.

Hypothesis 3: Revisions have a larger magnitude for data collected by democracies.

Hypothesis 4: Revisions do not have a larger magnitude for data collected by IMF borrowers.

4 Explaining Data Revisions

4.1 Outcome: Current GDP

As the most widely used source of macroeconomic data in political science (Goes, 2023), the WDI first appeared as a printed annex to the 1978 World Development Report and became a standalone publication in 1997 (World Bank, 2018). In 2018, the World Bank replaced print reports with a data portal that includes the WDI Database Archives, providing 104 electronic WDI releases from 1994 to 2021.[8]

The WDI are typically updated twice a year, around April and September, with additional updates as needed. As the World Bank shifted from print to digital, the number of annual updates increased: there were ten updates in 2017 (in every month except for January and February), compared to only one each year from 1997 to 2004 (always in April).

GDP, the value of all final goods and services produced in a country during a specific period, is "the superstar of indicators" (Hoekstra, 2019, 6) and the most ubiquitous measure of national wealth. In 2020, 205 out of 206 economies surveyed by the IMF compiled and published annual GDP statistics (Baer, Guerreiro and Silungwe, 2022); Eritrea was the lone exception. By comparison, only 109 compiled institutional sector accounts, such as deficit, debt, trade, and foreign direct investment (FDI). Therefore, I focus on revisions to GDP data — specifically, the indicator *GDP in current US dollars* (ID `NY.GDP.MKTP.CD`), the annual "sum of gross value added by all resident producers in the economy." The production approach is the most widely compiled and disseminated approach to GDP estimation (Baer, Guerreiro and Silungwe, 2022, 12). Current GDP enables comparisons across vintages, not across countries or over time, as it does not make PPP or inflation adjustments.[9]

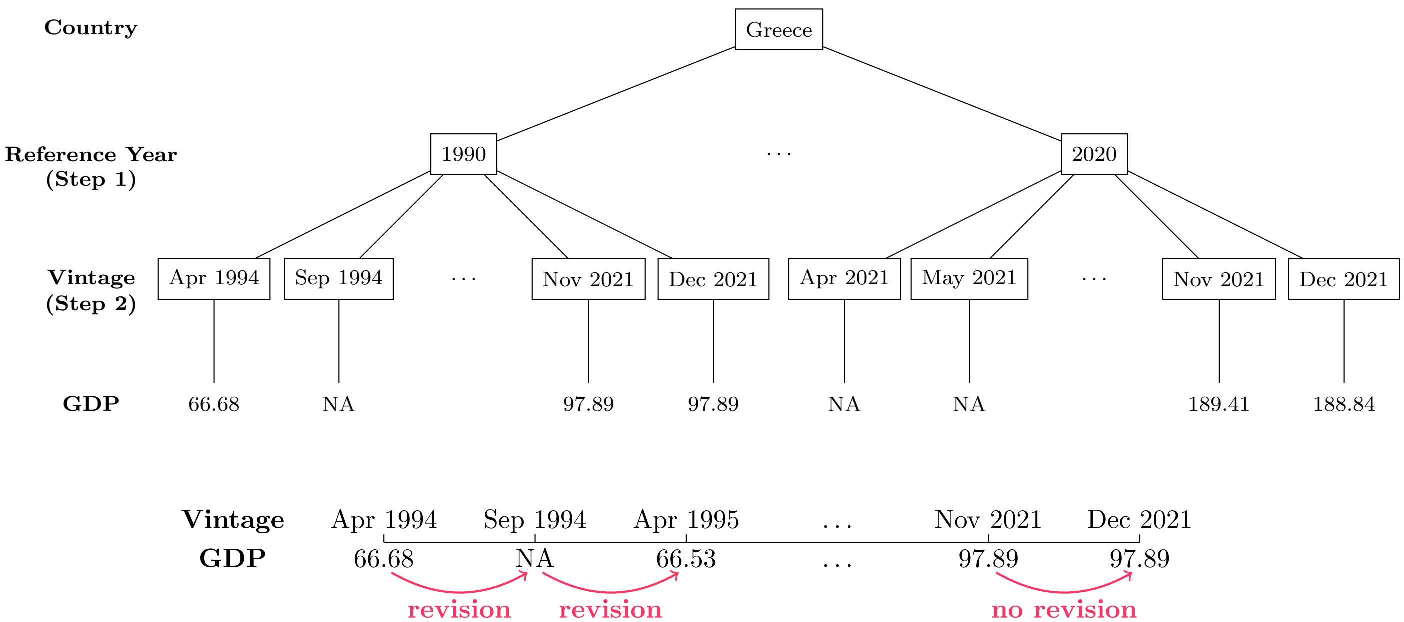
I use data from all vintages from 1994 to 2021 (the reporting year, or step 2), referring to all years from 1990 to 2020 (the reference year, or step 1). For each country and year, I speak of a revision when there is a change in the GDP value reported by two consecutive data releases. This change can reflect an addition, a deletion, or a change of any magnitude, in any direction. As a reminder, 39.8 percent of all economies surveyed by the IMF in 2020 reported annual GDP data within 90 days of the reference period, and 54.4 did so within 91 to 365 days. Correspondingly, the analysis includes all vintages released at least 90 days

[8] Though all releases since 1989 are available, the indicator of interest is missing from all releases before 1994, and no release is available for 1996.

[9] *GDP, PPP (current international $)* (ID `NY.GDP.MKTP.PP.CD`) allows for comparisons across countries, not across vintages, as the PPP conversion factor changes from one ICP round to another. *GDP in constant US dollars* (ID `NY.GDP.MKTP.KD`), calculated using the GDP deflator (the ratio of GDP in current local currency to GDP in constant local currency) to account for inflation, allows for comparisons over time, not across vintages.

after the end of the reference period. For the reference year 2020, say, this comprises all vintages since April 2021. Figure 4 presents the structure of the data, with GDP reported in billions of current US dollars.

Figure 4: Data Structure, Using the Example of Greece



Country	**Reference Year**	**Vintage**	**GDP**	**Revision**
Greece	1990	Apr 1994	66.68	NA
Greece	1990	Sep 1994	NA	1
Greece	1990	Apr 1995	66.53	1
...	...	...	...	...
Greece	1990	Nov 2021	97.89	0
Greece	1990	Dec 2021	97.89	0

As Figure 4 shows, the variable *Revision* takes the value of one if there is any discrepancy between two consecutive data releases (say, Greece in 1990 reported by the April 1994 WDI vs. the September 1995 WDI), and zero otherwise. Revisions occur in 17.17 percent of all observations. If a revision occurs, *Absolute % Change (Log)* quantifies its magnitude, ranging from a 17,488 percent increase (Myanmar, displayed in Figure 3) to a 100 percent reduction (the Democratic Republic of Congo, displayed in Figure 2). The average revision is relatively small, at only 3.6 percent. Since I am interested in the magnitude of revisions rather than their direction, I use the absolute value of percentage changes. To reduce the impact of extreme values, I apply a log transformation, ensuring that the distribution is less skewed.

4.2 Empirical Strategy

I use logistic regressions to estimate the probability of a *Revision*. Conditional on a revision occurring, I use linear regressions to estimate the magnitude of revisions, measured as *Absolute % Change (Log)*. This two-step modeling strategy accounts for the fact that different processes might determine whether a revision occurs and, if so, its magnitude. All models are estimated via maximum likelihood and include random intercepts for country, year, and vintage. This allows the baseline probability of *Revision* or the baseline *Absolute % Change (Log)* to vary across countries, years, and vintages, instead of assuming a single baseline for all observations.

For each outcome, I estimate two models: one that includes all vintages and another that focuses on each year's main scheduled update, in April,[10] to avoid overcounting similar data points and remove noise (after all, the information provided by the April 2021 WDI is likely similar to that provided by the May 2021 WDI). This also ensures that recent years (with more releases) do not receive undue weight relative to older years (with fewer releases).

My models are deliberately parsimonious, with few independent variables. This is because including too many predictors can lead to overfitting and multicollinearity: the model may capture noise instead of the true underlying relationships, and high correlations among predictors may destabilize coefficient estimates. In the appendix, I report the results of regularized regressions with fixed effects and a more extensive list of predictors. Regularized regressions mitigate overfitting and multicollinearity by performing variable selection: LASSO shrinks some coefficients to exactly zero, ridge regression shrinks them toward zero, and elastic net combines these approaches to select groups of related predictors. Still, these

[10]For 2020, this separate analysis uses August instead of April; that year, not a single country updated its data before August, presumably due to delays in data collection caused by the COVID-19 pandemic. The same applies to 2021, for which I use the September release.

methods do not directly allow for hypothesis testing, are harder to interpret, and lead to nearly identical conclusions, hence my choice to present simpler models in the main text.

4.3 Independent Variables

First, I assess a country's baseline ability to collect high-quality data, denoted as state capacity. Hanson and Sigman (2021) use Bayesian latent variable analysis to combine 21 indicators of extractive, coercive, and administrative capacity into a single index. Out of the 21 components, the V-Dem index *Rigorous Public Administration* has the most extensive coverage, which is why I use it in the main analysis. It ranges from 0 (the law is not respected by public officials) to 4 (the law is generally fully respected by public officials).

Conditional on a country's ability to collect high-quality data, I argue that domestic accountability agents increase the likelihood and magnitude of revisions. To quantify the existence of domestic accountability agents, V-Dem's *Polyarchy* index (Coppedge et al., 2023) measures the quality of electoral democracy, including extensive suffrage, fair elections, freedom of expression, and access to information (in an ordinal scale from 0 to 1). Larger values reflect more democratic regimes, which should be associated with more domestic accountability pressures — and thus higher odds of revision.

What is it about democracies that makes their data more susceptible to public scrutiny? My proposed pathway is *Freedom of Academic Expression*: in the absence of censorship and intimidation, experts can demand transparent data practices. This V-Dem index ranges from 0 (not respected by public authorities) to 4 (fully respected by public authorities). Alternative pathways include the *Political Corruption Index* (which ranges from 0 to 1, with larger values indicating more corruption) and the population share with *Suffrage*, both reported by V-Dem. In addition, I borrow three measures from Berliner (2014). *New Democracy* takes the value of one in the first five years after a democratic transition, using V-Dem data. *Opposition Strength* measures the vote share of the largest opposition party in the most recent legislative election, whereas *Turnover Frequency* indicates the number of changes in party control of the executive in the preceding five years (both calculated using data from the Database of Political Institutions, Cruz, Keefer and Scartascini 2021).

In terms of foreign accountability agents, I argue that IMF borrowers tend to be scrutinized more closely, creating external pressure to revise and refine macroeconomic data (without a corresponding effect on the magnitude of revisions). Therefore, models control for *IMF Program* participation (Kentikelenis and Stubbs, 2023) in addition to *SDDS Compliance*, a voluntary initiative associated with more data disclosure (Vadlamannati, Cooray and Brazys, 2018). Other important events may prompt data revisions: a financial crisis, natural disaster, or armed conflict (using data from Nguyen, Castro and Wood 2022, the Centre for Research on the Epidemiology of Disasters 2020, and Gleditsch et al. 2002, respectively).

Few measures of data quality are available for the entire period.[11] As an imperfect proxy, I use the *Difference Between Official and Alternative Exchange Rates*, scaled to have a mean of 0 and a standard deviation of 1. According to the WDI Metadata, "dollar figures for GDP are converted from domestic currencies using single year official exchange rates." However, if "the official exchange rate does not reflect the rate effectively applied to actual foreign exchange transactions," the World Bank applies an alternative exchange rate. A large discrepancy can signal exchange rate manipulation, dual exchange rates, or price controls. Whatever the cause, an official exchange rate that does not align with real economic transactions suggests the existence of underlying data problems.

Three final variables capture the effect of technical or methodological adjustments after the original data release. *SNA Change* takes the value of one for years in which countries adopted the SNA global standardization framework (partly or entirely)[12] or updated the SNA version in use, based on information provided by the UN National Accounts Statistics (complemented by WDI Metadata and IMF International Financial Statistics). The SNA

[11] For example, the World Bank's Statistical Capacity Indicators provide information about the Balance of Payments manual in use, but only for developing countries and only after 2004.

[12] Communist countries began to adopt the SNA around 1993. The one exception is North Korea, which still uses a Marxism-inspired alternative, the Material Product System (Herrera 2010, 23n8; van Heijster and DeRock 2022, 84n1). North Korea is the only country missing from all WDI releases and thus excluded from the analysis.

was initially published in 1953 and updated in 1968, 1993, and 2008. It can take years — even decades — for countries to adopt the latest SNA, but when they do, they often readjust previously published data to meet the new methodological refinements. For this reason, SNA updates likely drive ex post data revisions. *Data Management Error* takes the value of one for instances illustrated by Figures 2 and 3 or listed in the WDI errata (World Bank 2023; see appendix for an in-depth discussion). This variable provides a conservative estimate of idiosyncratic errors, which are likely much more widespread. Lastly, *Difference Between Vintage and Year* tallies the number of years elapsed between steps 1 and 2; if information-based revisions are most common up to one year after the initial data release, as Croushore and Stark (2003) find, the higher values of this variable should be associated with fewer revisions.

