When Countries Revise Their Data*

Iasmin Goes†

August 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 (BLS) 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 came as no surprise to experts: initial job estimates tend to rely on incomplete household and business surveys, and as survey funding and response rates decline over time, the uncertainty around initial estimates tends to increase. In relative terms, the magnitude of this revision was marginal: it meant that 159.2 million individuals, not 160 million individuals, were employed 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]

One year later, in August 2025, the BLS again published a downward revision of its preliminary employment data. That same day, President Trump fired the BLS commissioner, claiming on social media that the revisions had been "RIGGED in order to make the Republicans, and ME, look bad." Kevin Hassett, the director of the White House National Economic Council, complained that the BLS had been "revising numbers all over the place in a way that makes it so that I don't think anybody really can trust that the numbers are right."[2] These critical remarks were consistent with other decisions made by the Trump administration, which fired thousands of federal workers, deleted government datasets, and disbanded expert panels like the Federal Economic Statistics Advisory Committee.

The US is not unique in facing controversy over economic statistics. In 2009, Greece revised its planned deficit ratio from 3.7 to 12.5 percent of the Gross Domestic Product (GDP), triggering credit rating downgrades and a loss of market access that brought the country close to default (Aragão and Linsi, 2022). Between 2010 and 2014, Ghana, Kenya,

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

[2] Nick Niedzwiadek and Sam Sutton. "Trump Fires Statistics Chief After Soft Jobs Report." *Politico*. 1 August 2025.

Nigeria, and other African nations reported GDP increases of up to 60 percent after updating accounting methods and incorporating informal activities (Jerven and Ebo Duncan, 2012) — a shift 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). In 2016, Ireland's reported GDP growth of 26.3 percent (later traced to Apple's onshoring of intellectual property) drew ridicule as "leprechaun economics" (Polyak, 2023).

Measuring the economy is an iterative process: preliminary estimates are routinely revised to incorporate new data, improve methods, and correct mistakes (Carson, Khawaja and Morrison, 2004). Macroeconomic indicators like GDP, trade, inflation, 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. It might seem counterintuitive, but revisions are a statistical best practice, a natural and necessary component of the data production process. The International Monetary Fund (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. The resulting changes are typically minor technical adjustments.

Yet, as the case of the US shows, even revisions that are small in relative terms and seem like a technicality can have substantial political and economic consequences. Acknowledging past errors can undermine the government's reputation, imply a low commitment to transparency, trigger market instability, and erode public trust. The general public often misunderstands the purpose of revisions, and opposition parties capitalize on this misunderstanding. As Ghana, Kenya, and Nigeria show, even large *upward* revisions can be costly, weakening the government's bargaining position in international financial negotiations. Whether large or small, upward or downward, revision are both a technicality and a political liability. So what explains their likelihood and magnitude?

Producing official statistics is a two-stage process. First, governments collect and release preliminary estimates. Second, governments decide whether to revise the preliminary estimates, a choice shaped by the quality of the original data. I argue that the context of data *collection* is more important than the context of data *revision*. On average, countries with higher state capacity have better-trained statisticians, regular household surveys, and comprehensive administrative data systems that allow them to collect and report higher-quality data in a timely manner, setting the stage for future revisions. While state capacity provides the *technical ability* to detect and incorporate new information, governments must also be *willing* 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. The key point is not that democracies *want* to revise their data; rather, journalists, academics, and other experts *pressure* democracies to conduct revisions that are technically necessary but politically costly and would otherwise be overlooked.

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, requiring governments to improve data collection and disclose economic indicators in ways that conform to international standards. Even if governments comply reluctantly, data collected under IMF oversight 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 at the time of data collection are more likely to update (or "vintage") their data down the road.

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 with precision and minimal error, whereas reliable data provide consistent information across repeated measurements, across various sources and releases. I begin by reviewing a rich literature explaining why countries fail to disseminate valid data: due to low statistical capacity, high political interference, and poor data management. Moving on to reliability, I investigate what drives revisions to previously disseminated data — a less studied phenomenon[3] for which no systematic explanations exist. After developing a theory of data revisions, I find support for my expectations in an analysis of GDP data published by the World Development Indicators (WDI) between 1994 and 2021 and conclude by discussing research and policy implications.

My findings speak to the potential trade-off between validity and reliability, the credibility risks associated with data revisions, and the understudied consequences of revisions to the same data source. Indeed, while inconsistencies *across data sources* have been widely documented (Linsi, Burgoon and Mügge, 2023; Weikmans and Roberts, 2019; Ram and Ural, 2014; Kerner, 2014; Johnson et al., 2013; Michaelowa and Michaelowa, 2011; Amin Gutiérrez de Piñeres, 2006; Pellechio and Cady, 2006), I show that inconsistencies *across different versions of the same source* may be even more consequential.

2 Recording and Revising Economic Data

2.1 Recording Valid Data

Macroeconomic indicators often lack validity — they fail to capture the "correct" information with precision and minimal error — due to low statistical capacity, high political incentives,

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

and inadequate data management.[4] In terms of statistical capacity, national statistical offices (NSOs) are often underfunded, understaffed, use outdated methods, experience frequent turnover, and rely on outdated information about businesses and households. Many of their employees lack basic statistical competence or familiarity with the System of National Accounts (SNA), a global standardization framework (United Nations Economic Commission for Africa, 2005; Mejía Guerra et al., 2023). Informal sector activity — which accounts for up to 44 percent of the GDP in the developing world (Coyle, 2014, 110) — is difficult to measure (Olinto Ramos, Pastor and Rivas, 2008; Coyle, 2014). All this limits NSOs' ability to collect, standardize, and disseminate high-quality data.

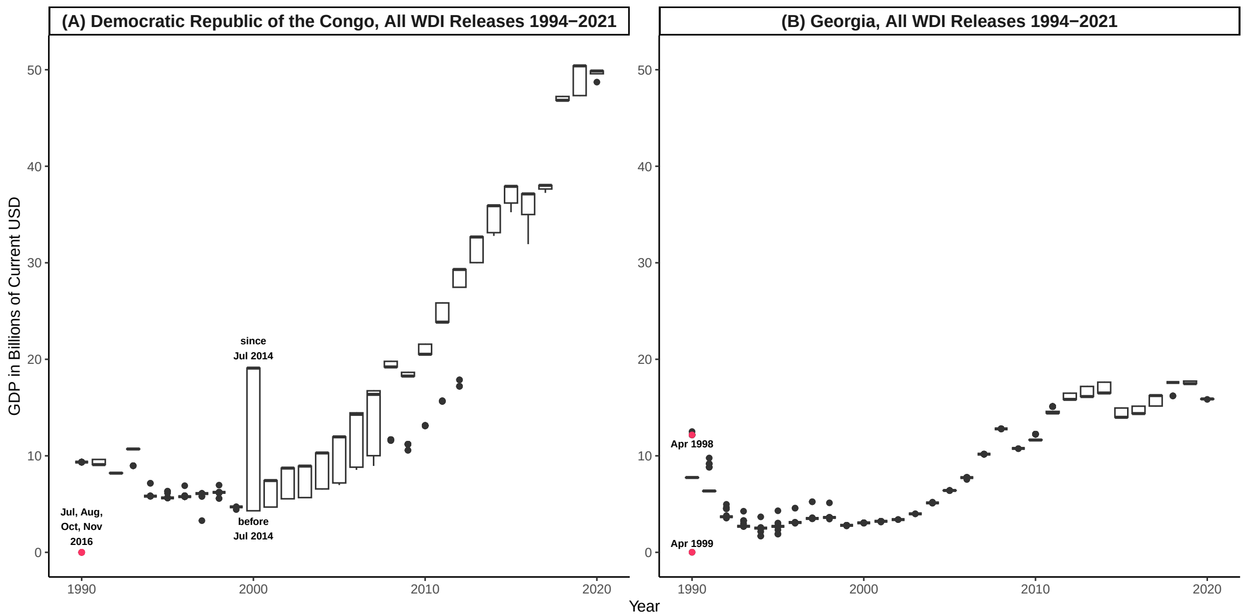
In terms of political incentives, autocracies are less likely to report policy-relevant data (Hollyer, Rosendorff and Vreeland, 2011) and may overstate growth rates or underreport COVID-19 deaths (Martínez, 2022; Magee and Doces, 2015; Wallace, 2014; Adiguzel, Cansunar and Corekcioglu, 2020). Some NSO directors are political appointees who lack autonomy and can be dismissed at will, generating incentives to misreport data. Returning to this study's opening example: after the BLS reported unfavorable employment data in August 2025, President Trump dismissed the BLS commissioner and appointed a replacement who proposed ending the agency's monthly jobs report.[5] Fiscal transfers, aid dependence, climate-conscious electorates, and EU rules create incentives to inflate population figures, understate economic growth, overstate climate aid, and violate deficit limits, respectively (Devarajan, 2013; Kerner, Jerven and Beatty, 2017; Michaelowa and Michaelowa, 2011; Alt, Lassen and Wehner, 2014). Conversely, political competition and frequent turnover increase uncertainty about the future, motivating incumbents to pass Freedom of Information (FOI) laws that enhance transparency (Berliner, 2014).

Idiosyncratic data management errors pose a final threat to validity. As Figure 1 shows, four different WDI vintages report the Democratic Republic of the Congo's 1990 GDP as

[4] Appendix C provides detailed descriptive evidence of data collection challenges across several countries.

[5] Natalie Sherman. "Trump's Pick to Lead Economic Data Agency Floats Ending Monthly Jobs Report." *BBC.* 12 August 2025.