Table 1: Summary of Independent Variables

Variable	**Underlying Concept**	**Period of Interest**
Rigorous Public Administration	Baseline State Capacity	Reference Year
Polyarchy *and components:*	Baseline Domestic Accountability Agents	Reference Year
– *Freedom of Academic Expression*		
– *Political Corruption Index*		
– *Suffrage*		
– *New Democracy*		
– *Opposition Strength*		
– *Turnover Frequency*		
IMF Program	Foreign Accountability Agents	Reference Year, Vintage
SDDS Compliance	Foreign Accountability Agents	Reference Year, Vintage
Financial Crisis	Exceptional Events	Reference Year, Vintage
Natural Disaster	Exceptional Events	Reference Year, Vintage
Armed Conflict	Exceptional Events	Reference Year, Vintage
Diff. Between Official and Alt. XR	Baseline Data Quality	Reference Year
SNA Change	Vintage-Specific Methodological Changes	Vintage
Data Management Error	Vintage-Specific Idiosyncratic Errors	Vintage
Diff. Between Vintage and Year	Time Since Initial Data Dissemination	Vintage – Reference Year

Table 1 summarizes all independent variables and their underlying concepts. As this table shows, some factors (like IMF programs and financial crises) may influence both initial reporting practices and subsequent revisions, so they are included for both the reference year and the vintage. Others (like state capacity and regime type) change slowly, hence their inclusion only for the reference year to avoid multicollinearity.[13] And some variables (like SNA change and idiosyncratic errors) are vintage-specific. Other than *Data Management Error* and *Difference Between Vintage and Year*, all variables are lagged by one year to avoid simultaneity bias. I do not include economic indicators like foreign aid, natural resource rents, or trade dependence, which are closely related to GDP and likely revised just as frequently.

5 Results

5.1 The Likelihood of Revisions

Table 2 presents the results of four logistic regressions with the dichotomous outcome *Revision*. Model 1 draws from all vintages released between 1994 and 2020. Due to the large sample size, even small effects can achieve statistical significance. Model 2 focuses on each year's main scheduled update, reducing noise and allowing for a more precise estimation of effects. As both models show, conditional on state capacity (*Rigorous Public Administration*), *Polyarchy* has a positive and significant effect on the outcome: information collected by more democratic states is more likely to be revised. These results are consistent with Hypothesis 1: democratic institutions encourage both the initial release of transparent statistics *and* the sustained pressure to later revise them. In line with Hypothesis 2, states under an IMF program ahead of data collection (step 1) produce data that are 7 to 8 percent more likely to be subsequently revised.

Once data are released, Model 2 shows that IMF programs, SDDS compliance, armed conflicts, and natural disasters do not systematically drive further adjustments (step 2), whereas financial crises have only a weakly significant effect ($p < 0.1$). Rather, time matters: for every additional year between the vintage year and the original data collection, a data point is 8 percent less likely to be revised. Older data undergo fewer revisions, denoting that

[13] For example, *Rigorous Public Administration* in step 1 and *Rigorous Public Administration* in step 2 are correlated at $\rho = 0.8325, p = 0.000$. *Polyarchy* in step 1 and *Polyarchy* in step 2 are correlated at $\rho = 0.879$ ($p = 0.000$).

estimates stabilize after an initial period of frequent adjustment to incorporate new data. Finally, according to Model 2, revisions are 60 percent more likely for vintages following a SNA change, an effect that is statistically significant. For instance, Greece updated its SNA in 2014; consistent with the pattern identified in Table 2, revisions for Greece increased in the 2015 vintages. Ex post data revisions are largely driven by economic shocks or changes in statistical methodology, rather than isolated political, natural, or economic events.

In lieu of a democracy index, Models 3 and 4 examine individual aspects of democracy likely to drive data revisions. Of these, only *Freedom of Academic Expression* has a positive and significant effect. Brambor et al. (2020) show that *Suffrage* is a proxy for information capacity: the more individuals are eligible to vote, the more a state must keep track of them. However, according to both models, the need and ability to collect information *reduce* the odds of data revisions. Widespread suffrage, frequent turnover, and a strong opposition allow citizens to hold governments accountable, but scrutinizing technical data depends more on specialized knowledge than broad electoral dynamics. Revisions are most likely when journalists, academics, and other experts have the freedom to demand data transparency and push for data corrections. The same holds when states face scrutiny from foreign experts, as indicated by IMF program participation.

In the appendix, several robustness checks corroborate these findings. First, I estimate full models with alternative measures of regime type (*Polity 2* and *Freedom House*), ensuring that the observed effects are not driven by a particular measure of democracy. Second, I replace *Rigorous Public Administration* with Hanson and Sigman's *State Capacity* index and other measures of bureaucratic quality (from the International Country Risk Guide, ICRG). Although these alternative measures cover fewer countries and years, the results are robust to their inclusion. The results are also robust to replacing random effects with country, year, and vintage fixed effects, though these alternative models are less efficient due to the large number of additional parameters.

Table 2: Predicting the Likelihood of Data Revisions

	Dependent Variable: Revision = 1			
	(1)	(2)	(3)	(4)
	All Vintages	*Only Main Vintages*	*All Vintages*	*Only Main Vintages*
Rigorous Public Administration	0.07^{***}	0.08^{***}	0.06^{***}	0.07^{***}
	(0.02)	(0.02)	(0.02)	(0.03)
Polyarchy	0.71^{***}	1.06^{***}		
	(0.08)	(0.11)		
Freedom of Academic Expression			0.18^{***}	0.21^{***}
			(0.01)	(0.02)
Suffrage			-0.62^{***}	-0.27^{**}
			(0.09)	(0.13)
New Democracy, Prev. 5 Yrs			-0.04	-0.08^{**}
			(0.03)	(0.04)
Opposition Strength			-0.00	-0.00
			(0.00)	(0.00)
Turnover Frequency, Prev. 5 Yrs			0.01	0.02
			(0.01)	(0.02)
Political Corruption Index			0.10	-0.15
			(0.09)	(0.12)
IMF Program (Step 1)	0.07^{***}	0.08^{***}	0.05^{***}	0.06^{**}
	(0.02)	(0.02)	(0.02)	(0.02)
IMF Program (Step 2)	-0.07^{***}	-0.00	-0.07^{***}	0.01
	(0.02)	(0.03)	(0.02)	(0.03)
SDDS Compliance (Step 1)	-0.27^{***}	-0.32^{***}	-0.27^{***}	-0.33^{***}
	(0.02)	(0.03)	(0.02)	(0.04)
SDDS Compliance (Step 2)	0.16^{***}	0.05	0.20^{***}	0.10^{**}
	(0.04)	(0.04)	(0.04)	(0.04)
Financial Crisis (Step 1)	0.05^{***}	0.05^{*}	0.07^{***}	0.06^{**}
	(0.02)	(0.02)	(0.02)	(0.03)
Financial Crisis (Step 2)	0.07^{***}	0.05^{*}	0.06^{***}	0.04
	(0.02)	(0.03)	(0.02)	(0.03)
Natural Disaster (Step 1)	0.05^{***}	0.05^{**}	0.06^{***}	0.05^{**}
	(0.02)	(0.02)	(0.02)	(0.02)
Natural Disaster (Step 2)	0.02	0.00	0.03	0.00
	(0.02)	(0.02)	(0.02)	(0.02)
Armed Conflict (Step 1)	-0.02	-0.01	0.00	0.02
	(0.03)	(0.04)	(0.03)	(0.04)
Armed Conflict (Step 2)	-0.08^{***}	-0.06	-0.08^{**}	-0.06
	(0.03)	(0.04)	(0.03)	(0.04)
Diff. Between Official and Alt. XR	-0.01^{**}	-0.00	-0.01^{**}	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
SNA Change	0.20^{***}	0.47^{***}	0.21^{***}	0.47^{***}
	(0.03)	(0.04)	(0.03)	(0.04)
Data Management Error	6.87^{***}	-0.18	6.91^{***}	-0.47
	(0.62)	(0.90)	(0.62)	(0.94)
Diff. Between Vintage and Year	-0.11^{***}	-0.08^{***}	-0.11^{***}	-0.08^{***}
	(0.01)	(0.01)	(0.01)	(0.01)
Intercept	-3.49^{***}	0.50^{*}	-3.04^{***}	0.83^{***}
	(0.48)	(0.26)	(0.49)	(0.30)
Log Likelihood	$-87,123.61$	$-41,793.10$	$-84,814.20$	$-40,583.76$
Observations	397,803	78,178	387,212	76,092
Number of Countries	170	170	167	167
Number of Years	31	31	31	31
Number of Vintages	103	26	103	26
Variance: Countries (Intercept)	0.29	0.38	0.31	0.37
Variance: Years (Intercept)	0.14	0.26	0.14	0.27
Variance: Vintages (Intercept)	19.95	1.31	19.88	1.35

This table presents the results of four logistic regressions with random intercepts for country, year, and vintage. Step 1 is the reference year, the year of data collection. Step 2 is the vintage year, the year of data revision. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

5.2 The Magnitude of Revisions

Having shown that regime type and IMF loans affect the *likelihood* of revisions, I now turn to their *magnitude*. Table 3 restricts the sample to instances when *Revision = 1*, hence the smaller number of observations. According to Models 1 and 2, an increase in democracy levels is associated with a significant increase in the magnitude of *Absolute % Change (Log)*. The implication is not that democracies produce intrinsically "worse" data requiring larger revisions. Data collected by autocracies may need revisions that are just as large. However, autocracies tend to withhold information or only disclose it selectively, so experts cannot identify errors and observe the resulting revisions in the first place. In contrast, the transparent data practices in democratic systems increase the pressure to acknowledge and correct potential inaccuracies, even if the resulting large-scale revisions reflect poorly on the government.

As expected, IMF program participation — whether at the time of original data collection (step 1) or during posterior data revision (step 2) — has a different effect. IMF programs might encourage more cautious revisions (that is, revisions of *smaller* magnitude), though these effects are not consistently significant across both models. Despite increasing the likelihood of revisions, IMF oversight does not necessarily lead to large-scale updates. Borrowers prefer to make incremental revisions that comply with reporting requirements without jeopardizing their economic stability or political reputation.

Finally, Models 3 and 4 decompose the effects of regime type on magnitude of revisions, finding again that *Freedom of Academic Expression* is the one aspect of democratic governance that matters most. Information collected in an environment without censorship or intimidation is most likely to be revised, and such revisions tend to have a larger magnitude. Relatedly, a higher *Political Corruption Index* is associated with revisions of significantly smaller magnitude, underscoring that large-scale revisions are most widespread in contexts with more transparent governance. Of note, *Data Management Error* drops out of Model 4: focusing only on revisions in the main vintages results in too few observations that qualify

Table 3: Predicting the Magnitude of Data Revisions

	Dependent Variable: Abs. % Change (Log)			
	(1)	(2)	(3)	(4)
	All Vintages	*Only Main Vintages*	*All Vintages*	*Only Main Vintages*
Rigorous Public Administration	0.09	0.06	-0.01	-0.11
	(0.08)	(0.10)	(0.08)	(0.10)
Polyarchy	1.34^{***}	1.86^{***}		
	(0.34)	(0.42)		
Freedom of Academic Expression			0.23^{***}	0.30^{***}
			(0.06)	(0.08)
Suffrage			0.13	0.70
			(0.40)	(0.49)
New Democracy, Prev. 5 Yrs			-0.16	-0.17
			(0.12)	(0.15)
Opposition Strength			-0.00^{**}	-0.01^{**}
			(0.00)	(0.00)
Turnover Frequency, Prev. 5 Yrs			0.06	0.06
			(0.06)	(0.07)
Political Corruption Index			-1.51^{***}	-2.25^{***}
			(0.38)	(0.45)
IMF Program (Step 1)	-0.13^{*}	-0.13	-0.07	-0.03
	(0.08)	(0.10)	(0.08)	(0.10)
IMF Program (Step 2)	-0.37^{***}	-0.33^{***}	-0.38^{***}	-0.29^{***}
	(0.09)	(0.11)	(0.09)	(0.11)
SDDS Compliance (Step 1)	-0.06	-0.38^{***}	-0.16	-0.51^{***}
	(0.11)	(0.14)	(0.11)	(0.14)
SDDS Compliance (Step 2)	0.54^{***}	0.89^{***}	0.56^{***}	0.88^{***}
	(0.15)	(0.17)	(0.15)	(0.17)
Financial Crisis (Step 1)	-0.17^{**}	-0.19^{*}	-0.15^{*}	-0.19^{*}
	(0.08)	(0.10)	(0.08)	(0.10)
Financial Crisis (Step 2)	0.55^{***}	-0.16	0.58^{***}	-0.17
	(0.09)	(0.11)	(0.09)	(0.11)
Natural Disaster (Step 1)	-0.12^{*}	-0.11	-0.09	-0.05
	(0.07)	(0.09)	(0.07)	(0.09)
Natural Disaster (Step 2)	-0.11	0.23^{**}	-0.10	0.24^{**}
	(0.08)	(0.10)	(0.08)	(0.10)
Armed Conflict (Step 1)	-0.36^{***}	-0.68^{***}	-0.32^{**}	-0.64^{***}
	(0.13)	(0.17)	(0.14)	(0.17)
Armed Conflict (Step 2)	0.27^{**}	0.56^{***}	0.33^{**}	0.61^{***}
	(0.14)	(0.16)	(0.14)	(0.16)
Diff. Between Official and Alt. XR	0.00	-0.06^{*}	0.00	-0.07^{*}
	(0.03)	(0.04)	(0.03)	(0.04)
SNA Change	1.68^{***}	2.62^{***}	1.73^{***}	2.63^{***}
	(0.12)	(0.15)	(0.12)	(0.15)
Data Management Error	12.14^{***}	15.11^{**}	12.09^{***}	
	(1.51)	(6.86)	(1.55)	
Diff. Between Vintage and Year	-0.24^{***}	-0.34^{***}	-0.24^{***}	-0.35^{***}
	(0.01)	(0.01)	(0.01)	(0.01)
Intercept	-5.91^{***}	-4.09^{***}	-4.93^{***}	-3.06^{***}
	(0.64)	(0.73)	(0.78)	(0.90)
Log Likelihood	$-210,649.32$	$-134,357.76$	$-206,053.37$	$-131,434.80$
Observations	62,916	40,119	61,551	39,263
Number of Countries	170	170	167	166
Number of Years	31	30	31	30
Number of Vintages	78	26	78	26
Variance: Countries (Intercept)	2.88	3.16	2.73	3.04
Variance: Years (Intercept)	0.15	0.19	0.16	0.18
Variance: Vintages (Intercept)	25.02	11.46	24.74	11.18

This table presents the results of four linear regressions with random intercepts for country, year, and vintage. Step 1 is the reference year, the year of data collection. Step 2 is the vintage year, the year of data revision. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

as coding errors.

As the appendix shows, these results are robust to other aggregate measures of regime type or alternative measures of state capacity. While *Absolute % Change (Log)* is a broad proxy for the magnitude of revisions, additional models in the appendix examine the *direction* of revisions. Compared to downward revisions, upward revisions become significantly more likely as the *Polyarchy* score increases. This suggests that initial reporting in democracies is more cautious, leading to understated estimates. As more complete information becomes available, subsequent updates result in larger revisions that tend to adjust figures upward.

6 Discussion and Conclusion

This study shows that revisions are more likely for data collected by democracies or IMF borrowers. Revisions are also systematically larger in democracies, but not during IMF programs. While the analysis is restricted to GDP data, I expect these predictions to hold for other macroeconomic indicators, like FDI, trade, deficit, and debt, which are all calculated relative to a country's GDP. The analysis uses the WDI due to their frequent updates and ubiquity in political science research, but the argument likely travels to other data sources. Since PWT, the Maddison Project Database, and others are not updated as frequently, they do not allow for as comprehensive a test. For example, between 1994 and 2024, there have been eleven PWT releases and five releases of the Maddison Project Database (including the original 2010 database produced by Angus Maddison), compared to over 100 WDI vintages available in the WDI Database Archives. Still, future research can employ other indicators and sources to validate the generalizability of my results, exploring variations in the frequency and magnitude of revisions across other compilation methodologies and institutional contexts.

These results have three important implications. First, data transparency generates a trade-off between validity and reliability. Autocracies release information with lower validity but higher reliability: there are many incentives to collect biased data and publish imprecise estimates, but once these estimates are published, they are set in stone and difficult to revisit. Democracies release information with higher validity but lower reliability: there are fewer political incentives to hide or lie about the data, but also greater incentives and institutional capacity to scrutinize — and, if necessary, alter — these data. Frequent revisions are not a problem; in fact, the international community *encourages* frequent revisions to increase the validity of the data. The assumption is that the initial measurement is fairly accurate, with each additional revision providing marginal accuracy gains. Yet this is not always the case. Large-magnitude revisions might signal that the initial measurement was inaccurate, reducing the perceived reliability of the data.

From a statistical perspective, correcting erroneous information is better than not correcting it. The Bureau of Labor Statistics no doubt made the right choice in revising its employment data as part of a scheduled annual update in August 2024. Yet the political backlash spearheaded by President Trump highlights a broader challenge: revisions can undermine public trust in official statistics, particularly in contexts of low data literacy and widespread misinformation. For this reason, transparency in revision processes must be accompanied by public communication strategies that frame revisions as a sign of statistical rigor. Revisions are not typically a sign of mismanagement; if anything, they correct earlier instances of mismanagement, as in Greece.

Second, what matters most is the environment under which the initial data were collected. Revisions might come later, driven by technical improvements yet constrained by the initial conditions. Subsequent technical updates may fine-tune the data and reduce noise, as Croushore and Stark (2003) show, but no amount of posterior adjustments can fully compensate for flaws in the original data collection. Investments in statistical capacity-building must begin at the initial stage of data collection; long-term data quality hinges on these early decisions. Ultimately, collecting, publishing, and revising data are political choices. If countries are willing to make these choices, international initiatives like the IMF's ROSC

Data Modules or the World Bank's former Trust Fund for Statistical Capacity Building can help mitigate data issues and enhance a country's statistical capacity. Yet these enhancements benefit future data collection and have limited effect on previously collected data. To reiterate: future revisions can only occur if the initial data collection process is structured to accommodate them. Policymakers and researchers must account for historical limitations in data quality, as present-day improvements to statistical capacity do not automatically fix past data. Cuts in statistical capacity — like those seen in the US under President Trump — will have long-lasting, irreversible consequences.

Finally, my results speak to existing work on the discrepancies between WDI and PWT vintages, or even between different vintages of the same data source (Goes, 2023; Johnson et al., 2013; Croushore and Stark, 2003). Replacing one source or vintage with another can significantly alter published research findings. At the same time, these existing studies have paid particular attention to non-democracies (Goes, 2023) or African states (Johnson et al., 2013). My results suggest that industrial democracies also warrant scrutiny: researchers and policymakers working with, say, EU members must be careful about their estimates, which are more fluid than one might think. In contrast, those working with less democratic states have different concerns: they are likely working with inaccurate information, but this information remains consistent across different releases of the same data. Ideally, countries would collect high-quality data at the outset, but researchers and policymakers must have realistic expectations for the data they use. As McMann et al. (2022) show, discussions about data quality often focus on validity and neglect reliability. Yet it is just as important to update, revise, improve, and compare estimates as better information and methodology become available.

References

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.

African Development Bank. 2013. *Situational Analysis of the Reliability of Economic Statistics in Africa: Special Focus on GDP Measurement.*
URL: *https://www.afdb.org/sites/default/files/documents/publications/economic_brief_-_situational_analysis_of_the_reliability_of_economic_statistics_in_africa-_special_focus_on_gdp_measurement.pdf*

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.

Amin Gutiérrez de Piñeres, Sheila. 2006. "What a Difference a Source Makes! An Analysis of Export Data." *Applied Economics Letters* 13(1):35–39.

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.

Baer, Andrew, Vanda Guerreiro and Anthony Silungwe. 2022. "2020 Global Stocktaking of National Accounts Statistics: Availability for Policy and Surveillance." *IMF Working Papers* (29):1–25.

Berliner, Daniel. 2014. "The Political Origins of Transparency." *Journal of Politics* 76(2):479–491.

Bolt, Jutta and Jan Luiten van Zanden. 2024. "Maddison-Style Estimates of the Evolution of the World Economy: A New 2023 Update." *Journal of Economic Surveys* pp. 1–41.

Brambor, Thomas, Agustín Goenaga, Johannes Lindvall and Jan Teorell. 2020. "The Lay of the Land: Information Capacity and the Modern State." *Comparative Political Studies* 53(2):175–213.

Carson, Carol S., Sarmad Khawaja and Thomas K. Morrison. 2004. "Revisions Policy for Official Statistics: A Matter of Governance." *IMF Working Papers* 4(87):1–35.

Centre for Research on the Epidemiology of Disasters. 2020. *EM-DAT: The International Disaster Database.*
URL: *https://public.emdat.be*

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Agnes Cornell, Sirianne Dahlum, Haakon Gjerløw, Adam Glynn, Allen Hicken, Joshua Krusell, Anna Lührmann, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Juraj Medzihorsky, Moa Olin, Pamela Paxton, Daniel Pemstein, Josefine Pernes, Johannes von Römer, Brigitte Seim, Rachel Sigman, Jeffrey Staton, Natalia Stepanova, Aksel Sundström, Eitan Tzelgov, Yi-ting Wang, Tore Wig, Steven Wilson and Daniel Ziblatt. 2023. *V-Dem Country-Year Dataset v13.*

Coyle, Diane. 2014. *GDP: A Brief But Affectionate History.* Princeton and Oxford: Princeton University Press.

Croushore, Dean and Tom Stark. 2003. "A Real-Time Data Set for Macroeconomists: Does the Data Vintage Matter?" *Review of Economics and Statistics* 85(3):605–617.

Cruz, Cesi, Philip Keefer and Carlos Scartascini. 2021. *Database of Political Institutions 2020.*

Deaton, Angus and Bettina Aten. 2017. "Trying to Understand the PPPs in ICP 2011: Why Are the Results So Different?" *American Economic Journal: Macroeconomics* 9(1):243–64.

DeRock, Daniel. 2021. "Hidden in Plain Sight: Unpaid Household Services and the Politics of GDP Measurement." *New Political Economy* 26(1):20–35.

Devarajan, Shantayanan. 2013. "Africa's Statistical Tragedy." *Review of Income and Wealth* 59(S1):9–15.

European Commission. 2010. *Eurostat Report on Greek Government Deficit and Debt Statistics- January 2010.*

URL: *http://ec.europa.eu/eurostat/documents/4187653/6404656/COM_2010_report_greek/c8523cfa-d3c1-4954-8ea1-64bb11e59b3a*

Fariss, Christopher J., Therese Anders, Jonathan N. Markowitz and Miriam Barnum. 2022. "New Estimates of Over 500 Years of Historic GDP and Population Data." *Journal of Conflict Resolution* 66(3):553–591.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg and Håvard Strand. 2002. "Armed Conflict 1946–2001: A New Dataset." *Journal of Peace Research* 39(5):615–637.

Goes, Iasmin. 2023. "New Data, New Results? How Data Vintaging Affects the Replicability of Research." *Research and Politics* (April-June):1–13.

Guriev, Sergei and Daniel Treisman. 2019. "Informational Autocrats." *Journal of Economic Perspectives* 33(4):100–127.

Hanson, Jonathan K. and Rachel Sigman. 2021. "Leviathan's Latent Dimensions: Measuring State Capacity for Comparative Political Research." *Journal of Politics* 83(4):1–16.

Herrera, Yoshiko M. 2010. *Mirrors of the Economy: National Accounts and International Norms in Russia and Beyond*. Ithaca: Cornell University Press.

Hoekstra, Rutger. 2019. *Replacing GDP by 2030: Towards a Common Language for the Well-Being and Sustainability Community*. Cambridge: Cambridge University Press.

Hollyer, James R., B. Peter Rosendorff and James Raymond Vreeland. 2011. "Democracy and Transparency." *Journal of Politics* 73(4):1191–1205.

Hollyer, James R., Peter Rosendorff and James Raymond Vreeland. 2018. *Information, Democracy, and Autocracy: Economic Transparency and Political (In)Stability.* Cambridge: Cambridge University Press.

Islam, Roumeen. 2006. "Does More Transparency Go Along With Better Governance?" *Economics and Politics* 18(2):121–167.

Jerven, Morten and Magnus Ebo Duncan. 2012. "Revising GDP Estimates in Sub-Saharan Africa: Lessons from Ghana." *African Statistical Journal* 15:13–24.

Johnson, Simon, William Larson, Chris Papageorgiou and Arvind Subramanian. 2013. "Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates." *Journal of Monetary Economics* 60(2):255–274.

Kentikelenis, Alexandros and Thomas Stubbs. 2023. *A Thousand Cuts: Social Protection in the Age of Austerity.* Oxford: Oxford University Press.

Kerner, Andrew. 2014. "What We Talk About When We Talk About Foreign Direct Investment." *International Studies Quarterly* 58(4):804–815.

Kerner, Andrew, Morten Jerven and Alison Beatty. 2017. "Does It Pay to Be Poor? Testing for Systematically Underreported GNI Estimates." *Review of International Organizations* 12(1):1–38.

Linsi, Lukas, Brian Burgoon and Daniel Mügge. 2023. "The Problem with Trade Measurement in IR." *International Studies Quarterly* 67(2):1–18.

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.

Martínez, Luis R. 2022. "How Much Should We Trust the Dictator's GDP Growth Estimates?" *Journal of Political Economy* 130(10):2731–2769.

McMann, Kelly, Daniel Pemstein, Brigitte Seim, Jan Teorell and Staffan Lindberg. 2022. "Assessing Data Quality: An Approach and an Application." *Political Analysis* 30(3):426–449.

Mejía Guerra, José Antonio, Christian Schuster, Magdalena Rojas Wettig, Kim Sass Mikkelsen and Jan Meyer-Sahling. 2023. *Making National Statistical Offices Work Better: Evidence from a Survey of 13,300 National Statistical Office (NSO) Employees in 14 Latin American and Caribbean Countries.* Washington, D.C.: Inter-American Development Bank.

Michaelowa, Axel and Katharina Michaelowa. 2011. "Coding Error or Statistical Embellishment? The Political Economy of Reporting Climate Aid." *World Development* 39(11):2010–2020.