Figure 1: Current GDP of the Democratic Republic of the Congo and Georgia, 1990–2020

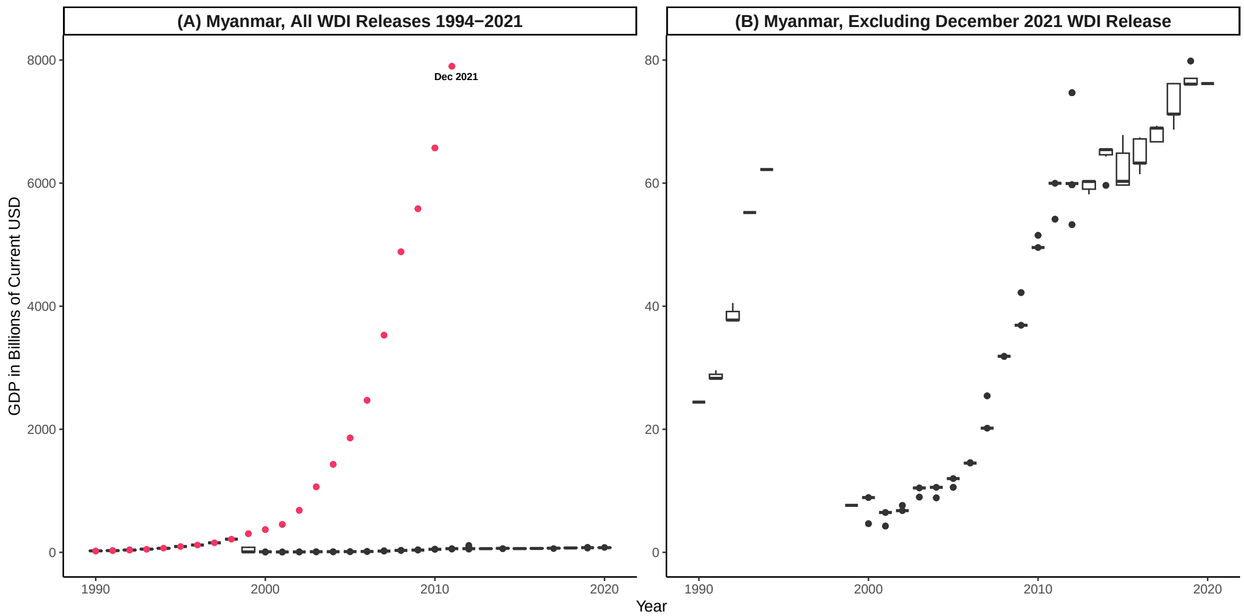


These boxplots present the distribution of current GDP estimates from 1990 to 2020 for (A) the Democratic Republic of the Congo and (B) Georgia, using data drawn from the 104 WDI releases from April 1994 to December 2021. Four WDI releases (in pink) reported a GDP of zero for the Democratic Republic of the Congo in 1990. 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 WDI vintages and 12.1707 *billion* in others (in pink). Section 4 discusses the data in detail.

zero; two other releases report it as missing. Georgia's 1990 GDP — reported to be around 12.1707 *billion* current US dollars until April 1998 — "lost" three digits in the April 1999 and April 2000 vintages, shrinking to 12.1707 *million*.[6] In the December 2021 update, Myanmar's 2011 GDP was over 100 times higher than in any other vintage: 7.899 *trillion*, versus 54 to 59 *billion* (Figure 2). Other trivial errors include swapped country names, mislabelled columns, and entire indicators mistakenly blanked (see World Bank 2023 for errata). The most plausible explanation for these singular discrepancies is a data management error — perhaps the worst kind of error, as it is impossible to predict.

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

Figure 2: 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.

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.[7] About 39.8 percent of all economies surveyed by the IMF in 2020 first released 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

[7]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 days of each quarter's end using incomplete data.

recommended.

Revisions can be information-based or structural (Croushore and Stark, 2003). Information-based revisions occur when countries incorporate newly available data, allowing for more precise measurement. For instance, Greece revised its deficit ratio in 2009 after "discovering" new social security and military expenses. Structural data revisions occur when there are updates to the definition of concepts, the base year, or the aggregation method. 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). For quarterly data, at least, information-based revisions are most common up to one year after the initial data release. The more time has elapsed since the original data release, the more structural the nature of revisions (Croushore and Stark, 2003).

To illustrate the prevalence of such revisions, consider the IMF-produced Report on the Observance of Standards and Codes (ROSC) Data Module,[8] which reviews countries' data capabilities at their 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 module, including high-income democracies like France and Norway, but also autocracies at various income levels, such as Belarus and Tajikistan. At face value, many governments appear 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). "Even preliminary data, with the understanding that these are subject to revisions, would be useful" (Uruguay,

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

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. Admittedly, this assumption could be wrong: revisions could reflect a *decrease* in validity. Governments could be revising their data in bad faith, introducing errors into accurate preliminary data. While hard to observe directly, evidence shows that this is not typically the case: today's ill-intentioned authorities do not tend to cook the books ex post (Guriev and Treisman, 2019). Instead of retroactively replacing "correct" information with fictitious information, they withhold statistics (Venezuela since 2015), postpone the initial release (Zimbabwe in 2019), or release doctored numbers from the outset (Argentina in 2007–2008). Already-published statistics might be erased from public records (as in the US in 2025), but do not tend to be replaced with falsified information.

Revisions highlight a trade-off between validity and reliability. Governments that prioritize validity will update their data each time better information and updated measurements become available, even if this comes at the expense of reliability. Governments that prioritize reliability may sacrifice validity, keeping numbers unchanged to avoid any perception of inconsistency. 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 the 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 revisions. On the one hand, they enhance credibility, signaling a commitment to improving already valid data in line with international standards. On the other hand, they can signal deep institutional problems, including a lack of "independence, integrity and accountability of the national statistical authorities" (according to the aforementioned European Commission report). This holds true not only for large revisions (like Greece in 2009) but also for comparatively minor ones (like the US in 2025). Statistical officers face a delicate balance: they must release information quickly and update it regularly, knowing that such updates can undermine the NSO's reputation and erode public trust by suggesting low reliability and low validity. Given this challenge, it is not obvious 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

Producing official statistics is a two-stage process. The first stage is the initial data collection and dissemination. The second stage is the decision to revise (or "vintage") previously published data, conditional on the first stage. This second-order decision is shaped both by the validity of the initial information and by domestic or international incentives to adjust it. Understanding the drivers of revisions requires attention to both stages, but particularly the first: when collecting the initial data, states can adopt strategies that enable future revisions. I argue that this initial decision, influenced by experts, academics, journalists, and foreign creditors, is difficult to reverse, as it shapes the scope and feasibility of any subsequent corrections.

In the first stage, state capacity sets the foundation. Countries with higher state capacity can train and retain skilled 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). These capabilities enhance the quality of initial data, making subsequent revisions easier: in the second stage, high-capacity countries are better able to detect and correct errors. At the same time, they are less likely to make large errors in the first place, so most revisions should consist of smaller adjustments. Thus, higher capacity should increase the likelihood of revisions without a systematic effect on the relative magnitude of revisions.

Hypothesis 1a: Higher state capacity in the first stage increases the odds of revisions in the second stage.

Hypothesis 1b: Higher state capacity in the first stage has no systematic effect on the magnitude of revisions in the second stage.

While Hypotheses 1a and 1b might appear self-evident, they outline a necessary — not sufficient — condition for revisions. Several high-income nations that could afford to release and revise their data choose not to (Williams, 2009). A report by the African Development Bank (2013, 5) underscores this point: "rich countries are not guaranteed to have good statistics; neither are poor countries condemned to have bad data ... Achieving high-quality data entails a political choice and a firm commitment to invest in statistics that will support informed evidence-based decisions."

Democracies and constrained executives release more data (Hollyer, Rosendorff and Vreeland, 2011; Williams, 2009) and of higher validity (Magee and Doces, 2015; Martínez, 2022). 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. Democracy is associated not only with more and better data, but also — I argue — with a greater likelihood of data revisions. Beyond the *ability* to revise, countries must be *compelled* to do so. Democracies enable the work of experts who serve precisely this function: they are accountability agents that pressure for regular revisions in both stages, no matter how high the political cost.

Democracies are more willing 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. Put simply, democracies release data that are easier to "correct" in the second stage. 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. One possibility is that democracies face such intense transparency pressures that they rush to release incomplete data, creating a greater need for subsequent revisions. Another possibility is that they are more cautious and initially publish understated estimates (as opposed to dictators, who — according to Martínez (2022) — tend to exaggerate their numbers). Either way, democratic leaders have limited control over revisions, which will be as frequent and as large as experts deem necessary.

In contrast, autocracies 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. Autocracies like Azerbaijan, Belarus, Oman, and Tajikistan might recognize the instrumental value of data transparency, having requested multiple ROSC Data Modules in the past. Yet their data releases 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). "Informational autocrats" gain legitimacy by disseminating selective information about economic successes while concealing information about their economic failures (Guriev and Treisman, 2019). They prioritize control over timeliness, delaying or avoiding publication of data deemed politically unsafe. 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. When freedom of speech is limited, journalists and academics may be hesitant to question official statistics or pressure for corrections. Overall, I expect regime type to affect not only the probability but also the magnitude of revisions.

Hypothesis 2a: Higher democracy levels in the first stage increase the odds of revisions in the second stage.
Hypothesis 2b: Higher democracy levels in the first stage increase the magnitude of revisions in the second stage.

This two-stage theory applies to domestic, but also to foreign accountability agents. In the first stage, IMF borrowers are more likely to disseminate data (Hollyer, Rosendorff and Vreeland, 2011). I expect data disseminated by IMF borrowers to be more prone to revisions in the second stage. 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 were met before subsequent loan tranches can be disbursed. As Kentikelenis and Stubbs (2023) show, these conditions include: "develop a monitoring system to verify the quality of the accounting data ... in terms of data consistency and accuracy" (Brazil, 1998); create "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). To comply with reporting requirements and secure the timely disbursement of funds, IMF borrowers might feel pressured to quickly publish preliminary figures that require later revision.[9] Additionally, loan agreements are often attached to technical assistance that improve borrowers' data collection standards and fill statistical gaps, in turn enabling subsequent revisions.

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 signal economic mismanagement and jeopardize the borrower's credibility. 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, borrowers have at least some control over the revision process; they can make minor revisions to satisfy reporting requirements without drastically altering past figures, minimizing potential political or economic fallout. Therefore, as with state capacity, the expected effect of IMF participation on the magnitude of revisions is ambiguous: while external oversight can push for comprehensive updates, governments under IMF programs likely prefer minor, incremental adjustments to avoid reputational costs.

Hypothesis 3a: IMF program participation in the first stage increases the odds of revisions in the second stage.

[9] Voluntary multilateral initiatives like the SDDS (see footnote 7) are associated with greater information disclosure (Vadlamannati, Cooray and Brazys, 2018), partly due to technical assistance. While SDDS compliance could also drive revisions, countries that are already inclined to update their data may self-select into the SDDS. I expect the IMF to have more leverage when direct money is on the line.

Hypothesis 3b: IMF program participation in the first stage has no systematic effect on the magnitude of revisions in the second stage.

Importantly, my theory builds on research about data *validity* but is agnostic about its effects on data *reliability*. Some countries might revisit their data because they rushed to release low-quality preliminary information that must be corrected ex post. Others might only need to fine-tune their data, having released high-quality information from the start. Whatever the underlying, unobserved validity, I argue that revisions are best explained by contextual factors at the first stage (when data are initially collected), not the second stage (when data are potentially revised). Conditional on state capacity, foreign and domestic accountability pressures generate incentives to release transparent information from the start, enabling subsequent revisions.

4 Operationalizing 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.[10]

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).

[10] 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.

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 releases, not across countries or over time, as it does not make PPP or inflation adjustments.[11]

Recall the two stages of data production: first, the reference year, when country c collected and disseminated its data ($y = 1990, \dots, 2020$), and second, the WDI vintage, when the same country may revise its data ($v =$ April 1994, September 1994, ..., December 2021). Figure 3 illustrates this structure using GDP in billions of current US dollars for Myanmar as an example. In this figure, each WDI vintage corresponds to a separate spreadsheet containing all available reference years.[12]

For country c and reference year y, I define a revision as any change in GDP between two consecutive WDI vintages, $v - 1$ and v. Revisions can reflect an addition, deletion, or change of any magnitude, in any direction. As Figure 3 illustrates, Myanmar's 1992 GDP was reported as 37.7 billion in April 1994, 37.7 billion in September 1994 (*Revision = 0*), missing in April 1995 (*Revision = 1*), and 40.5 billion in September 1995 (*Revision = 1*).