Mügge, Daniel. 2022. "Economic Statistics as Political Artefacts." *Review of International Political Economy* 29(1):1–22.

Nguyen, Thanh Cong, Vítor Castro and Justine Wood. 2022. "A New Comprehensive Database of Financial Crises: Identification, Frequency, and Duration." *Economic Modelling* 108:105770.

Olinto Ramos, Roberto, Gonzalo Pastor and Lisbeth Rivas. 2008. "Latin America: Highlights from the Implementation of the System of National Accounts 1993 (1993 SNA)." *IMF Working Paper* 8(239):1–51.

Pardo, Claudio A. 2011. *2011 Review of the Standards and Codes Initiative.* Washington, D.C.: International Monetary Fund.

Pellechio, Anthony and John Cady. 2006. "Differences in IMF Data: Incidence and Implications." *IMF Staff Papers* 53(2):326–349.

Polyak, Palma. 2023. "Jobs and Fiction: Identifying the Effect of Corporate Tax Avoidance Inflating Export Measures in Ireland." *Journal of European Public Policy* 30(10):2143–2164.

Ram, Rati and Secil Ural. 2014. "Comparison of GDP Per Capita Data in Penn World Table and World Development Indicators." *Social Indicators Research* 116(2):639–646.

United Nations Economic Commission for Africa. 2005. *Assessment of the Implementation of the 1993 System of National Accounts in Africa.*
URL: *https://repository.uneca.org/handle/10855/33251*

Vadlamannati, Krishna Chaitanya, Arusha Cooray and Samuel Brazys. 2018. "Nothing to Hide: Commitment to, Compliance With, and Impact of the Special Data Dissemination Standard." *Economics and Politics* 30(1):55–77.

van Heijster, Joan and Daniel DeRock. 2022. "How GDP Spread to China: The Experimental Diffusion of Macroeconomic Measurement." *Review of International Political Economy* 29(1):65–87.

Wallace, Jeremy L. 2014. "Juking the Stats? Authoritarian Information Problems in China." *British Journal of Political Science* 46(1):11–29.

Weikmans, Romain and J. Timmons Roberts. 2019. "The International Climate Finance Accounting Muddle: Is There Hope on the Horizon?" *Climate and Development* 11(2):97–111.

Williams, Andrew. 2009. "On the Release of Information by Governments: Causes and Consequences." *Journal of Development Economics* 89(1):124–138.

World Bank. 2018. *World Development Indicators: The Story.*

URL: *https://datatopics.worldbank.org/world-development-indicators/stories/world-development-indicators-the-story.html*

World Bank. 2023. *Data Updates and Errata.*

URL: *https://datahelpdesk.worldbank.org/knowledgebase/articles/906522-data-updates-and-errata*

Appendix for
When Countries Revise Their Data

March 2025

Contents

A Countries Included in the Main Analysis

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

B Variables Included in the Main Analysis

Table B.1: Variables Included in the Main Analysis

Variable	Description	Coverage	Source
Absolute % Change	Percentage change in the GDP value reported by two consecutive data releases for the same country-year pair	1990–2021	WDI
Armed Conflict	Was there an armed conflict? Yes = 1	1990–2021	Gleditsch et al. (2002)
Data Management Error	Coded 1 for the following observations: Armenia, 1992, April 1998 WDI; China, 2007 and 2008, May 2009 WDI; Democratic Republic of the Congo, 1990, July 2016 to April 2017 WDI; Estonia, 1995, April 1997 WDI; Myanmar, all years, December 2021 WDI	1990–2021	Own Coding, based on World Bank (2023)
Diff. Between Official and Alt. XR	Difference between the official exchange rate, `PA.NUS.FCRF`, and the DEC alternative conversion factor, `PA.NUS.ATLS` (both in LCU per US$)	1990–2021	WDI
Diff. Between Vintage and Year	Number of years elapsed between vintage and data collection	1990–2021	WDI

Financial Crisis	Was there a banking, currency, or debt crisis? Yes = 1	1990–2019	Nguyen, Castro and Wood (2022)
Freedom of Academic Expression	Is there academic freedom and freedom of cultural expression related to political issues? (not = 0, weakly = 1, somewhat = 2, mostly = 3, fully = 4)	1990–2021	Coppedge et al. (2023)
IMF Program	Was there an IMF program? Yes = 1	1990–2021	Kentikelenis and Stubbs (2023)
Natural Disaster	Was there a biological (epidemic), climatological (drought, wildfire), meteorological (storm, extreme temperature), hydrological (flood, landslide), or geophysical (earthquake, volcanic activity) disaster? Yes = 1	1990–2021	Centre for Research on the Epidemiology of Disasters (2020)
New Democracy	Was there a democratic transition in the previous five years? Yes = 1	1990–2021	Coppedge et al. (2023)
Opposition Strength	Vote share of the largest opposition party in the most recent legislative election	1990–2021	Cruz, Keefer and Scartascini (2021)
Political Corruption Index	Average levels of public sector, executive, legislative, and judicial corruption, from 0 to 1, from less corrupt to more corrupt	1990–2021	Coppedge et al. (2023)
Polyarchy	Electoral democracy index	1990–2021	Coppedge et al. (2023)
Revision	Is there change in the GDP value reported by two consecutive data releases? Yes = 1	1990–2021	WDI
Rigorous Public Administration	Are public officials rigorous and impartial in the performance of their duties? (no = 0, weakly = 1, modestly = 2, mostly = 3, fully = 4)	1990–2022	Coppedge et al. (2023)
SDDS Compliance	Does the state comply with the IMF's Special Data Dissemination Standard (SDDS) specifications for the coverage, periodicity, and timeliness of data dissemination? Yes = 1	1990–2021	IMF Dissemination Standards Bulletin Board
SNA Change	Was the SNA in use updated this year?	1994–2021	UN National Accounts Statistics, complemented by WDI Metadata and IMF International Financial Statistics
Suffrage	Share of adult citizens that have the legal right to vote in national elections	1990–2021	Coppedge et al. (2023)
Turnover Frequency	Number of changes in party control of the executive in the preceding five years	1990–2021	Cruz, Keefer and Scartascini (2021)

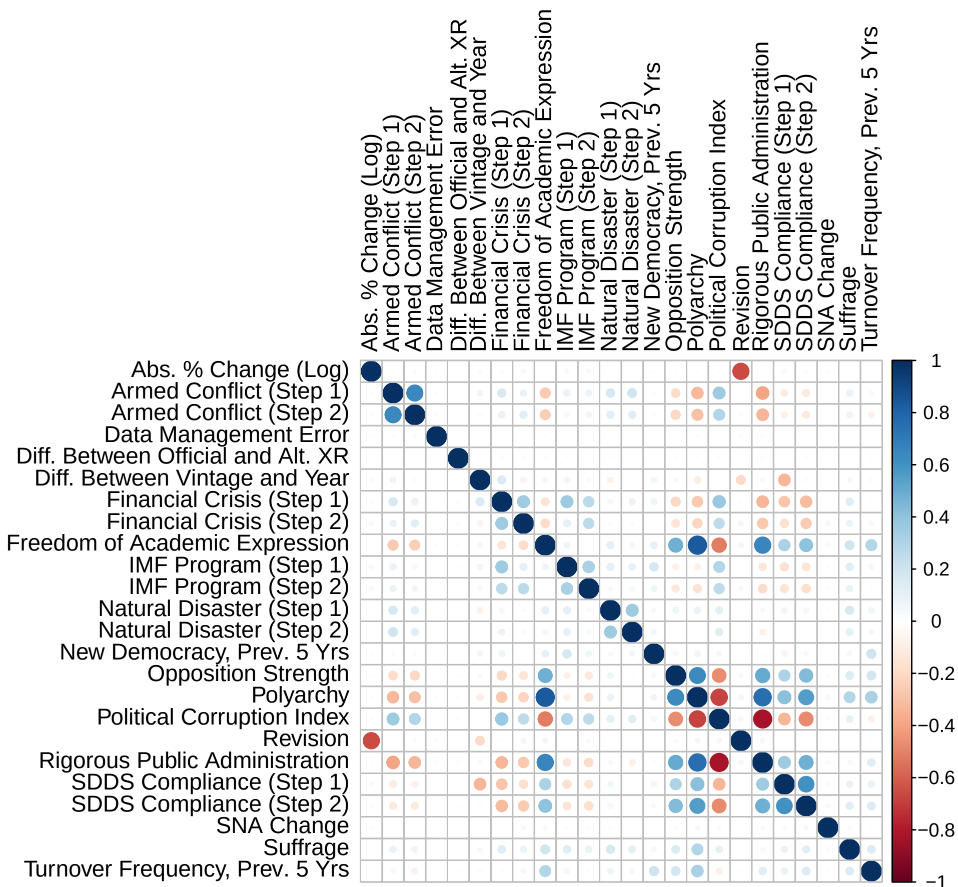
Table B.2: Summary Statistics, All Vintages

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Abs. % Change (Log)	64378	-9.411	9.059	-30.02	-17.89	-0.9071	9.769
Armed Conflict (Step 1)	406880						
... 0	298793	73.44%					
... 1	108087	26.56%					
Armed Conflict (Step 2)	409711						
... 0	333167	81.32%					
... 1	76544	18.68%					
Data Management Error	409711						
... 0	409662	99.99%					
... 1	49	0.01%					
Diff. Between Official and Alt. XR	403725	-0.00002156	1.001	-70.65	0.01416	0.01416	0.01458
Diff. Between Vintage and Year	409711	12.71	7.808	0.3333	6	18.83	31
Financial Crisis (Step 1)	406880						
... 0	237433	58.35%					
... 1	169447	41.65%					
Financial Crisis (Step 2)	409711						
... 0	318437	77.72%					
... 1	91274	22.28%					
Freedom of Academic Expression	402054						
... 0	35089	8.73%					
... 1	46431	11.55%					
... 2	61187	15.22%					
... 3	144852	36.03%					
... 4	114495	28.48%					
IMF Program (Step 1)	406880						
... 0	273071	67.11%					
... 1	133809	32.89%					
IMF Program (Step 2)	409711						
... 0	313569	76.53%					
... 1	96142	23.47%					
Natural Disaster (Step 1)	406880						
... 0	157087	38.61%					
... 1	249793	61.39%					
Natural Disaster (Step 2)	409711						
... 0	144624	35.3%					
... 1	265087	64.7%					
New Democracy, Prev. 5 Yrs	396796						
... 0	369798	93.2%					
... 1	26998	6.8%					
Opposition Strength	401606	19.9	22.18	0	0	41.3	99.5
Polyarchy	402054	0.4958	0.2748	0.013	0.245	0.759	0.922
Political Corruption Index	400486	0.5145	0.3029	0.002	0.212	0.79	0.966
Revision	409711						
... 0	339589	82.89%					
... 1	70122	17.11%					
Rigorous Public Administration	402054						
... 0	13452	3.35%					
... 1	113384	28.2%					
... 2	110833	27.57%					
... 3	105477	26.23%					
... 4	58908	14.65%					
SDDS Compliance (Step 1)	406880						
... 0	327657	80.53%					
... 1	79223	19.47%					
SDDS Compliance (Step 2)	409711						
... 0	246227	60.1%					
... 1	163484	39.9%					
SNA Change	407819						
... 0	381615	93.57%					
... 1	26204	6.43%					
Suffrage	402054	0.9696	0.163	0	1	1	1
Turnover Frequency, Prev. 5 Yrs	401368						
... 0	228040	56.82%					
... 1	150385	37.47%					
... 2	22782	5.68%					
... 3	161	0.04%					