[11] *GDP, PPP (current international $)* (ID `NY.GDP.MKTP.PP.CD`) allows for comparisons across countries, not across releases, 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 releases.

[12] To ensure that each reference year has had time to be compiled and published, the analysis includes only vintages released at least 90 days after the end of the reference period. For example, for the reference year 2020, this corresponds to all WDI releases starting in April 2021. I use 90 days as a cutoff because 39.8 percent of all economies surveyed by the IMF in 2020 reported annual GDP data within 90 days of the reference period (Baer, Guerreiro and Silungwe, 2022, 17).

Figure 3: Illustration of the Data Structure, Using the Case of Myanmar

Vintage: April 1994

Country	Year	GDP
Myanmar	1990	24.4
Myanmar	1991	28.3
Myanmar	1992	37.7
...	...	

Vintage: September 1994

Country	Year	GDP
Myanmar	1990	24.4
Myanmar	1991	28.3
Myanmar	1992	37.7
...	...	

Vintage: April 1995

Country	Year	GDP
Myanmar	1990	NA
Myanmar	1991	NA
Myanmar	1992	NA
...	...	

Vintage: September 1995

Country	Year	GDP
Myanmar	1990	24.4
Myanmar	1991	29.6
Myanmar	1992	40.5
...	...	...

...

This figure uses the case of Myanmar to illustrate the three-dimensional structure of the data: the unit of analysis is country c, year y, and vintage v. In this notation, $y = 1990, \ldots, 2020$ refers to the first stage, when the data are initially collected and disseminated, whereas $v =$ April 1994, September 1994, $\ldots$, December 2021 refers to the second stage (i.e., each subsequent WDI release), when previously published data may be revised.

Revisions occur in 17.2 percent of all observations.

If a revision occurs and no missing values are involved, I also examine its relative size, ranging from a 100 percent reduction (the Democratic Republic of Congo, Figure 1) to a 17,488 percent increase (Myanmar, Figure 2). The average revision is relatively small: 3.6 percent. Since the focus is on magnitude rather than direction, I use the absolute value of percentage changes, logged to reduce the influence of extreme values. Again, using Figure 3 as an illustration: since Myanmar's 1992 GDP did not change in September 1994 relative to April 1994, the value of *Absolute % Change (Log)* for Myanmar-1992-September 1994 is zero.

4.2 Empirical Strategy

I estimate two sets of models by maximum likelihood. First, I use logistic regressions to capture the probability that a given country-year observation (say, Myanmar-1992) is revised between two consecutive WDI vintages:

$$\Pr\left(\text{Revision}_{c,y,v} = 1\right) = \text{logit}^{-1}(\alpha + \mathbf{X}_{c,y,v}\beta + u_c + u_y + u_v) \tag{1}$$

where $\text{Revision}_{c,y,v}$ equals one if the GDP value for country c in reference year y differs between vintage $v-1$ and vintage v; $\mathbf{X}_{c,y,v}$ denotes a set of independent variables (described in the next section); and u_c, u_y, and u_v are random intercepts for country, year, and vintage, respectively.

Second, conditional on a revision occurring, with no missing values involved, I use linear regressions to estimate its magnitude, measured as the logged absolute percentage change:

$$\text{Absolute \% Change (Log)}_{c,y,v} = \alpha + \mathbf{X}_{c,y,v}\beta + u_c + u_y + u_v + \varepsilon_{c,y,v}, \tag{2}$$

where $\varepsilon_{c,y,v}$ is an idiosyncratic error term. This two-step approach allows me to separately model whether a revision occurs and, if so, its size.[13]

4.3 Independent Variables

My theory focuses on three determinants of data revisions: state capacity, regime type, and IMF program participation. Regarding state capacity, Hanson and Sigman (2021) use Bayesian latent variable analysis to combine 21 indicators of extractive, coercive, statistical, informational, and administrative capacity into a single index. Of these, statistical or informational capacity are the most relevant dimensions, but existing measures have limited

[13] I model the two steps separately because Tobit or Heckman selection models assume independent observations and cannot incorporate random intercepts.

coverage.[14] For this reason, I rely on V-Dem index *Rigorous Public Administration* as the main proxy for capacity. This measure ranges from 0 (the law is not respected by public officials) to 4 (the law is generally fully respected by public officials). While it primarily captures bureaucratic impartiality and rule-following rather than statistical expertise, it has the broadest temporal and spatial coverage.

Conditional on a country's ability to collect high-quality data, domestic accountability agents increase the likelihood and magnitude of revisions. I measure domestic accountability using V-Dem's *Polyarchy* index (Coppedge et al., 2023), which captures the quality of electoral democracy (including extensive suffrage, fair elections, freedom of expression, and access to information) on an ordinal scale from 0 to 1. Higher values indicate more democratic regimes, associated with more domestic accountability pressures.

Why are democracies more subject to public scrutiny? The key mechanism I propose is *Freedom of Academic Expression*: in the absence of censorship and intimidation, experts can scrutinize official statistics and demand transparent data practices. This V-Dem index ranges from 0 (not respected by public authorities) to 4 (fully respected by public authorities). While this is my main hypothesized pathway, I also assess alternative mechanisms, such as the *Political Corruption Index* (from 0 to 1, with larger values indicating more corruption) and the population share with *Suffrage*, both reported by V-Dem. Additionally, I borrow three measures from Berliner (2014). *New Democracy* equals one during 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 executive party control over the preceding five years (both calculated using the Database of Political Institutions, Cruz, Keefer and Scartascini 2021).

In terms of foreign accountability, I argue that IMF borrowers tend to be scrutinized

[14]For example, the World Bank's Statistical Performance Indicators provide information about the Balance of Payments manual in use, but only for developing countries and only after 2004 (Dang et al., 2023). Likewise, Brambor et al.'s (2020) measure of information capacity is only available for 85 countries and only until 2012.

more closely, creating external pressure to revise and refine macroeconomic data (without a corresponding effect on the magnitude of revisions). Therefore, models include a dichotomous indicator of *IMF Program* participation (Kentikelenis and Stubbs, 2023). Other important events may prompt data revisions: compliance with voluntary data disclosure initiatives like the SDDS, a financial crisis, natural disaster, or armed conflict (using data from the IMF Dissemination Standards Bulletin Board, Nguyen, Castro and Wood 2022, the Centre for Research on the Epidemiology of Disasters 2020, and Gleditsch et al. 2002, respectively).

Since GDP is reported in billions of current US dollars, the conversion from local currency may itself introduce discrepancies. To account for this potential source of error, 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 that might trigger ex post revisions. *SNA Change* takes the value of one in years when countries adopted or updated the SNA in use, using data from the UN National Accounts Statistics (complemented by WDI Metadata and IMF International Financial Statistics). *Data Management Error* takes the value of one for documented reporting mistakes (like those in Figures 1 or 2; see Appendix C for a discussion of such cases). Lastly, *Difference Between Vintage and Year* measures the gap (in years) between the first and the second stages, $v - y$; larger values might correlate with fewer revisions, since information-based revisions are most common up to one year after the initial data release (Croushore and Stark, 2003).

Table 1 summarizes all independent variables and their underlying concepts. Some fac-

Table 1: Summary of Independent Variables

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

tors, such as IMF programs and financial crises, are included for the first *and* second stages, y and v, because they may affect both initial reporting and subsequent revisions. Others, like state capacity and regime type, change slowly over time and are included only for the first stage y to avoid multicollinearity and reverse causality.[15] Certain variables, such as SNA changes and idiosyncratic errors, are vintage-specific (v). Besides *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 because they are revised in parallel with GDP.

[15]For example, *Rigorous Public Administration* in stage 1 and *Rigorous Public Administration* in stage 2 are correlated at $\rho = 0.8325, p = 0.000$. *Polyarchy* in stage 1 and *Polyarchy* in stage 2 are correlated at $\rho = 0.879$ ($p = 0.000$). In addition, using these variables at stage 1 could introduce reverse causality; for example, a country's willingness or ability to release data at step 1 may affect its observed state capacity at step 2.

5 Determinants of Data Revisions

5.1 Main Results

In Table 2, Models 1 and 2 are logistic regressions with the dichotomous outcome *Revision*. Consistent with Hypothesis 1a, higher capacity (*Rigorous Public Administration*) in the first stage is associated with higher likelihood of subsequent revisions. Conditional on state capacity, *Polyarchy* has a positive and significant effect on the outcome: democratic institutions encourage not only the initial release of transparent statistics (Hollyer, Rosendorff and Vreeland, 2014), but also the sustained pressure to later revise them. This supports Hypothesis 2a.

In lieu of a democracy index, Model 2 examines individual aspects of democracy likely to drive data revisions. Of these, only *Freedom of Academic Expression* has a positive and significant effect. Although *Suffrage* has been linked to information capacity (Brambor et al., 2020), it *reduces* the odds of a data revision. *New Democracy* (indicating a democratic transition in the previous five years) has no significant effect on revisions, which is consistent with the proposed mechanism: expert scrutiny depends on institutionalized freedoms that build over time, rather than emerging immediately after a regime change. Widespread suffrage, democratization, frequent turnover, and a strong opposition allow citizens to hold governments accountable, but scrutinizing technical data depends more on specialized knowledge than on general democratic pressures. 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*: on average, program participation ahead of data collection (first stage) increases the odds of subsequent revisions by 5 to 8 percent, in line with Hypothesis 3a.