Table B.3: Summary Statistics, Only Main Vintages

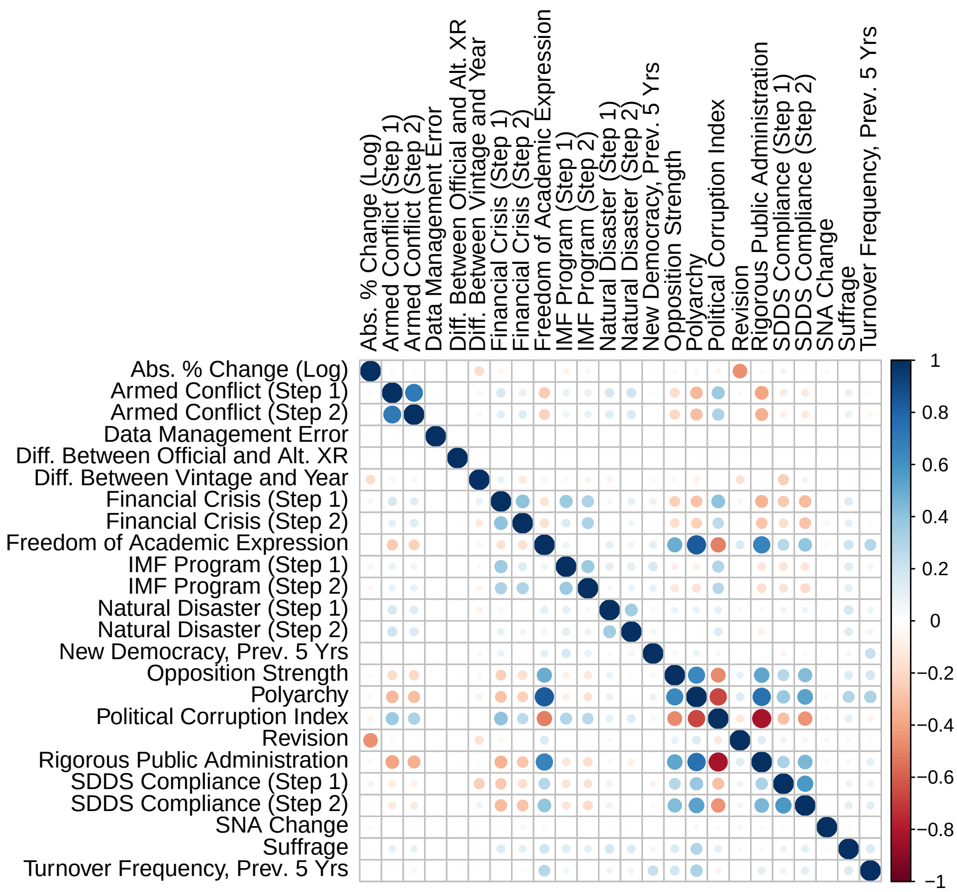
Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Abs. % Change (Log)	41167	-7.538	8.624	-30.02	-13.22	-0.07151	6.93
Armed Conflict (Step 1)	80517						
... 0	58442	72.58%					
... 1	22075	27.42%					
Armed Conflict (Step 2)	81211						
... 0	64044	78.86%					
... 1	17167	21.14%					
Data Management Error	81211						
... 0	81202	99.99%					
... 1	9	0.01%					
Diff. Between Official and Alt. XR	79852	-0.00009406	1.004	-81.88	0.01221	0.01221	0.0127
Diff. Between Vintage and Year	81211	10.58	7.364	0.3333	4.333	15.67	30.75
Financial Crisis (Step 1)	80517						
... 0	45213	56.15%					
... 1	35304	43.85%					
Financial Crisis (Step 2)	81211						
... 0	57928	71.33%					
... 1	23283	28.67%					
Freedom of Academic Expression	79557						
... 0	7234	9.09%					
... 1	9342	11.74%					
... 2	12051	15.15%					
... 3	28608	35.96%					
... 4	22322	28.06%					
IMF Program (Step 1)	80517						
... 0	53310	66.21%					
... 1	27207	33.79%					
IMF Program (Step 2)	81211						
... 0	59408	73.15%					
... 1	21803	26.85%					
Natural Disaster (Step 1)	80517						
... 0	32020	39.77%					
... 1	48497	60.23%					
Natural Disaster (Step 2)	81211						
... 0	27989	34.46%					
... 1	53222	65.54%					
New Democracy, Prev. 5 Yrs	78361						
... 0	72624	92.68%					
... 1	5737	7.32%					
Opposition Strength	79530	19.67	22.18	0	0	41.23	99.5
Polyarchy	79557	0.4885	0.2768	0.013	0.239	0.757	0.922
Political Corruption Index	79197	0.5131	0.3015	0.002	0.222	0.79	0.966
Revision	81211						
... 0	34899	42.97%					
... 1	46312	57.03%					
Rigorous Public Administration	79557						
... 0	2686	3.38%					
... 1	22551	28.35%					
... 2	22416	28.18%					
... 3	20230	25.43%					
... 4	11674	14.67%					
SDDS Compliance (Step 1)	80517						
... 0	68108	84.59%					
... 1	12409	15.41%					
SDDS Compliance (Step 2)	81211						
... 0	52316	64.42%					
... 1	28895	35.58%					
SNA Change	80362						
... 0	75743	94.25%					
... 1	4619	5.75%					
Suffrage	79557	0.9672	0.1692	0	1	1	1
Turnover Frequency, Prev. 5 Yrs	79500						
... 0	44675	56.19%					
... 1	30110	37.87%					
... 2	4685	5.89%					
... 3	30	0.04%					

Figure B.1: Correlation Plot, All Vintages



This figure shows the correlation plot for all variables included in the main analysis, for all vintages. To generate this figure, *Abs. % Change (Log)* was coded as zero for all instances of *Revision = 1*. In the actual analysis, *Abs. % Change (Log)* takes the value of zero for all instances of *Revision = 1*.

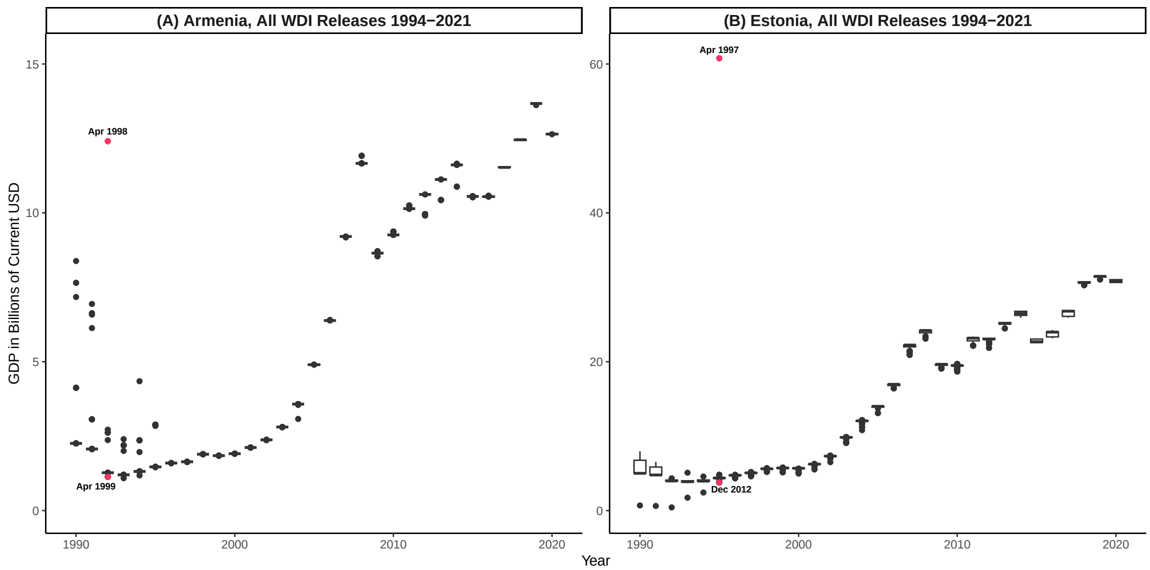
Figure B.2: Correlation Plot, Only Main Vintages



This figure shows the correlation plot for all variables included in the main analysis, only for the main vintages. To generate this figure, *Abs. % Change (Log)* was coded as zero for all instances of *Revision = 1*. In the actual analysis, *Abs. % Change (Log)* takes the value of zero for all instances of *Revision = 1*.

C Additional Descriptive Information

Figure C.1: Current GDP of Armenia and Estonia, 1990–2020



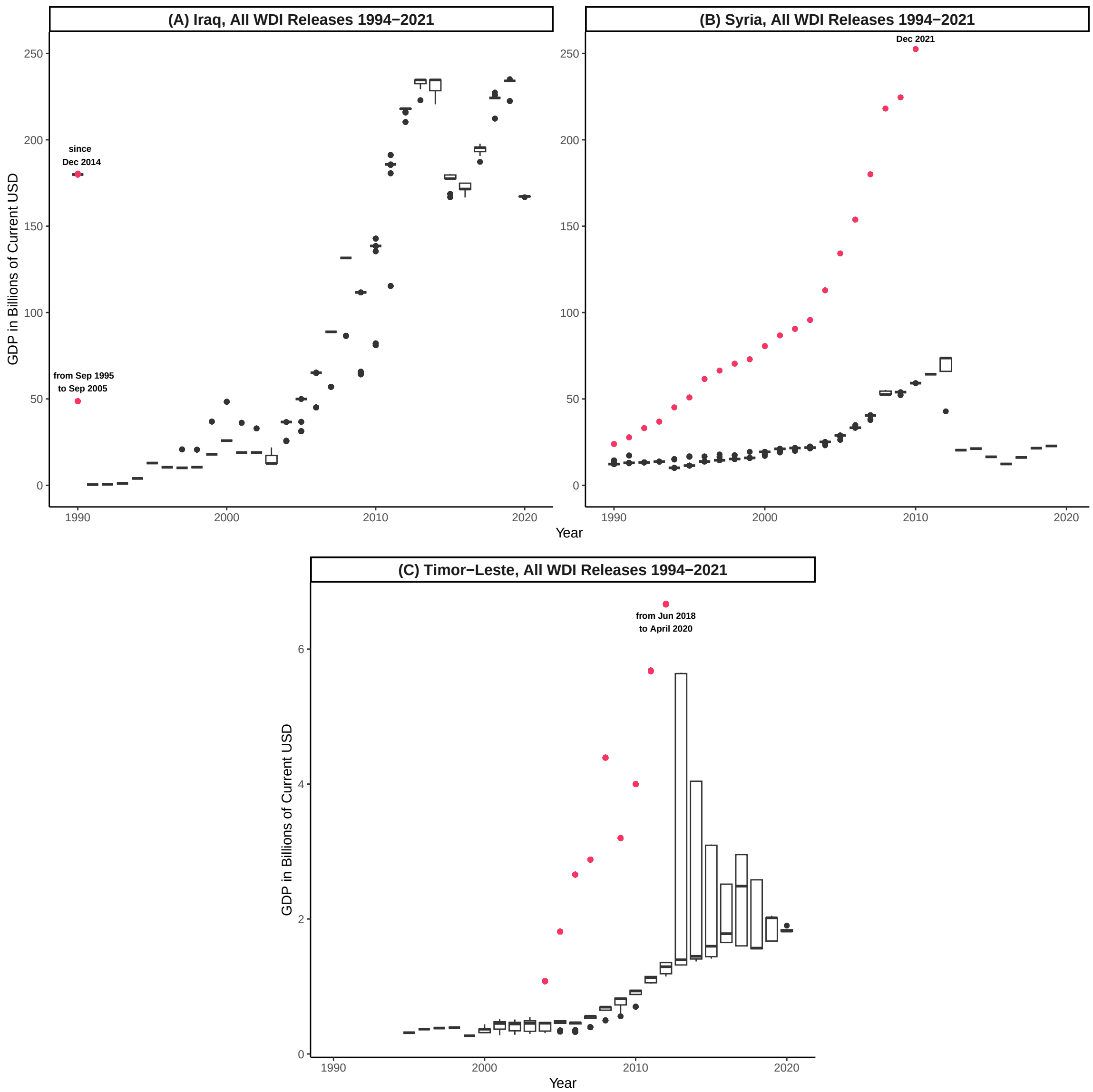
These boxplots present the distribution of current GDP estimates for (A) Armenia and (B) Estonia, from 1990 to 2020, using data drawn from the 104 WDI releases from April 1994 to December 2021. Section 4 discusses the data in detail.

To give readers a clearer grasp of the variation in the data, Figure C.1 presents the GDP of two former Soviet republics, Armenia and Estonia (both of which gained independence in 1991). According to the April 1998 WDI, Armenia had a GDP of 12.4 billion in 1992 — a number over four times as large as what any other WDI release reports. According to the April 1997 WDI, Estonia had a GDP of 60.8 billion in 1995 — a number at least 13 times as large as what other releases report. Since these extreme values only appear once, I assume they are the product of a data management error corrected in subsequent vintages.

Figure C.2 presents the GDP of two war-plagued countries in the Middle East, Iraq and Syria, and a newly independent country, Timor-Leste. Iraq in 1990 is an interesting case: this country-year pair first enters the WDI in September 1995 and takes the value of 48.66 billion until September 2005, at which point it ceases to be included. It reappears in the December 2014 WDI, at which point it is reported to be nearly four times as large: 179.91 billion. Syria's GDP from 1990 to 2010 is considerably larger in the December 2021 WDI than in other vintages. As of 2024, these values have not been revised; they are the most up-to-date values, suggesting that the WDI has not recognized them as erroneous.

As Figure C.2 further shows, all vintages report a GDP between 1.1 and 1.2 billion for Timor-Leste in 2012, with the exception of 16 vintages between June 2018 and April 2020 that report a number six times larger. Indeed, these 16 vintages provide exceptionally large values for Timor-Leste for all years from 2004 until 2012.

Figure C.2: Current GDP of Iraq, Syria, and Timor-Leste, 1990–2020



These boxplots present the distribution of current GDP estimates for (A) Iraq, (B) Syria, and (C) Timor-Leste from 1990 to 2020, using data drawn from the 104 WDI releases from April 1994 to December 2021. Section 4 discusses the data in detail.

Iraq, Syria, and Timor-Leste's unusual values appear in multiple vintages. They are not listed by the World Bank in its Data Updates and Errata website and the WDI team did not respond to my inquiries about these specific observations, so I do not code them as an error. In intentionally setting a high bar for an error, I likely underestimate the prevalence of such errors.