Models 3 and 4 turn to the *magnitude* of revisions, restricting the sample to instances when *Revision = 1* (hence the smaller number of observations). Now, state capacity has no systematic effect on the outcome, which is consistent with Hypothesis 1b. Rather, an increase

Table 2: Determinants of the Likelihood and Magnitude of Data Revisions

	Dependent Variable:			
	Revision = 1		Abs. % Change (Log)	
	(1)	(2)	(3)	(4)
Rigorous Public Administration, y	0.07***	0.06***	0.09	−0.01
	(0.02)	(0.02)	(0.08)	(0.08)
Polyarchy, y	0.71***		1.34***	
	(0.08)		(0.34)	
Freedom of Academic Expression, y		0.18***		0.23***
		(0.01)		(0.06)
Suffrage, y		−0.62***		0.13
		(0.09)		(0.40)
New Democracy, Prev. 5 Yrs, y		−0.04		−0.16
		(0.03)		(0.12)
Opposition Strength, y		−0.00		−0.00**
		(0.00)		(0.00)
Turnover Frequency, Prev. 5 Yrs, y		0.01		0.06
		(0.01)		(0.06)
Political Corruption Index, y		0.10		−1.51***
		(0.09)		(0.38)
IMF Program, y	0.07***	0.05***	−0.13*	−0.07
	(0.02)	(0.02)	(0.08)	(0.08)
IMF Program, v	−0.07***	−0.07***	−0.37***	−0.38***
	(0.02)	(0.02)	(0.09)	(0.09)
SDDS Compliance, y	−0.27***	−0.27***	−0.06	−0.16
	(0.02)	(0.02)	(0.11)	(0.11)
SDDS Compliance, v	0.16***	0.20***	0.54***	0.56***
	(0.04)	(0.04)	(0.15)	(0.15)
Financial Crisis, y	0.05***	0.07***	−0.17**	−0.15*
	(0.02)	(0.02)	(0.08)	(0.08)
Financial Crisis, v	0.07***	0.06***	0.55***	0.58***
	(0.02)	(0.02)	(0.09)	(0.09)
Natural Disaster, y	0.05***	0.06***	−0.12*	−0.09
	(0.02)	(0.02)	(0.07)	(0.07)
Natural Disaster, v	0.02	0.03	−0.11	−0.10
	(0.02)	(0.02)	(0.08)	(0.08)
Armed Conflict, y	−0.02	0.00	−0.36***	−0.32**
	(0.03)	(0.03)	(0.13)	(0.14)
Armed Conflict, v	−0.08***	−0.08**	0.27**	0.33**
	(0.03)	(0.03)	(0.14)	(0.14)
Diff. Between Official and Alt. XR, v	−0.01**	−0.01**	0.00	0.00
	(0.01)	(0.01)	(0.03)	(0.03)
SNA Change, v	0.20***	0.21***	1.68***	1.73***
	(0.03)	(0.03)	(0.12)	(0.12)
Data Management Error, v	6.87***	6.91***	12.14***	12.09***
	(0.62)	(0.62)	(1.51)	(1.55)
Diff. Between Vintage and Year, $v - y$	−0.11***	−0.11***	−0.24***	−0.24***
	(0.01)	(0.01)	(0.01)	(0.01)
Intercept	−3.49***	−3.04***	−5.91***	−4.93***
	(0.48)	(0.49)	(0.64)	(0.78)
Observations	397,803	387,212	62,916	61,551
Log Likelihood	−87,123.61	−84,814.20	−21,0649.32	−20,6053.37
Number of Countries	170	167	170	167
Number of Years	31	31	31	31
Number of Vintages	103	103	78	78
Variance: Countries (Intercept)	0.29	0.31	2.88	2.73
Variance: Years (Intercept)	0.14	0.14	0.15	0.16
Variance: Vintages (Intercept)	19.95	19.88	25.02	24.74

This table presents the results of two logistic regressions (Models 1 and 2) and two linear regressions (Models 3 and 4) with random intercepts for country, year, and vintage. y is the reference year, when the data are initially collected and disseminated. v is the vintage, when previously published data may be revised. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

in democracy levels is associated with a significant increase in *Absolute % Change (Log)*, as Hypothesis 2b predicts. 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 contrast, 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. Indeed, when Model 4 decomposes the effects of regime type on magnitude of revisions, it finds 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 (Model 2), and such revisions tend to have a larger magnitude (Model 4). Relatedly, more corruption is associated with smaller revisions, as indicated by the negative coefficient for the *Political Corruption Index*.

As expected, IMF program participation — whether at the time of original data collection (y) or during posterior data revision (v) — has a different effect than before. IMF programs increase the odds of revisions without systematically increasing their size. One interpretation is that borrowers prefer to make incremental revisions that comply with reporting requirements without jeopardizing their economic stability or political reputation. For brevity, I do not discuss the control variables in detail, but they behave largely as expected; for instance, SNA changes and data management errors are associated with higher odds of revisions and revisions of larger magnitude.

5.2 Robustness Checks

I conduct additional tests that correspond to each part of the argument. First, I examine whether the results hold under alternative measures of state capacity (Hypotheses 1a and 1b). My main proxy for state capacity does not directly capture statistical or informational expertise. To address this concern, Table 3 presents models that replace *Rigorous Public*

Table 3: Determinants of the Likelihood and Magnitude of Data Revisions: Alternative Measures of State Capacity

	Dependent Variable:			
	Revision = 1		Abs. % Change (Log)	
	(1)	(2)	(3)	(4)
State Capacity, y	0.07**		0.02	
	(0.03)		(0.13)	
Bureaucratic Quality, y		0.07***		0.09
		(0.01)		(0.07)
Polyarchy, y	0.76***	0.74***	1.41***	1.33***
	(0.08)	(0.08)	(0.33)	(0.35)
IMF Program, y	0.05***	0.05***	−0.12	−0.06
	(0.02)	(0.02)	(0.08)	(0.09)
IMF Program, v	−0.09***	−0.07***	−0.40***	−0.28***
	(0.02)	(0.02)	(0.09)	(0.10)
SDDS Compliance, y	−0.26***	−0.25***	0.02	−0.17
	(0.03)	(0.03)	(0.11)	(0.12)
SDDS Compliance, v	0.11***	0.23***	0.61***	1.05***
	(0.04)	(0.04)	(0.16)	(0.17)
Financial Crisis, y	0.05***	0.03	−0.18**	−0.16*
	(0.02)	(0.02)	(0.08)	(0.09)
Financial Crisis, v	0.07***	0.02	0.62***	0.38***
	(0.02)	(0.02)	(0.09)	(0.10)
Natural Disaster, y	0.05***	0.07***	−0.12	−0.12
	(0.02)	(0.02)	(0.07)	(0.08)
Natural Disaster, v	0.05***	0.05***	−0.06	−0.09
	(0.02)	(0.02)	(0.08)	(0.09)
Armed Conflict, y	−0.04	−0.07*	−0.40***	−0.18
	(0.03)	(0.04)	(0.14)	(0.16)
Armed Conflict, v	−0.10***	−0.21***	0.33**	−0.61***
	(0.03)	(0.03)	(0.14)	(0.16)
Diff. Between Official and Alt. XR, v	−0.01**	−0.01**	0.00	0.00
	(0.01)	(0.01)	(0.03)	(0.03)
SNA Change, v	0.22***	0.16***	1.71***	1.40***
	(0.03)	(0.03)	(0.12)	(0.13)
Data Management Error, v	6.95***	6.92***	13.68***	11.87***
	(0.62)	(0.61)	(1.64)	(1.55)
Diff. Between Vintage and Year, $v - y$	−0.08***	−0.11***	−0.23***	−0.25***
	(0.00)	(0.01)	(0.01)	(0.01)
Intercept	−3.73***	−3.49***	−5.90***	−5.95***
	(0.48)	(0.47)	(0.65)	(0.64)
Observations	367,567	312,598	58,486	49,992
Log Likelihood	−80,262.22	−68,241.24	−195,954.25	−167,336.77
Number of Countries	163	136	163	136
Number of Years	27	31	27	31
Number of Vintages	103	103	76	78
Variance: Countries (Intercept)	0.33	0.34	3.01	2.98
Variance: Years (Intercept)	0.03	0.15	0.16	0.14
Variance: Vintages (Intercept)	20.79	19.33	25.31	21.87

This table presents the results of two logistic regressions (Models 1 and 2) and two linear regressions (Models 3 and 4) with random intercepts for country, year, and vintage. y is the reference year, when the data are initially collected and disseminated. v is the vintage, when previously published data may be revised. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Administration with Hanson and Sigman's *State Capacity* index (Models 1 and 3) or with the International Country Risk Guide (ICRG) measure of *Bureaucratic Quality* (Models 2 and 4). In both cases, higher values indicate higher capacity. While these alternative measures are also imperfect and cover fewer countries and years, they allow me to triangulate the effect of state capacity. The results remain robust across these specifications.

Second, I reports models with alternative measures of regime type, ensuring that the observed effects are not driven by a particular measure of democracy (Hypotheses 2a and 2b). Table 4 replaces *Polyarchy* with *Polity 2* (Marshall and Gurr 2020, Models 1 and 3) and the *Freedom House* index (Models 2 and 4). As with *Polyarchy*, larger values of *Polity 2* reflect more democratic regimes. *Freedom House* typically ranges from 1 (most free) to 7 (least free); I flip this scale so its interpretation mirrors that of the other regime type variables, with larger values reflecting more freedom. Again, the results are robust to these alternative measures.

Finally, I estimate a staggered difference-in-differences (DiD) design to probe Hypotheses 3a and 3b. This design compares countries before and after entering an IMF program to countries without any program. Since IMF programs can themselves affect bureaucratic capacity (Reinsberg et al., 2019) and democracy (Nelson and Wallace, 2017), the main models could be conflating or absorbing the direct effect of program participation on data revisions. The DiD isolates this channel by leveraging the staggered and reversible nature of IMF agreements, which countries enter and exit at different times. I summarize this approach here, with full results and diagnostic tests in Appendix D.

Since DiD estimators require a two-dimensional panel (unit-time), I collapse the country-year-vintage data into *% Revisions*. For country c and reference year y, this outcome indicates the percentage of all vintages that included a revision. Using Figure 3 as a guide, this corresponds to aggregating all values for Myanmar's 1992 GDP, across all available spreadsheets. This outcome captures the *frequency* of revisions, not their *likelihood* or *magnitude*, but it provides a comparable measure for identification. The treatment is a binary indicator of *IMF Program* participation at y (the first stage). I use Liu, Wang and Xu's (2024) fixed effects counterfactual estimator to estimate the following two-way fixed effects (TWFE) model:

$$\%\ \text{Revisions}_{c,y} = \text{IMF Program}_{c,y}\,\beta + \alpha_c + \gamma_y + \epsilon_{c,y}, \tag{3}$$

Table 4: Determinants of the Likelihood and Magnitude of Data Revisions: Alternative Measures of Regime Type