D Models With Alternative Measures

D.1 Regime Type

Table D.1: Predicting the Likelihood of Data Revisions (Alt. Measures of Regime Type)

	Dependent Variable: Revision = 1			
	(1) *All Vintages*	(2) *All Vintages*	(3) *Only Main Vintages*	(4) *Only Main Vintages*
Rigorous Public Administration	0.09***	0.10***	0.12***	0.13***
	(0.02)	(0.02)	(0.02)	(0.02)
Polity	0.02***		0.02***	
	(0.00)		(0.00)	
Freedom House		0.07***		0.09***
		(0.01)		(0.01)
IMF Program (Step 1)	0.04**	0.08***	0.05**	0.09***
	(0.02)	(0.02)	(0.02)	(0.02)
IMF Program (Step 2)	−0.08***	−0.07***	−0.02	0.00
	(0.02)	(0.02)	(0.03)	(0.03)
SDDS Compliance (Step 1)	−0.23***	−0.29***	−0.27***	−0.34***
	(0.02)	(0.02)	(0.04)	(0.03)
SDDS Compliance (Step 2)	0.14***	0.16***	0.06	0.07
	(0.04)	(0.04)	(0.05)	(0.04)
Financial Crisis (Step 1)	0.04**	0.05***	0.02	0.05*
	(0.02)	(0.02)	(0.03)	(0.02)
Financial Crisis (Step 2)	0.05***	0.07***	0.03	0.05*
	(0.02)	(0.02)	(0.03)	(0.03)
Natural Disaster (Step 1)	0.06***	0.06***	0.05**	0.05**
	(0.02)	(0.02)	(0.02)	(0.02)
Natural Disaster (Step 2)	0.04**	0.02	0.01	−0.00
	(0.02)	(0.02)	(0.02)	(0.02)
Armed Conflict (Step 1)	−0.02	−0.02	−0.01	−0.00
	(0.03)	(0.03)	(0.04)	(0.04)
Armed Conflict (Step 2)	−0.09***	−0.09***	−0.07*	−0.07*
	(0.03)	(0.03)	(0.04)	(0.04)
Diff. Between Official and Alt. XR	−0.01**	−0.01**	−0.00	−0.00
	(0.01)	(0.01)	(0.01)	(0.01)
SNA Change	0.22***	0.20***	0.47***	0.47***
	(0.03)	(0.03)	(0.04)	(0.04)
Data Management Error	6.91***	6.88***	−0.10	−0.14
	(0.62)	(0.62)	(0.89)	(0.89)
Diff. Between Vintage and Year	−0.10***	−0.11***	−0.07***	−0.08***
	(0.01)	(0.01)	(0.01)	(0.01)
Intercept	−3.35***	−2.95***	0.76***	1.25***
	(0.48)	(0.48)	(0.25)	(0.27)
Log Likelihood	−82123.87	−87035.02	−39461.27	−41757.32
Observations	376061	397331	74057	78081
Number of Countries	164	170	164	170
Number of Years	30	31	30	31
Number of Vintages	103	103	26	26
Variance: Countries (Intercept)	0.35	0.30	0.44	0.38
Variance: Years (Intercept)	0.09	0.13	0.14	0.26
Variance: Vintages (Intercept)	20.05	19.91	1.33	1.32

This table presents the results of four logistic regressions with random intercepts for country, year, and vintage. Step 1 is the reference year, the year of data collection. Step 2 is the vintage year, the year of data revision. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

To quantify the effect of regime type on revisions, the full models use V-Dem's *Polyarchy* index (Coppedge et al., 2023), which measures the quality of electoral democracy, including extensive suffrage, fair elections, freedom of expression, and access to information (in an ordinal scale from 0 to 1). Tables D.1 and D.2

Table D.2: Predicting the Magnitude of Data Revisions (Alt. Measures of Regime Type)

	Dependent Variable: Abs. % Change (Log)			
	(1) *All Vintages*	(2) *All Vintages*	(3) *Only Main Vintages*	(4) *Only Main Vintages*
Rigorous Public Administration	0.17**	0.18**	0.20**	0.18*
	(0.07)	(0.07)	(0.09)	(0.09)
Polity	0.03**		0.03**	
	(0.01)		(0.01)	
Freedom House		0.09**		0.13**
		(0.05)		(0.06)
IMF Program (Step 1)	−0.10	−0.11	−0.08	−0.11
	(0.08)	(0.08)	(0.10)	(0.10)
IMF Program (Step 2)	−0.38***	−0.38***	−0.31***	−0.33***
	(0.09)	(0.09)	(0.11)	(0.11)
SDDS Compliance (Step 1)	0.02	−0.08	−0.25*	−0.42***
	(0.11)	(0.11)	(0.14)	(0.14)
SDDS Compliance (Step 2)	0.59***	0.56***	0.89***	0.93***
	(0.15)	(0.15)	(0.17)	(0.17)
Financial Crisis (Step 1)	−0.18**	−0.17**	−0.18*	−0.21**
	(0.08)	(0.08)	(0.10)	(0.10)
Financial Crisis (Step 2)	0.58***	0.55***	−0.16	−0.16
	(0.09)	(0.09)	(0.11)	(0.11)
Natural Disaster (Step 1)	−0.14*	−0.11	−0.08	−0.09
	(0.07)	(0.07)	(0.09)	(0.09)
Natural Disaster (Step 2)	−0.09	−0.10	0.21**	0.24**
	(0.08)	(0.08)	(0.10)	(0.10)
Armed Conflict (Step 1)	−0.35**	−0.36***	−0.72***	−0.67***
	(0.14)	(0.14)	(0.17)	(0.17)
Armed Conflict (Step 2)	0.37***	0.26*	0.64***	0.54***
	(0.14)	(0.14)	(0.16)	(0.16)
Diff. Between Official and Alt. XR	0.00	0.00	−0.06*	−0.07*
	(0.03)	(0.03)	(0.04)	(0.04)
SNA Change	1.76***	1.68***	2.65***	2.62***
	(0.12)	(0.12)	(0.15)	(0.15)
Data Management Error	12.55***	12.16***	15.16**	15.65**
	(1.54)	(1.51)	(6.84)	(6.86)
Diff. Between Vintage and Year	−0.23***	−0.24***	−0.35***	−0.35***
	(0.01)	(0.01)	(0.01)	(0.01)
Intercept	−5.66***	−5.08***	−3.67***	−2.92***
	(0.64)	(0.67)	(0.74)	(0.78)
Log Likelihood	−200025.51	−210457.05	−127952.64	−134239.17
Observations	59745	62858	38238	40082
Number of Countries	164	170	164	170
Number of Years	30	31	30	30
Number of Vintages	78	78	26	26
Variance: Countries (Intercept)	3.05	2.99	3.35	3.31
Variance: Years (Intercept)	0.16	0.15	0.20	0.18
Variance: Vintages (Intercept)	25.11	24.97	11.71	11.44

This table presents the results of four linear regressions with random intercepts for country, year, and vintage. Step 1 is the reference year, the year of data collection. Step 2 is the vintage year, the year of data revision. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

present the full results using two other indices that capture different dimensions of democracy. *Polity 2* (Marshall and Gurr, 2020) prioritizes regime characteristics that reflect democratic or autocratic patterns, such as political competition or checks and balances (from –10 to +10). As with *Polyarchy*, larger values of *POLCON III* and *Polity 2* reflect more democratic regimes, which should be associated with a higher likelihood of revisions. *Freedom House* is a civil liberties index that typically ranges from 1 (most free) to 7 (least free); I flipped this scale so that its interpretation mirrors that of the other regime type variables, with larger values reflecting more freedom. The results are robust to these alternative measures, despite their less extensive coverage.

D.2 State Capacity

Table D.3: Predicting the Likelihood of Data Revisions (Alt. Measures of State Capacity)

	Dependent Variable: Revision = 1			
	(1) *All Vintages*	(2) *All Vintages*	(3) *Only Main Vintages*	(4) *Only Main Vintages*
State Capacity	0.07**		0.11***	
	(0.03)		(0.04)	
Bureaucratic Quality		0.07***		0.10***
		(0.01)		(0.02)
Polyarchy	0.76***	0.74***	1.09***	1.08***
	(0.08)	(0.08)	(0.10)	(0.11)
IMF Program (Step 1)	0.05***	0.05***	0.07***	0.07**
	(0.02)	(0.02)	(0.02)	(0.03)
IMF Program (Step 2)	−0.09***	−0.07***	−0.01	−0.03
	(0.02)	(0.02)	(0.03)	(0.03)
SDDS Compliance (Step 1)	−0.26***	−0.25***		−0.30***
	(0.03)	(0.03)		(0.04)
SDDS Compliance (Step 2)	0.11***	0.23***		0.10*
	(0.04)	(0.04)		(0.05)
Financial Crisis (Step 1)	0.05***	0.03	0.06**	0.03
	(0.02)	(0.02)	(0.03)	(0.03)
Financial Crisis (Step 2)	0.07***	0.02	0.01	0.03
	(0.02)	(0.02)	(0.03)	(0.03)
Natural Disaster (Step 1)	0.05***	0.07***	0.05**	0.06**
	(0.02)	(0.02)	(0.02)	(0.03)
Natural Disaster (Step 2)	0.05***	0.05***	0.03	−0.02
	(0.02)	(0.02)	(0.02)	(0.03)
Armed Conflict (Step 1)	−0.04	−0.07*	0.00	−0.05
	(0.03)	(0.04)	(0.04)	(0.05)
Armed Conflict (Step 2)	−0.10***	−0.21***	−0.06	−0.22***
	(0.03)	(0.03)	(0.04)	(0.05)
Diff. Between Official and Alt. XR	−0.01**	−0.01**	−0.00	−0.00
	(0.01)	(0.01)	(0.01)	(0.01)
SNA Change	0.22***	0.16***	0.47***	0.45***
	(0.03)	(0.03)	(0.04)	(0.04)
Data Management Error	6.95***	6.92***	−0.16	−0.57
	(0.62)	(0.61)	(0.91)	(0.97)
Diff. Between Vintage and Year	−0.08***	−0.11***	−0.05***	−0.08***
	(0.00)	(0.01)	(0.00)	(0.01)
Intercept	−3.73***	−3.49***	0.33	0.48*
	(0.48)	(0.47)	(0.25)	(0.26)
Log Likelihood	−80262.22	−68241.24	−38465.00	−32637.04
Observations	367567	312598	73069	61324
Number of Countries	163	136	163	136
Number of Years	27	31	27	31
Number of Vintages	103	103	26	26
Variance: Countries (Intercept)	0.33	0.34	0.39	0.41
Variance: Years (Intercept)	0.03	0.15	0.03	0.30
Variance: Vintages (Intercept)	20.79	19.33	1.46	1.28

This table presents the results of four logistic regressions with random intercepts for country, year, and vintage. Step 1 is the reference year, the year of data collection. Step 2 is the vintage year, the year of data revision. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Hanson and Sigman (2021) use Bayesian latent variable analysis to combine 21 indicators of extractive, coercive, and administrative capacity into a single *State Capacity* index. Out of the 21 components, the V-

Table D.4: Predicting the Magnitude of Data Revisions (Alt. Measures of State Capacity)

	Dependent Variable: Abs. % Change (Log)			
	(1) *All Vintages*	(2) *All Vintages*	(3) *Only Main Vintages*	(4) *Only Main Vintages*
State Capacity	0.02		−0.01	
	(0.13)		(0.15)	
Bureaucratic Quality		0.09		0.19**
		(0.07)		(0.08)
Polyarchy	1.41***	1.33***	1.78***	1.52***
	(0.33)	(0.35)	(0.40)	(0.41)
IMF Program (Step 1)	−0.12	−0.06	−0.10	0.01
	(0.08)	(0.09)	(0.10)	(0.11)
IMF Program (Step 2)	−0.40***	−0.28***	−0.36***	−0.10
	(0.09)	(0.10)	(0.11)	(0.12)
SDDS Compliance (Step 1)	0.02	−0.17	−0.25*	−0.50***
	(0.11)	(0.12)	(0.14)	(0.15)
SDDS Compliance (Step 2)	0.61***	1.05***	0.84***	1.61***
	(0.16)	(0.17)	(0.18)	(0.19)
Financial Crisis (Step 1)	−0.18**	−0.16*	−0.19*	−0.22*
	(0.08)	(0.09)	(0.10)	(0.11)
Financial Crisis (Step 2)	0.62***	0.38***	−0.14	−0.23*
	(0.09)	(0.10)	(0.11)	(0.12)
Natural Disaster (Step 1)	−0.12	−0.12	−0.08	−0.11
	(0.07)	(0.08)	(0.09)	(0.10)
Natural Disaster (Step 2)	−0.06	−0.09	0.23**	0.35***
	(0.08)	(0.09)	(0.10)	(0.11)
Armed Conflict (Step 1)	−0.40***	−0.18	−0.73***	−0.49**
	(0.14)	(0.16)	(0.17)	(0.19)
Armed Conflict (Step 2)	0.33**	−0.61***	0.56***	−0.32*
	(0.14)	(0.16)	(0.16)	(0.18)
Diff. Between Official and Alt. XR	0.00	0.00	−0.06*	−0.06*
	(0.03)	(0.03)	(0.04)	(0.04)
SNA Change	1.71***	1.40***	2.59***	2.38***
	(0.12)	(0.13)	(0.15)	(0.16)
Data Management Error	13.68***	11.87***	14.85**	
	(1.64)	(1.55)	(6.87)	
Diff. Between Vintage and Year	−0.23***	−0.25***	−0.36***	−0.36***
	(0.01)	(0.01)	(0.01)	(0.01)
Intercept	−5.90***	−5.95***	−3.86***	−4.56***
	(0.65)	(0.64)	(0.74)	(0.73)
Log Likelihood	−195954.25	−167336.77	−126856.66	−106390.82
Observations	58486	49992	37860	31865
Number of Countries	163	136	163	136
Number of Years	27	31	27	30
Number of Vintages	76	78	26	26
Variance: Countries (Intercept)	3.01	2.98	3.21	2.97
Variance: Years (Intercept)	0.16	0.14	0.20	0.12
Variance: Vintages (Intercept)	25.31	21.87	11.72	10.54

This table presents the results of four linear regressions with random intercepts for country, year, and vintage. Step 1 is the reference year, the year of data collection. Step 2 is the vintage year, the year of data revision. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Dem index *Rigorous Public Administration* has the most extensive coverage, which is why I use it in the main analysis. In Tables D.3 and D.4, I replace *Rigorous Public Administration* with Hanson and Sigman's *State Capacity* index, which ranges from –22.31 to 2.96 (with higher values indicating higher state capacity). I also estimate models with *Bureaucratic Quality*, from the International Country Risk Guide (ICRG), ranging from 0 (low) to 4 (high), which measures the extent to which a country's bureaucracy has the strength and expertise to govern without drastic changes in policy or interruptions in government services. Both variables have worse coverage, but all main effects are robust to their inclusion. Notably, *Data Management Error* does not drop out of any model estimated with *State Capacity*.