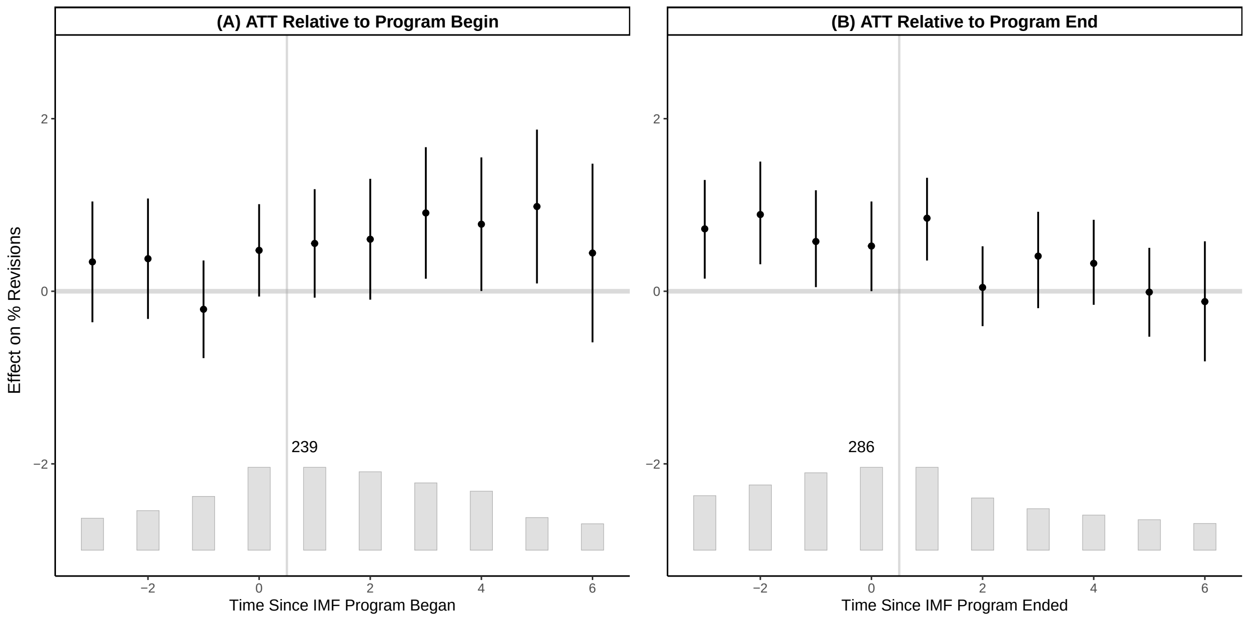
	Dependent Variable:			
	Revision = 1		Abs. % Change (Log)	
	(1)	(2)	(3)	(4)
Rigorous Public Administration, y	0.09***	0.10***	0.17**	0.18**
	(0.02)	(0.02)	(0.07)	(0.07)
Polity 2, y	0.02***		0.03**	
	(0.00)		(0.01)	
Freedom House, y		0.07***		0.09**
		(0.01)		(0.05)
IMF Program, y	0.04**	0.08***	−0.10	−0.11
	(0.02)	(0.02)	(0.08)	(0.08)
IMF Program, v	−0.08***	−0.07***	−0.38***	−0.38***
	(0.02)	(0.02)	(0.09)	(0.09)
SDDS Compliance, y	−0.23***	−0.29***	0.02	−0.08
	(0.02)	(0.02)	(0.11)	(0.11)
SDDS Compliance, v	0.14***	0.16***	0.59***	0.56***
	(0.04)	(0.04)	(0.15)	(0.15)
Financial Crisis, y	0.04**	0.05***	−0.18**	−0.17**
	(0.02)	(0.02)	(0.08)	(0.08)
Financial Crisis, v	0.05***	0.07***	0.58***	0.55***
	(0.02)	(0.02)	(0.09)	(0.09)
Natural Disaster, y	0.06***	0.06***	−0.14*	−0.11
	(0.02)	(0.02)	(0.07)	(0.07)
Natural Disaster, v	0.04**	0.02	−0.09	−0.10
	(0.02)	(0.02)	(0.08)	(0.08)
Armed Conflict, y	−0.02	−0.02	−0.35**	−0.36***
	(0.03)	(0.03)	(0.14)	(0.14)
Armed Conflict, v	−0.09***	−0.09***	0.37***	0.26*
	(0.03)	(0.03)	(0.14)	(0.14)
Diff. Between Official and Alt. XR, v	−0.01**	−0.01**	0.00	0.00
	(0.01)	(0.01)	(0.03)	(0.03)
SNA Change, v	0.22***	0.20***	1.76***	1.68***
	(0.03)	(0.03)	(0.12)	(0.12)
Data Management Error, v	6.91***	6.88***	12.55***	12.16***
	(0.62)	(0.62)	(1.54)	(1.51)
Diff. Between Vintage and Year, $v - y$	−0.10***	−0.11***	−0.23***	−0.24***
	(0.01)	(0.01)	(0.01)	(0.01)
Intercept	−3.35***	−2.95***	−5.66***	−5.08***
	(0.48)	(0.48)	(0.64)	(0.67)
Observations	376,061	397,331	59,745	62,858
Log Likelihood	−82,123.87	−87,035.02	−200,025.51	−210,457.05
Number of Countries	164	170	164	170
Number of Years	30	31	30	31
Number of Vintages	103	103	78	78
Variance: Countries (Intercept)	0.35	0.30	3.05	2.99
Variance: Years (Intercept)	0.09	0.13	0.16	0.15
Variance: Vintages (Intercept)	20.05	19.91	25.11	24.97

This table presents the results of two logistic regressions (Models 1 and 2) and two linear regressions (Models 3 and 4) with random intercepts for country, year, and vintage. y is the reference year, when the data are initially collected and disseminated. v is the vintage, when previously published data may be revised. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

where β is the DiD estimate of the effect of IMF program participation; α_c and γ_y are country and year fixed effects, respectively; and $\epsilon_{c,y}$ is the idiosyncratic error term.

Figure 4 presents the estimated period-wise average treatment effect on the treated (ATT), with 95 percent confidence intervals, relative to entering or exiting an IMF agree-

Figure 4: Dynamic Effects of IMF Program Participation on Data Revisions



This figure shows the estimated period-wise average treatment effect on the treated (ATT), with 95 percent confidence intervals, relative to (A) entering or (B) exiting the treatment — that is, relative to beginning or ending an IMF program. The bar plots show how many countries are still in the treatment group (i.e., under an IMF program) at each time period relative to when the treatment began. Since IMF programs rarely last over four years, few countries remain treated for longer periods, so the bars shrink over time.

ment. According to Panel (A), entering an IMF program in the first stage increases the *frequency* of revisions. While different, this finding is consistent with the expectation that program participation in the first stage increases the *odds* of revisions in the second stage (Hypothesis 3a). According to Panel (B), estimates gradually decline back toward zero after programs end, suggesting that the IMF's accountability pressures are temporary.

Additional robustness checks corroborate the main findings. While the main models include data on all available vintages, Appendix E uses only each year's main scheduled release, typically in April.[16] Restricting the analysis to the main release reduces noise (since the information provided by two consecutive releases is likely very similar) and ensures that recent years (with many releases) do not disproportionately influence the results compared

[16]For 2020, I use August instead of April, and for 2021 I use September, due to pandemic-related delays in data collection and dissemination.

to earlier years (with fewer releases). Appendix F replaces random effects with country, year, and vintage fixed effects, though these alternative models are less efficient due to the large number of additional parameters. Appendix G reports the results from more complex specifications, estimated via regularized regressions (LASSO, ridge, and elastic net). Finally, Appendix H examines the *direction* of revisions. Compared to downward revisions, upward revisions are significantly more likely at higher *Polyarchy* scores. 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, conditional on state capacity, revisions are more likely for GDP data collected by democracies or IMF borrowers. When revisions occur, they are systematically larger in democracies, whereas state capacity and IMF programs do not appear to affect revision size. Beyond GDP data, I expect these predictions to hold for other macroeconomic indicators, like FDI, trade, inflation, and unemployment, all of which are regularly revised. Beyond WDI data, 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 systematic a test.[17] Still, future research can validate the generalizability of my results by exploring variations in the frequency and magnitude of revisions across other indicators and data sources.

These results have four key implications. First, they add nuance to existing evidence that democracies produce data with higher validity. Democratically elected leaders might be less likely to lie about growth rates or COVID-19 deaths (Martínez, 2022; Adiguzel, Cansunar and Corekcioglu, 2020), yet they face greater incentives to scrutinize and revise

[17] For example, between 1994 and 2024, there have been eleven PWT releases and five releases of the Maddison Project Database, compared to over 100 WDI vintages available in the WDI Database Archives.

information. As a result, democracies produce data with lower reliability: their numbers are more likely to change over time. Even revisions undertaken in good-faith can create a perception of mismanagement — 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 a credibility loss. In contrast, autocrats might collect biased data and publish imprecise estimates, but once released, these estimates are rarely revised. For different reasons, all data are flawed — in autocracies and democracies alike.

Second, and relatedly, transparency in revision processes must be accompanied by public communication strategies that frame revisions as a sign of statistical rigor. Revisions are not inherently problematic; in fact, the international community *encourages* revisions as a sign of statistical rigor, assuming that the initial measurement is fairly accurate and that subsequent revisions provide marginal improvements in data quality. Of course, this is not always the case, but from a statistical perspective, correcting erroneous information is better than letting errors stand. In the US, the backlash against BLS employment data updates in 2024 and 2025 illustrates how, in contexts of low data literacy and widespread misinformation, revisions can erode confidence in official statistics, undermining the government's authority as the main producer of macroeconomic data and driving individuals toward private data providers.[18]

Third, what matters most is the environment under which the initial data were collected. Revisions 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. Long-term data quality hinges on investments in statistical capacity-building at the initial stage of data collection. Put simply, 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 might have

[18]Geoffrey Morgan and Matthew Griffin. "Wall Street Leaning Harder on Private Data After Trump BLS Spat." *Bloomberg*. 23 August 2025.

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. While much of the literature has focused on non-democracies (Goes, 2023) or African states (Johnson et al., 2013), my results suggest that industrial democracies also warrant scrutiny: their data may be more valid but less stable than assumed. In contrast, those working with less democratic states often confront inaccurate information, but at least this information remains fairly consistent across different vintages. 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. This study demonstrates that reliability matters just as much.

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.

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*.

Dang, Hai Anh H., John Pullinger, Umar Serajuddin and Brian Stacy. 2023. "Statistical Performance Indicators and Index—a New Tool to Measure Country Statistical Capacity." *Scientific Data* 10(1):1–14.

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.

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., B. Peter Rosendorff and James Raymond Vreeland. 2014. "Measuring Transparency." *Political Analysis* 22(4):413–434.

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.

Liu, Licheng, Ye Wang and Yiqing Xu. 2024. "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data." *American Journal of Political Science* 68(1):160–176.

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.

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

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.

Nelson, Stephen C. and Geoffrey P.R. R Wallace. 2017. "Are IMF Lending Programs Good or Bad for Democracy?" *Review of International Organizations* 12(4):523–558.

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.

Reinsberg, Bernhard, Thomas Stubbs, Alexander Kentikelenis and Lawrence King. 2019.

"The Political Economy of Labor Market Deregulation During IMF Interventions." *International Interactions* 45(3):532–559.

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.