E Models With Fixed Effects

Table E.1: Predicting the Likelihood of Data Revisions (Fixed Effects)

	Dependent Variable: Revision = 1	
	(1) *All Vintages*	(2) *Only Main Vintages*
Rigorous Public Administration	0.06	0.06
	(0.05)	(0.05)
Polyarchy	0.69***	1.01***
	(0.26)	(0.26)
IMF Program (Step 1)	0.07*	0.08**
	(0.03)	(0.04)
IMF Program (Step 2)	−0.07	−0.00
	(0.06)	(0.07)
SDDS Compliance (Step 1)	−0.27***	−0.33***
	(0.07)	(0.08)
SDDS Compliance (Step 2)	0.14	0.02
	(0.10)	(0.11)
Financial Crisis (Step 1)	0.06	0.05
	(0.05)	(0.06)
Financial Crisis (Step 2)	0.08	0.06
	(0.06)	(0.07)
Natural Disaster (Step 1)	0.05***	0.04**
	(0.02)	(0.02)
Natural Disaster (Step 2)	0.02	−0.00
	(0.05)	(0.07)
Armed Conflict (Step 1)	−0.04	−0.03
	(0.07)	(0.09)
Armed Conflict (Step 2)	−0.09	−0.08
	(0.12)	(0.14)
Diff. Between Official and Alt. XR	−0.01***	−0.00**
	(0.00)	(0.00)
SNA Change	0.20**	0.47***
	(0.09)	(0.10)
Data Management Error	6.98***	−0.22
	(1.23)	(0.46)
Log Likelihood	−86235.26	−41234.79
Observations	295399	78178
Number of Countries	170	170
Number of Years	31	31
Number of Vintages	80	26

This table presents the results of two logistic regressions with fixed effects for country, year, and vintage and standard errors clustered by country. Step 1 is the reference year, the year of data collection. Step 2 is the vintage year, the year of data revision. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Given the large number of units (170 countries), time periods (31 years), and vintages (103), models with country, year, and vintage fixed effects are computationally inefficient: they lead to a substantial loss of degrees of freedom, and the variable *Diff. Between Vintage and Year* drops out due to collinearity. Despite the reduced sample size, the main results for the dependent variable *Revision* are robust to the use of fixed effects instead of random effects, as Table E.1 shows. However, the main results for the dependent variable

Abs. % Change (Log) are not, as Table E.2 shows.

Table E.2: Predicting the Magnitude of Data Revisions (Fixed Effects)

	Dependent Variable: Abs. % Change (Log)	
	(1)	(2)
	All Vintages	*Only Main Vintages*
Rigorous Public Administration	−0.02	−0.08
	(0.13)	(0.17)
Polyarchy	0.71	1.05
	(0.72)	(0.84)
IMF Program (Step 1)	−0.04	−0.02
	(0.13)	(0.15)
IMF Program (Step 2)	−0.32	−0.28
	(0.29)	(0.33)
SDDS Compliance (Step 1)	−0.11	−0.40
	(0.29)	(0.27)
SDDS Compliance (Step 2)	0.25	0.80
	(0.56)	(0.56)
Financial Crisis (Step 1)	−0.14	−0.17
	(0.14)	(0.15)
Financial Crisis (Step 2)	0.60**	−0.13
	(0.29)	(0.30)
Natural Disaster (Step 1)	−0.14*	−0.09
	(0.07)	(0.07)
Natural Disaster (Step 2)	−0.12	0.24
	(0.24)	(0.25)
Armed Conflict (Step 1)	−0.37	−0.63**
	(0.26)	(0.26)
Armed Conflict (Step 2)	0.31	0.65
	(0.70)	(0.67)
Diff. Between Official and Alt. XR	0.01	−0.06***
	(0.00)	(0.00)
SNA Change	1.68***	2.62***
	(0.40)	(0.50)
Data Management Error	12.02***	15.59***
	(1.03)	(0.76)
R^2 (full model)	0.43	0.37
Observations	62916	40119
Number of Countries	170	170
Number of Years	31	30
Number of Vintages	78	26

This table presents the results of two linear regressions with fixed effects for country, year, and vintage and standard errors clustered by country. Step 1 is the reference year, the year of data collection. Step 2 is the vintage year, the year of data revision. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

F Direction of Revisions

When revisions occur, what explains their direction? Table F.1 restricts the analysis to instances where there was a revision; now, the baseline is a downward revision, and the outcome reflects the occurrence of an *Upward Revision*. Relative to downward revisions, upward revisions are significantly as states become more democratic (that is, as the *Polyarchy* score increases) or in cases of SNA change. *Data Management Error* drops out of Model 2: focusing only on revisions in the main vintages results in too few observations that qualify as coding errors.

Table F.1: Predicting the Likelihood of Upward Data Revisions

	Dependent Variable: Upward Revision = 1	
	(1)	(2)
	All Vintages	*Only Main Vintages*
Rigorous Public Administration	−0.01	−0.01
	(0.02)	(0.02)
Polyarchy	0.14**	0.22**
	(0.07)	(0.09)
IMF Program (Step 1)	0.01	0.01
	(0.02)	(0.03)
IMF Program (Step 2)	0.00	−0.03
	(0.02)	(0.03)
SDDS Compliance (Step 1)	0.08***	0.07*
	(0.03)	(0.04)
SDDS Compliance (Step 2)	−0.01	0.03
	(0.03)	(0.04)
Financial Crisis (Step 1)	−0.04**	−0.05*
	(0.02)	(0.03)
Financial Crisis (Step 2)	−0.04*	−0.03
	(0.02)	(0.03)
Natural Disaster (Step 1)	0.01	−0.01
	(0.02)	(0.03)
Natural Disaster (Step 2)	−0.06***	−0.03
	(0.02)	(0.03)
Armed Conflict (Step 1)	0.03	0.00
	(0.03)	(0.04)
Armed Conflict (Step 2)	−0.08**	−0.07*
	(0.04)	(0.04)
Diff. Between Official and Alt. XR	−0.01	−0.37
	(0.01)	(3001.19)
SNA Change	0.15***	0.54***
	(0.04)	(0.05)
Data Management Error	0.80	
	(0.49)	
Diff. Between Vintage and Year	−0.00	−0.01***
	(0.00)	(0.00)
Intercept	0.26***	0.41
	(0.08)	(36.65)
Log Likelihood	−42035.89	−26516.81
Observations	62916	40119
Number of Countries	170	170
Number of Years	31	30
Number of Vintages	78	26
Variance: Countries (Intercept)	0.02	0.05
Variance: Years (Intercept)	0.00	0.00
Variance: Vintages (Intercept)	0.23	0.08

This table presents the results of two logistic regressions with random intercepts for country, year, and vintage. Step 1 is the reference year, the year of data collection. Step 2 is the vintage year, the year of data revision. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

G Regularized Regressions

The predictors included in the main analysis are not exhaustive: many other correlated factors might explain GDP revisions. However, including too many predictors relative to the sample size in linear or logistic regressions can lead to overfitting and multicollinearity: the model might capture noise instead of the true underlying relationships, and high correlations among predictors might destabilize coefficient estimates. To address these issues, I also estimate regularized regressions with the original variables listed in Table B.1

and the additional variables listed in Table G.1, each lagged and included for both steps 1 and 2 (with the exception of *Data Management Error*, *Diff. Between Vintage and Year*, and the time-invariant variables *Former European Colony*, *Island*, *Post-Soviet State*, and *Tax Haven*).

Table G.1: Additional Independent Variables Included in Regularized Regressions

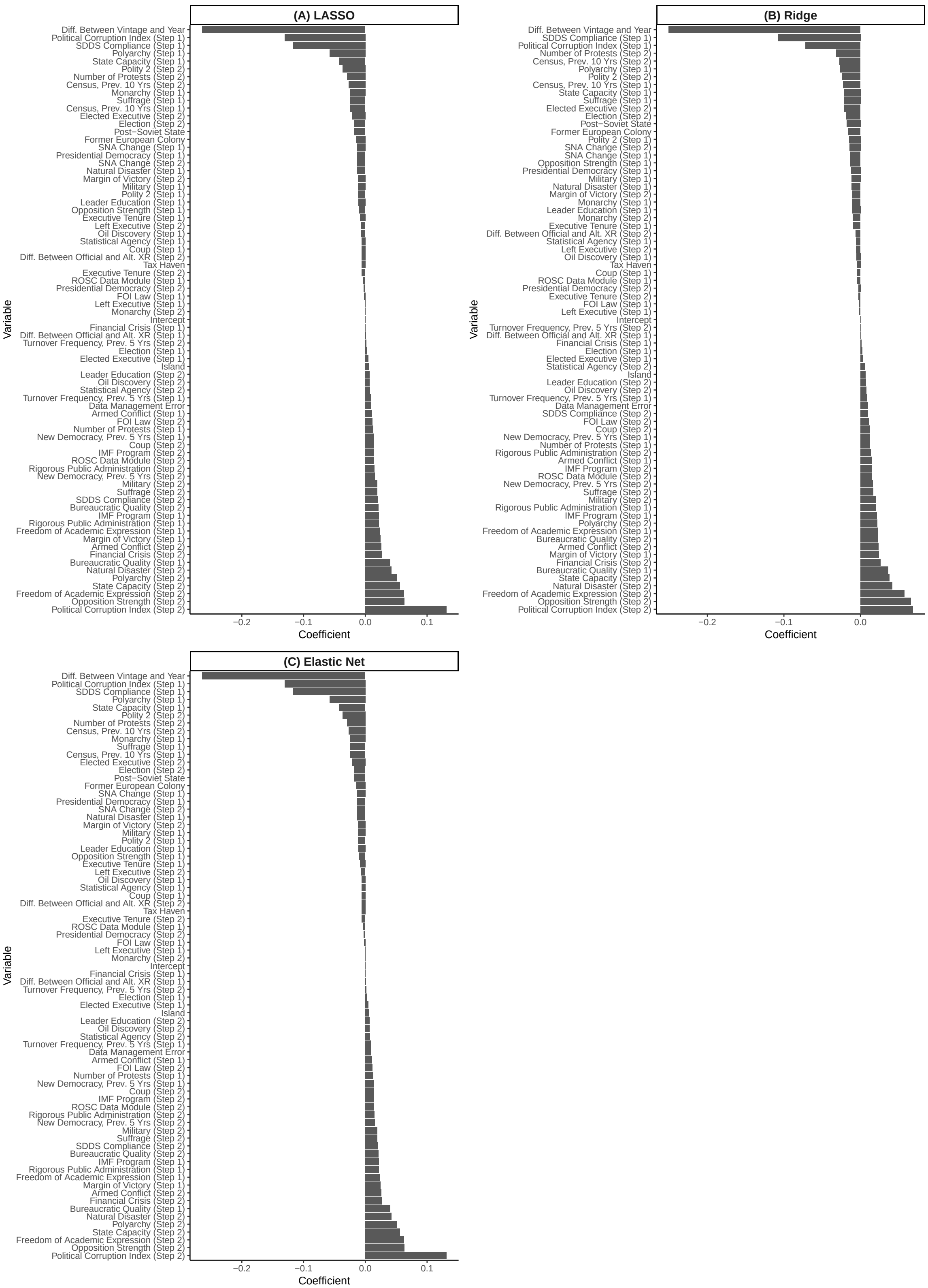
Variable	Description	Coverage	Source
Bureaucratic Quality	To what extent does the country's bureaucracy have the strength and expertise to govern without drastic changes in policy or interruptions in government services? Low = 0, High = 4	1990–2021	The PRS Group (2022)
Census	Was there a national census in the previous 10 years? Yes = 1	1990–2021	World Bank's Statistical Performance Indicators
Coup	Did a coup d'etat occur? Yes = 1	1990–2021	Coppedge et al. (2023)
Election	Did a presidential, legislative, or constituent assembly election take place? Yes = 1	1990–2021	For Brunei and Belize, Cruz, Keefer and Scartascini (2021); for all other countries, Coppedge et al. (2023)
Executive Tenure So Far	Number of years a leader has been in power during their current tenure	1990–2020	Bell, Besaw and Frank (2021)
Executive Was Elected	Was the executive leader elected to office? Yes = 1	1990–2020	Bell, Besaw and Frank (2021)
FOI Law	Was a Freedom of Information law (also known as a Right to Information law) passed? Yes = 1	1990–2021	Global Right to Information Rating
Former European Colony	Is this country a former colony of Belgium, France, Germany, Great Britain, Italy, Netherlands, Portugal, or Spain? Yes = 1	1990–2021	Becker (2019)
Island	Is the country an island? Yes = 1	1990–2021	Own coding
Leader Education	Leader's level of education summarized in eight categories	1990–2020	Dreher et al. (2020)
Left Executive	Party orientation of the executive with respect to economic policy. Left = 1	1990–2020	Cruz, Keefer and Scartascini (2021)
Margin of Victory	Difference in the vote share between the largest government party and the largest opposition party in the most recent legislative election	1990–2021	Cruz, Keefer and Scartascini (2021)
Military	Direct or indirect military regime. Yes = 1	1990–2020	Bell, Besaw and Frank (2021)
Monarchy	Monarchy. Yes = 1	1990–2020	Bell, Besaw and Frank (2021)
Number of Protests	Number of recorded protests	1990–2020	Clark and Regan (2020)