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

August 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

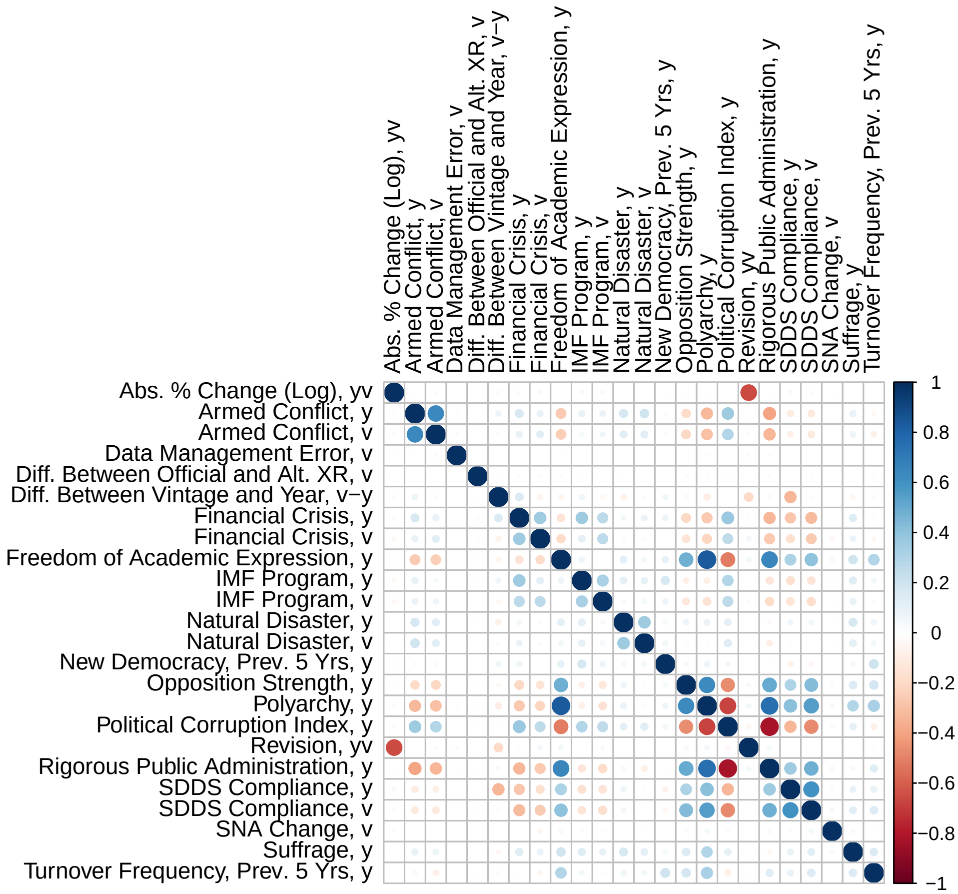
Variable	Description	Coverage	Source
Absolute % Change	Percentage change in the GDP value reported by two consecutive data releases for the same country-year pair	1990–2021	WDI
Armed Conflict	Was there an armed conflict? Yes = 1	1990–2021	Gleditsch et al. (2002)
Data Management Error	Coded 1 for the 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

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

Figure B.1: Correlation Plot



This figure shows the correlation plot for all variables included in the main analysis. 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

C.1 Recording Valid Data

Unlike latent concepts like election integrity, state capacity, or democratic consolidation, which rely on expert coding that might be ideologically biased or improperly aggregated (Giannone, 2010; Martínez i Coma and van Ham, 2015; Hanson and Sigman, 2021), macroeconomic indicators purportedly measure observable, neutral, "objective" outputs. Still, these indicators are often accused of 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). Sometimes different data sources oversimplify the world in different ways. As a result, 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). The US is among the few countries without a centralized NSO; its federal statistical system consists of over 100 statistical agencies, units, and programs that might provide conflicting information. And even an oversimplified world is difficult to measure. This section provides more detailed information about the challenges of measuring valid macroeconomic data.

One challenge, as mentioned in the main text, is low statistical capacity. 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 System of National Accounts (SNA), a global standardization framework developed by the International Comparison Program (ICP) to enable cross-country comparisons. Communist countries did not begin to adopt the SNA until 1993, and North Korea still uses a Marxism-inspired alternative, the Material Product System (Herrera 2010, 23n8; van Heijster and DeRock 2022, 84n1), rendering its data incomparable to other countries.

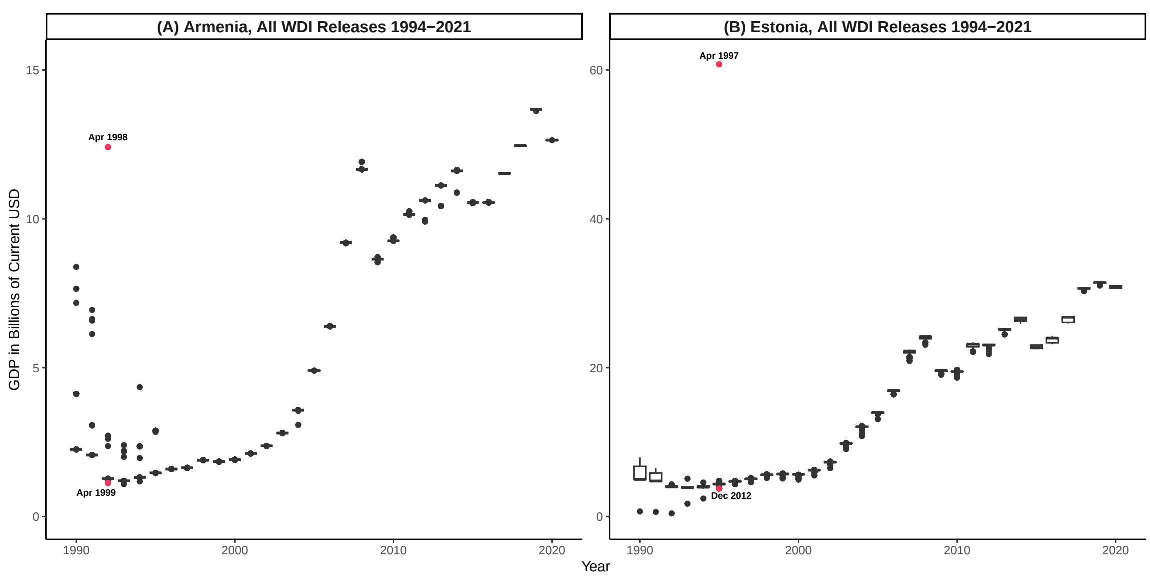
Even countries that follow the SNA might struggle with difficult-to-measure concepts like imputed rent, thus underestimating household final consumption expenditure — an important component of GDP (Olinto Ramos, Pastor and Rivas, 2008). Moreover, it is difficult to quantify the size of the informal economy, which accounts for up to 44 percent of the GDP in the developing world (Coyle, 2014, 110). Population figures tend to be extrapolated from the last census and can grow progressively inaccurate over time; Lebanon, for instance, last conducted a census in 1932 (Devarajan, 2013). Some countries struggle with overly ambitious data-collection efforts: Chile's 2012 census, which spanned three months, failed to account for nearly 10 percent of the population and was replaced by a single-day census in 2017.

Another challenge to measuring valid macroeconomic data relates to political incentives. 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).

A final challenge relates to idiosyncratic data management errors. These cannot be predicted systematically but are fairly easy to identify. Some 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). Below, I provide concrete examples.

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.

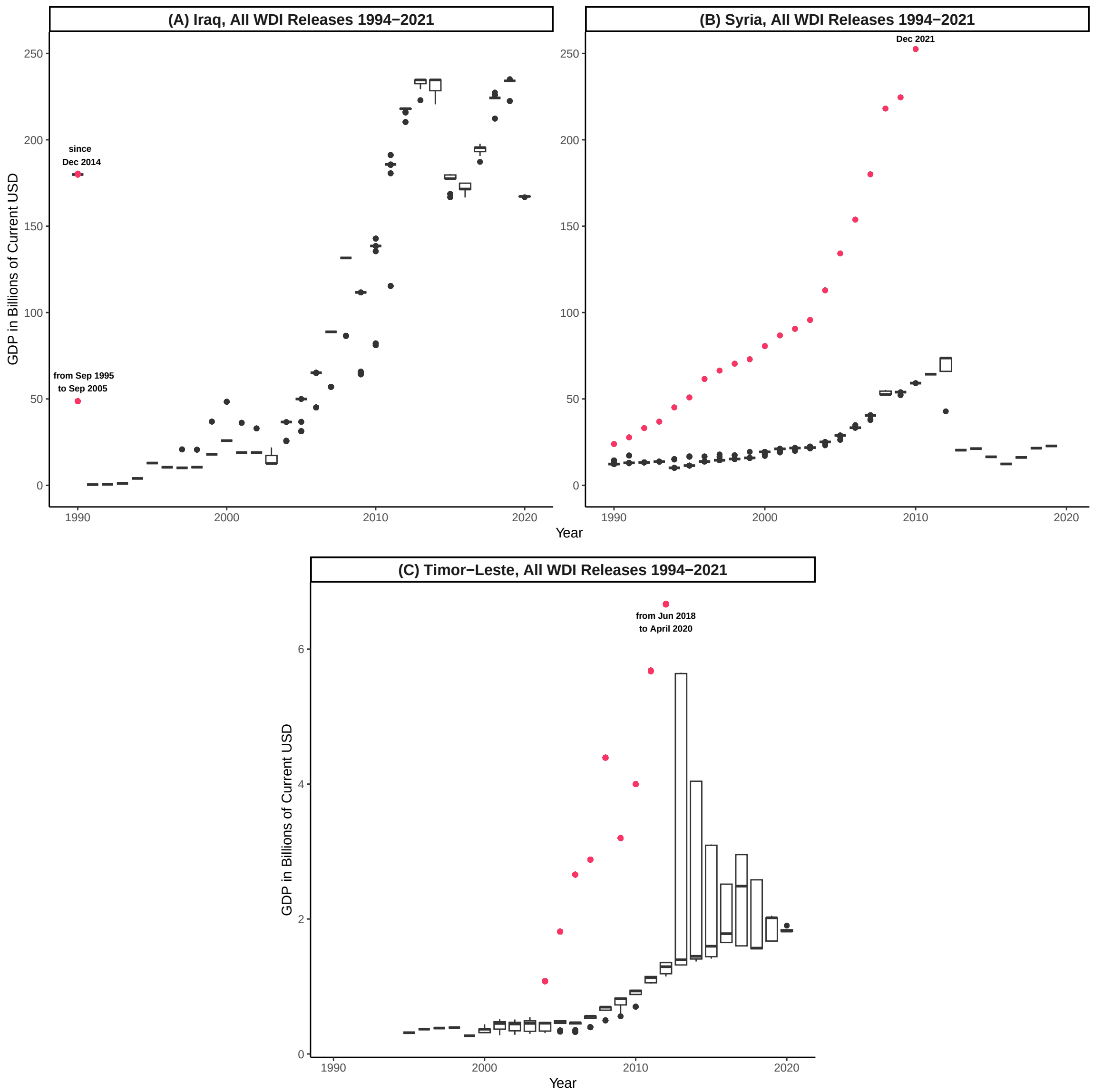
Figure C.1: Current GDP of Armenia and Estonia, 1990–2020



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.

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

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.

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. The World Bank (2023) warns that *some* of its data series for Syria "should not be used for analysis or assessment" because they "are incomplete and based on estimates," but it does not specify *which* data series are affected by these issues.

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.

Iraq, Syria, and Timor-Leste's unusual values appear in multiple vintages. They are not specifically listed by the World Bank in its Data Updates and Errata website. 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, but doing so did not rectify these issues, and the WDI team did not respond to my follow-up inquiries about these specific observations. Thus, 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.

C.2 Correcting the Record

This section provides additional information about prominent data revisions. I begin by providing examples of what Croushore and Stark (2003) deems information-based revisions, which are related to the "discovery" of new data. In these cases, countries revise their data to incorporate newly available information that allows for more precise measurement.

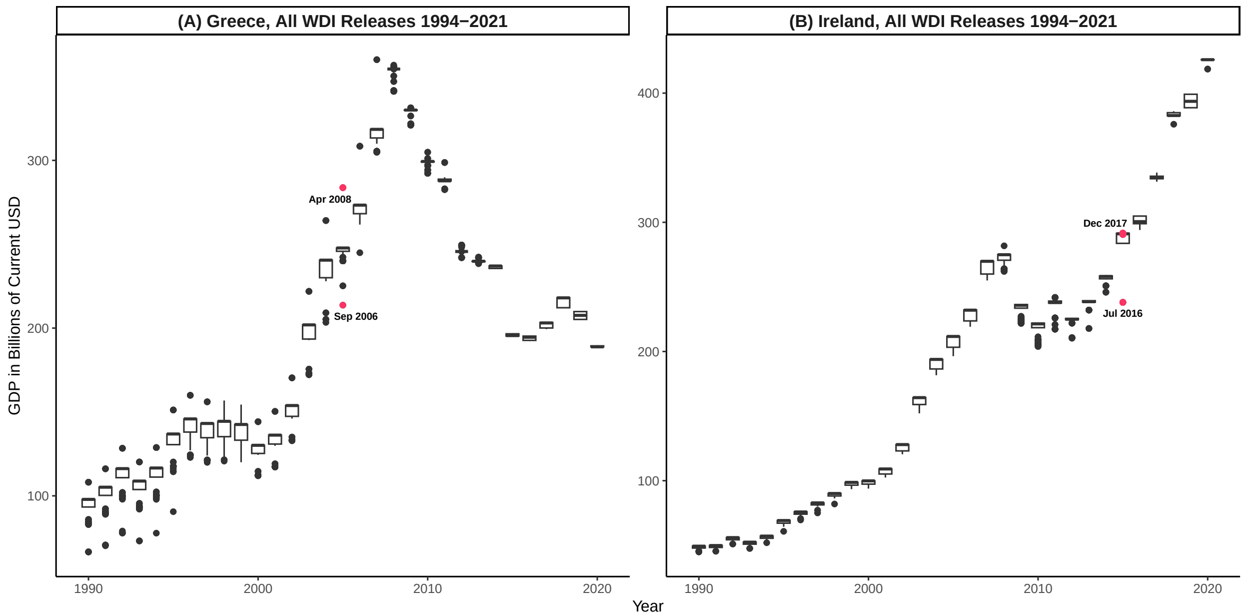
In 2009, Prime Minister George Papandreou came to power in Greece and requested help from Eurostat and the International Monetary Fund (IMF) to revise the country's finances.[1] As these revisions quickly established, previous administrations had overestimated tax revenues, disregarded social security 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 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). Figure C.3 illustrates the cases of Greece and Ireland. Between the September 2006 and April 2008 WDI

[1] In fact, information-based revisions are often tied to international actors; for example, subscribing to the IMF's Special Data Dissemination Standard (SDDS) can increase transparency (Vadlamannati, Cooray and Brazys, 2018).

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.

Figure C.3: 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.

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).

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. However, price surveys in China were only conducted in urban areas, introducing yet another potential measurement issue (Bolt and van Zanden, 2024). Ultimately, ICP rounds still disagree with each other due to differences in relative prices, consumption patterns, region-specific PPP adjustments, and accounting or reporting practices (Deaton and Aten, 2017).

Besides information-based revisions, countries may also conduct structural revisions, changing definitions, base years, or aggregation methods. 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). Nicaragua revised its national accounts in 2003, changing the base year from 1980 to 1994 and implementing the 1993 SNA; as a result, the country's current GDP for the year 2000 increased by 70 percent (Olinto Ramos, Pastor and Rivas, 2008, 9).

D Difference-in-Differences: Full Results and Diagnostics

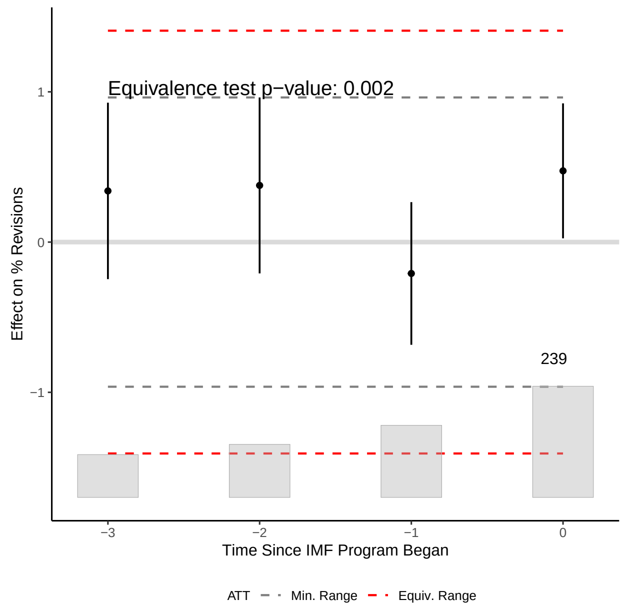
I estimate the DiD model using the `R` package `fect` (Liu, Wang and Xu, 2024). Table D.1 presents the result, reporting estimates for the average treatment effect on the treated (ATT) where treated observations are weighted equally, and alternatively where treated units (countries) are weighted equally. Figure D.1 tests for the existence of a pre-trend; the `fect` package offers two pre-trend tests, the F-test and the equivalence test. In large samples with few outliers, the F-test tends to reject the null hypothesis of no pre-trend even for substantively negligible pre-trend differences (Liu, Wang and Xu, 2024, 10-11). Therefore, I rely on the equivalence test, which directly assesses whether pre-treatment differences are significantly meaningful. A smaller p-value suggests a better pre-trend fitting. The model passes the equivalence test at $p < 0.01$, indicating good pre-treatment balance between treated and untreated groups.

Table D.1: Average Treatment Effect on the Treated

	Dependent Variable:
	% Revisions
	(1)
IMF Program (Weighted by Obs.)	0.545^{*}
	(0.322)
IMF Program (Weighted by Country)	0.801^{**}
	(0.314)
Number of Countries	173
Number of Years	31

Model includes bootstrapped standard errors. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Figure D.1: Pre-Trend Test



This figure tests for the existence of a pre-trend. The model passes the pre-trend test at $p < 0.01$, rejecting the null that the treated and untreated groups display a pre-specified difference.

E Reduced Sample: Only Main Releases

While the main models include data on all available vintages, Table E.1 uses only each year's main scheduled release, typically in April.[2] Restricting the analysis to the main release reduces noise (since the information provided by two consecutive releases is likely very similar) and ensures that recent years (with many releases) do not disproportionately influence the results compared to earlier years (with fewer releases). The results are robust to these changes, though *Data Management Error* drops out of Model 4: focusing only on revisions in the main vintages results in too few observations that qualify as coding errors.

[2]For 2020, I use August instead of April, and for 2021 I use September, due to pandemic-related delays in data collection and dissemination.

Table E.1: Determinants of the Likelihood and Magnitude of Data Revisions: Only Main Scheduled Releases

	Dependent Variable:			
	Revision = 1		Abs. % Change (Log)	
	(1)	(2)	(3)	(4)
Rigorous Public Administration, y	0.08^{***}	0.07^{***}	0.06	−0.11
	(0.02)	(0.03)	(0.10)	(0.10)
Polyarchy, y	1.06^{***}		1.86^{***}	
	(0.11)		(0.42)	
Freedom of Academic Expression, y		0.21^{***}		0.30^{***}
		(0.02)		(0.08)
Suffrage, y		-0.27^{**}		0.70
		(0.13)		(0.49)
New Democracy, Prev. 5 Yrs, y		-0.08^{**}		−0.17
		(0.04)		(0.15)
Opposition Strength, y		−0.00		-0.01^{**}
		(0.00)		(0.00)
Turnover Frequency, Prev. 5 Yrs, y		0.02		0.06
		(0.02)		(0.07)
Political Corruption Index, y		−0.15		-2.25^{***}
		(0.12)		(0.45)
IMF Program, y	0.08^{***}	0.06^{**}	−0.13	−0.03
	(0.02)	(0.02)	(0.10)	(0.10)
IMF Program, v	−0.00	0.01	-0.33^{***}	-0.29^{***}
	(0.03)	(0.03)	(0.11)	(0.11)
SDDS Compliance, y	-0.32^{***}	-0.33^{***}	-0.38^{***}	-0.51^{***}
	(0.03)	(0.04)	(0.14)	(0.14)
SDDS Compliance, v	0.05	0.10^{**}	0.89^{***}	0.88^{***}
	(0.04)	(0.04)	(0.17)	(0.17)
Financial Crisis, y	0.05^{*}	0.06^{**}	-0.19^{*}	-0.19^{*}
	(0.02)	(0.03)	(0.10)	(0.10)
Financial Crisis, v	0.05^{*}	0.04	−0.16	−0.17
	(0.03)	(0.03)	(0.11)	(0.11)
Natural Disaster, y	0.05^{**}	0.05^{**}	−0.11	−0.05
	(0.02)	(0.02)	(0.09)	(0.09)
Natural Disaster, v	0.00	0.00	0.23^{**}	0.24^{**}
	(0.02)	(0.02)	(0.10)	(0.10)
Armed Conflict, y	−0.01	0.02	-0.68^{***}	-0.64^{***}
	(0.04)	(0.04)	(0.17)	(0.17)
Armed Conflict, v	−0.06	−0.06	0.56^{***}	0.61^{***}
	(0.04)	(0.04)	(0.16)	(0.16)
Diff. Between Official and Alt. XR, v	−0.00	−0.00	-0.06^{*}	-0.07^{*}
	(0.01)	(0.01)	(0.04)	(0.04)
SNA Change, v	0.47^{***}	0.47^{***}	2.62^{***}	2.63^{***}
	(0.04)	(0.04)	(0.15)	(0.15)
Data Management Error, v	−0.18	−0.47	15.11^{**}	
	(0.90)	(0.94)	(6.86)	
Diff. Between Vintage and Year, $v-y$	-0.08^{***}	-0.08^{***}	-0.34^{***}	-0.35^{***}
	(0.01)	(0.01)	(0.01)	(0.01)
Intercept	0.50^{*}	0.83^{***}	-4.09^{***}	-3.06^{***}
	(0.26)	(0.30)	(0.73)	(0.90)
Observations	78,178	76,092	40,119	39,263
Log Likelihood	−41,793.10	−40,583.76	−134,357.76	−131,434.80
Number of Countries	170	167	170	166
Number of Years	31	31	30	30
Number of Vintages	26	26	26	26
Variance: Countries (Intercept)	0.38	0.37	3.16	3.04
Variance: Years (Intercept)	0.26	0.27	0.19	0.18
Variance: Vintages (Intercept)	1.31	1.35	11.46	11.18