Oil Discovery	Did this country discover a giant, megagiant, or super-giant oil or gas field? Yes = 1	1990–2020	Horn (2014); Cust, Mihalyi and Rivera-Ballesteros (2021)
Polity	Revised combined Polity score, from –10 (hereditary monarchy) to +10 (consolidated democracy)	1990–2018	Marshall and Gurr (2020)
Population Density	Total population, `SP.POP.TOTL`, divided by land area (sq. km), `AG.LND.TOTL.K2`	1990–2021	WDI
Post-Soviet State	Former Republic of the Union of Soviet Socialist Republics	1990–2021	Own coding
Presidential Democracy	Presidential democracy. Yes = 1	1990–2020	Bell, Besaw and Frank (2021)
ROSC Data Module	Was a ROSC Data Module conducted? Yes = 1	1990–2021	IMF
State Capacity	Estimate of state capacity by Hanson/Sigman	1990–2015	Hanson and Sigman (2021)
Statistical Agency	Is there a national statistical agency? Yes = 1	1990–2022	Coppedge et al. (2023); UN Statistics Division
Tax Haven	Does the US Department of Treasury consider this country a tax haven? Yes = 1	1990–2021	Graham et al. (2018); Graham and Tucker (2019)

Regularized regressions mitigate overfitting and multicollinearity by performing variable selection. LASSO adds a penalty to the absolute values of the coefficients (L1 regularization) that encourages most coefficients to become exactly zero, effectively performing feature selection by eliminating irrelevant variables. In contrast, ridge regression adds a penalty to the squared values of the coefficients (L2 regularization) that discourages large coefficients but does not force any coefficients to become exactly zero. Elastic net balances the strengths of LASSO and ridge regression, retaining groups of correlated variables. Since regularized regressions can be sensitive to the magnitude of predictors, I center and scale all the predictors in Tables B.1 and G.1, such that they all have a mean of zero and a standard deviation of one. This prevents variables with larger ranges from disproportionately influencing the models.

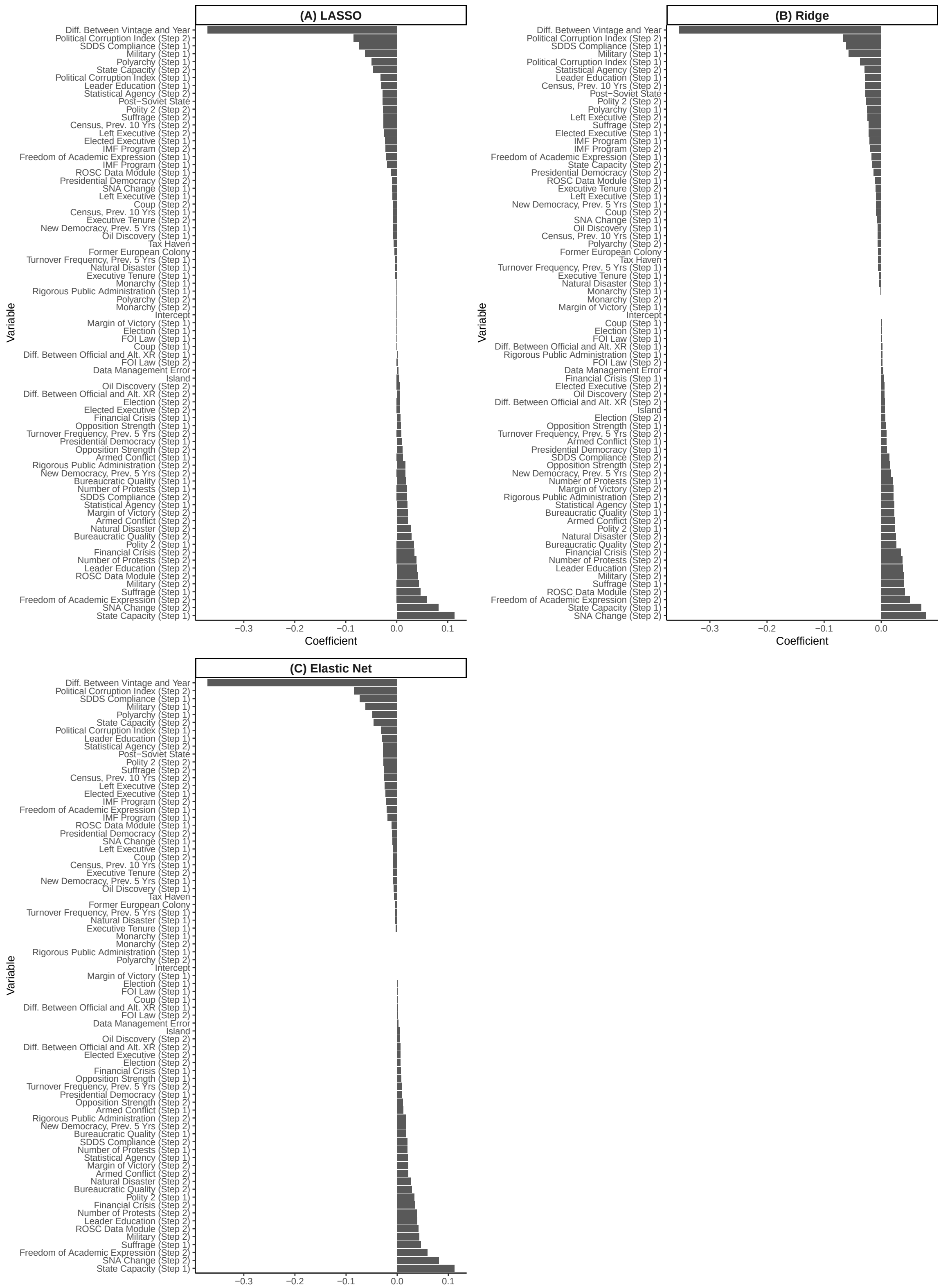
Consider the outcome *Revision*. For each model, Figure G.1 indicates the relative importance of all predictors. Because LASSO shrinks the coefficients of irrelevant variables to exactly zero, its importance rankings are sparse. The elastic net plot is similarly sparse. The ridge regression plot reflects a more distributed influence, as this model assigns nonzero coefficients to all variables. Still, all three models concur that *Political Corruption Index*, *Opposition Strength*, and *Freedom of Academic Expression* have the strongest positive association with the outcome, whereas *Difference Between Vintage and Year* has the strongest negative association. To be clear, this does not mean that *Political Corruption Index* or *Difference Between Vintage and Year* is the "best" or most significant predictor of variation in the outcome. Regularized regressions cannot be used to test hypotheses; they do not provide standard errors, so it is not possible to calculate p-values. In shrinking coefficients toward zero (or exactly to zero), these regressions distort the true relationship between predictors and the outcome. But since all predictors are scaled, Figure G.1 allows me to say that various measures of regime type have the largest positive effect on the outcome and *Difference Between Vintage and Year* has the largest negative outcome, as measured by standardized units. By this metric, the original models include the "right" variables, that is, the variables with the largest effects on the outcome. This conclusion is reinforced by Figure G.2, which identifies *SNA Change* and *State Capacity* as the variables with the strongest positive association with the outcome *Abs. % Change (Log)* and *Difference Between Vintage and Year* as the variable with the strongest negative association.

Figure G.1: Variable Importance Plots, Outcome: *Revision*, All Vintages



These variable importance plots highlight the predictors selected by each model: (A) LASSO, (B) ridge regression, and (C) elastic net. Variables with positive (negative) coefficients are associated with an increase (decrease) in the outcome, *Revision*. Variables with a coefficient of zero (or close to zero) are considered irrelevant: the model either excludes them (LASSO) or minimizes their impact (ridge and elastic net) to improve generalization and reduce overfitting.

Figure G.2: Variable Importance Plots, Outcome: *Abs. % Change (Log)*, All Vintages



These variable importance plots highlight the predictors selected by each model: (A) LASSO, (B) ridge regression, and (C) elastic net. Variables with positive (negative) coefficients are associated with an increase (decrease) in the outcome, *Abs. % Change (Log)*. Variables with a coefficient of zero (or close to zero) are considered irrelevant: the model either excludes them (LASSO) or minimizes their impact (ridge and elastic net) to improve generalization and reduce overfitting.

References

Becker, Bastian. 2019. "Introducing COLDAT: The Colonial Dates Dataset." *SOCIUM/SFB1342 Working Paper Series* (2):1–17.

Bell, Curtis, Clayton Besaw and Matthew Frank. 2021. *The Rulers, Elections, and Irregular Governance (REIGN) Dataset.*
URL: *https://oefdatascience.github.io/REIGN.github.io/*

Centre for Research on the Epidemiology of Disasters. 2020. *EM-DAT: The International Disaster Database.*
URL: *https://public.emdat.be*

Clark, David and Patrick Regan. 2020. *Mass Mobilization Protest Data.*
URL: *https://massmobilization.github.io/*

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Agnes Cornell, Sirianne Dahlum, Haakon Gjerløw, Adam Glynn, Allen Hicken, Joshua Krusell, Anna Lührmann, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Juraj Medzihorsky, Moa Olin, Pamela Paxton, Daniel Pemstein, Josefine Pernes, Johannes von Römer, Brigitte Seim, Rachel Sigman, Jeffrey Staton, Natalia Stepanova, Aksel Sundström, Eitan Tzelgov, Yi-ting Wang, Tore Wig, Steven Wilson and Daniel Ziblatt. 2023. *V-Dem Country-Year Dataset v13.*

Cruz, Cesi, Philip Keefer and Carlos Scartascini. 2021. *Database of Political Institutions 2020.*

Cust, James, David Mihalyi and Alexis Rivera-Ballesteros. 2021. The Economic Effects of Giant Oil and Gas Discoveries. In *Giant Fields of the Decade: 2010-2020*, ed. Charles A. Sternbach, Robert K. Merrill and John C. Dolson. Tulsa: AAPG pp. 21–36.

Dreher, Axel, Andreas Fuchs, Andreas Kammerlander, Lennart Kaplan, Charlott Robert and Kerstin Unfried. 2020. *The Political Leaders' Affiliation Database.*

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg and Håvard Strand. 2002. "Armed Conflict 1946–2001: A New Dataset." *Journal of Peace Research* 39(5):615–637.

Graham, Benjamin A. T. and Jacob R. Tucker. 2019. "The International Political Economy Data Resource." *Review of International Organizations* 14:149–161.

Graham, Benjamin A.T., Raymond Hicks, Helen Milner and Lori D. Bougher. 2018. *World Economics and Politics Dataverse.*
URL: *https://ncgg.princeton.edu/wep/dataverse.html*

Hanson, Jonathan K. and Rachel Sigman. 2021. "Leviathan's Latent Dimensions: Measuring State Capacity for Comparative Political Research." *Journal of Politics* 83(4):1–16.

Horn, Myron K. 2014. *Giant Oil and Gas Fields of the World.*
URL: *https://edx.netl.doe.gov/dataset/aapg-datapages-giant-oil-and-gas-fields-of-the-world*

Kentikelenis, Alexandros and Thomas Stubbs. 2023. *A Thousand Cuts: Social Protection in the Age of Austerity.* Oxford: Oxford University Press.

Marshall, Monty G. and Ted Robert Gurr. 2020. *Polity5: Political Regime Characteristics and Transitions, 1800–2018.*
URL: *http://www.systemicpeace.org/inscrdata.html*

Nguyen, Thanh Cong, Vítor Castro and Justine Wood. 2022. "A New Comprehensive Database of Financial Crises: Identification, Frequency, and Duration." *Economic Modelling* 108:105770.

The PRS Group. 2022. *International Country Risk Guide (ICRG).*

World Bank. 2023. *Data Updates and Errata.*
URL: *https://datahelpdesk.worldbank.org/knowledgebase/articles/906522-data-updates-and-errata*