This table presents the results of two logistic regressions (Models 1 and 2) and two linear regressions (Models 3 and 4) with random intercepts for country, year, and vintage. y is the reference year, when the data are initially collected and disseminated. v is the vintage, when previously published data may be revised. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

F Models With Fixed Effects

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. As Table F.1 shows, not all results remain robust under these specifications. However, these models are not comparable

Table F.1: Determinants of the Likelihood and Magnitude of Data Revisions: Fixed Effects

	Dependent Variable:			
	Revision = 1		Abs. % Change (Log)	
	(1)	(2)	(3)	(4)
Rigorous Public Administration, y	0.06	0.06	-0.02	-0.07
	(0.05)	(0.05)	(0.13)	(0.14)
Polyarchy, y	0.69^{***}		0.71	
	(0.26)		(0.72)	
Freedom of Academic Expression, y		0.17^{***}		0.22^{*}
		(0.06)		(0.12)
Suffrage, y		-0.68		0.08
		(0.52)		(0.99)
New Democracy, Prev. 5 Yrs, y		-0.03		-0.11
		(0.04)		(0.19)
Opposition Strength, y		-0.00		-0.01^{**}
		(0.00)		(0.00)
Turnover Frequency, Prev. 5 Yrs, y		0.01		0.09
		(0.02)		(0.10)
Political Corruption Index, y		0.27		0.08
		(0.29)		(0.89)
IMF Program, y	0.07^{*}	0.05	-0.04	0.00
	(0.03)	(0.03)	(0.13)	(0.13)
IMF Program, v	-0.07	-0.07	-0.32	-0.33
	(0.06)	(0.06)	(0.29)	(0.29)
SDDS Compliance, y	-0.27^{***}	-0.27^{***}	-0.11	-0.16
	(0.07)	(0.07)	(0.29)	(0.30)
SDDS Compliance, v	0.14	0.18^{*}	0.25	0.30
	(0.10)	(0.10)	(0.56)	(0.56)
Financial Crisis, y	0.06	0.07	-0.14	-0.15
	(0.05)	(0.05)	(0.14)	(0.14)
Financial Crisis, v	0.08	0.07	0.60^{**}	0.62^{**}
	(0.06)	(0.06)	(0.29)	(0.30)
Natural Disaster, y	0.05^{***}	0.06^{***}	-0.14^{*}	-0.11
	(0.02)	(0.02)	(0.07)	(0.07)
Natural Disaster, v	0.02	0.03	-0.12	-0.10
	(0.05)	(0.05)	(0.24)	(0.24)
Armed Conflict, y	-0.04	-0.01	-0.37	-0.32
	(0.07)	(0.07)	(0.26)	(0.27)
Armed Conflict, v	-0.09	-0.09	0.31	0.38
	(0.12)	(0.12)	(0.70)	(0.70)
Diff. Between Official and Alt. XR, v	-0.01^{***}	-0.01^{***}	0.01	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
SNA Change, v	0.20^{**}	0.21^{**}	1.68^{***}	1.73^{***}
	(0.09)	(0.09)	(0.40)	(0.40)
Data Management Error, v	6.98^{***}	7.02^{***}	12.02^{***}	11.95^{***}
	(1.23)	(1.23)	(1.03)	(1.09)
Observations	295,399	287,529	62,916	61,551
Log Likelihood	$-86,235.26$	$-83,926.17$		
Pseudo R^2	0.46	0.46		
R^2			0.43	0.43
Number of Countries	170	167	170	167
Number of Years	31	31	31	31
Number of Vintages	80	80	78	78

This table presents the results of two logistic regressions (Models 1 and 2) and two linear regressions (Models 3 and 4) with fixed effects for country, year, and vintage, and standard errors clustered by country. y is the reference year, when the data are initially collected and disseminated. v is the vintage, when previously published data may be revised. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

to the random effects models because of the reduced sample size. This issue is especially pronounced for Models 1 and 2 (logistic regressions), which suffer from perfect separation: if some countries have all zeros or ones in some vintage, for example, the logistic regression automatically drops their fixed effects.

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 the first and the second stage (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

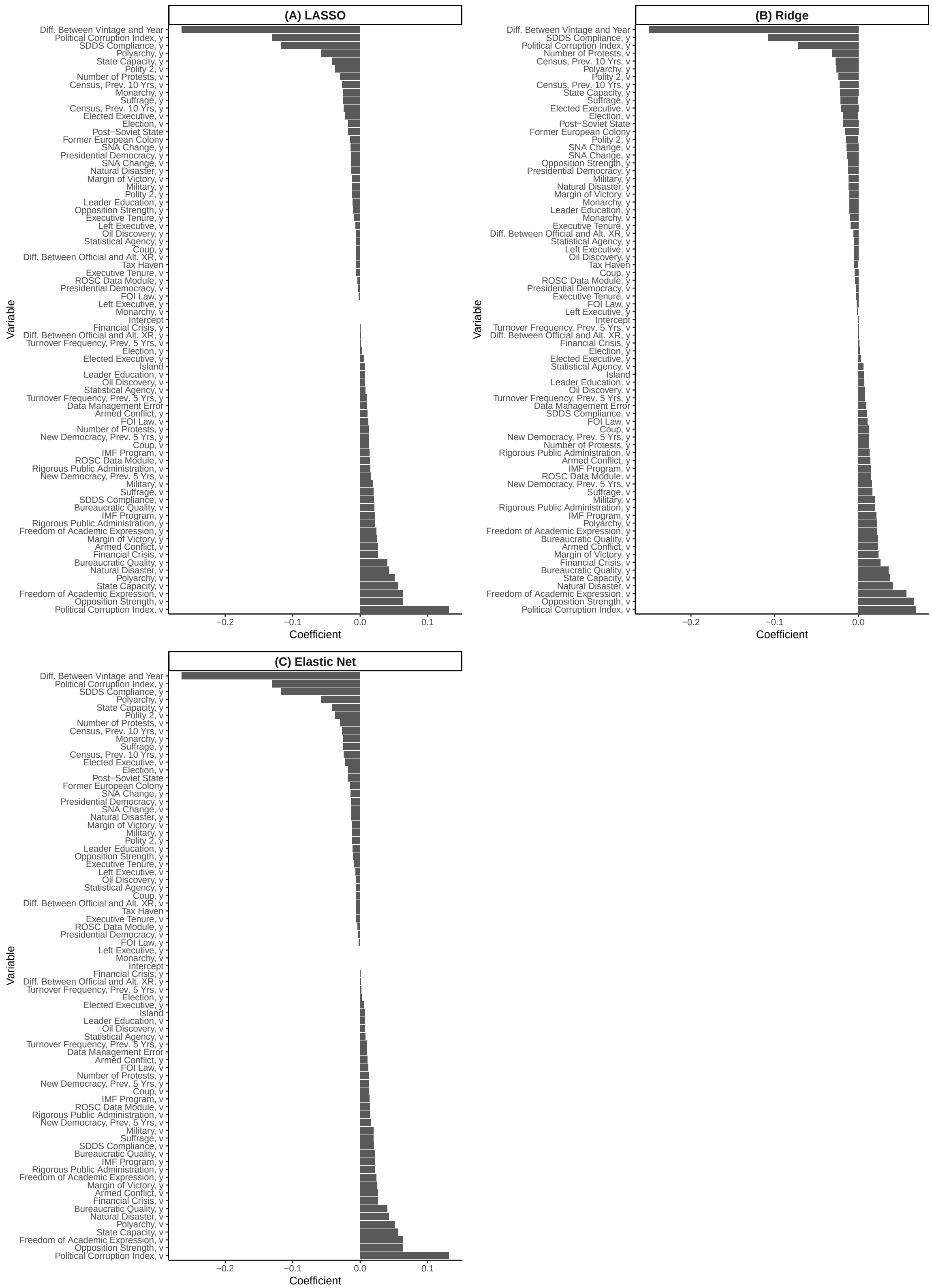
Variable	Description	Coverage	Source
Bureaucratic Quality	To what extent does the country's bureaucracy have the strength and expertise to govern without drastic changes in policy or interruptions in government services? Low = 0, High = 4	1990–2021	The PRS Group (2022)
Census	Was there a national census 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* — all at the second stage, v — 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* at v and *State Capacity* at y 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.

H Direction of Revisions

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

	Dependent Variable: Upward Revision = 1
	(1)
Rigorous Public Administration	−0.01
	(0.02)
Polyarchy	0.14**
	(0.07)
IMF Program (Step 1)	0.01
	(0.02)
IMF Program (Step 2)	0.00
	(0.02)
SDDS Compliance (Step 1)	0.08***
	(0.03)
SDDS Compliance (Step 2)	−0.01
	(0.03)
Financial Crisis (Step 1)	−0.04**
	(0.02)
Financial Crisis (Step 2)	−0.04*
	(0.02)
Natural Disaster (Step 1)	0.01
	(0.02)
Natural Disaster (Step 2)	−0.06***
	(0.02)
Armed Conflict (Step 1)	0.03
	(0.03)
Armed Conflict (Step 2)	−0.08**
	(0.04)
Diff. Between Official and Alt. XR	−0.01
	(0.01)
SNA Change	0.15***
	(0.04)
Data Management Error	0.80
	(0.49)
Diff. Between Vintage and Year	−0.00
	(0.00)
Intercept	0.26***
	(0.08)
Observations	62,916
Log Likelihood	−42,035.89
Number of Countries	170
Number of Years	31
Number of Vintages	78
Variance: Countries (Intercept)	0.02
Variance: Years (Intercept)	0.00
Variance: Vintages (Intercept)	0.23

This table presents the results of a logistic regression with random intercepts for country, year, and vintage. Step 1 is the reference year, the year of data collection. y is the reference year, when the data are initially collected and disseminated. v is the vintage, when previously published data may be revised. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

When revisions occur, what explains their direction? Table H.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.

References

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.

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/*

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.

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.*

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*.

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.

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.

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

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*

Giannone, Diego. 2010. "Political and Ideological Aspects in the Measurement of Democracy: The Freedom House Case." *Democratization* 17(1):68–97.

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.

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.

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

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*

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.

Liu, Licheng, Ye Wang and Yiqing Xu. 2024. "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data." *American Journal of Political Science* 68(1):160–176.

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

Martínez i Coma, Ferran and Carolien van Ham. 2015. "Can Experts Judge Elections? Testing the Validity of Expert Judgments for Measuring Election Integrity." *European Journal of Political Research* 54(2):305–325.

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.

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.

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

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.

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.

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