New Estimates of Over 500 Years of Historic GDP and Population Data

Christopher J. Fariss
Department of Political Science
University of Michigan
cjf0006@gmail.com *

Therese Anders
School of International Relations
University of Southern California
tanders.usc@gmail.com

Jonathan N. Markowitz
School of International Relations
University of Southern California
jnmarkowitz@gmail.com

Miriam Barnum
School of International Relations
University of Southern California
mbarnum@usc.edu

Journal of Conflict of Resolution (Forthcoming)

Abstract

Gross Domestic Product (GDP), GDP per capita, and population are central to the study of politics and economics broadly, and conflict processes in particular. Despite the prominence of these variables in empirical research, existing data lack historical coverage and are assumed to be measured without error. We develop a latent variable modeling framework that expands data coverage (1500 A.D–2018 A.D) and, by making use of multiple indicators for each variable, provides a principled framework to estimate uncertainty for values for all country-year variables relative to one another. Expanded temporal coverage of estimates provides new insights about the relationship between development and democracy, conflict, repression, and health. We also demonstrate how to incorporate uncertainty in observational models. Results show that the relationship between repression and development is weaker than models that do not incorporate uncertainty suggest. Future extensions of the latent variable model can address other forms of systematic measurement error with new data, new measurement theory, or both.

Keywords: *Gross Domestic Product, population, GDP per capita, latent variables, measurement, construct validity*

*Contact Author. We thank Pablo Barberá, Megan Becker, Michael Beckley, John Gerring, Kristian Gleditsch, Ben Graham, David Lake, and James Lo for helpful comments on this project. All data and code necessary to replicate the model and predicted data presented in the article will be publicly available upon publication at https://dataverse.harvard.edu/dataverse/LatentGDP. This research was supported by the Security and Political Economy (SPEC) Lab at the University of Southern California.

Introduction

Gross Domestic Product (GDP), GDP per capita (GDPPC), and population play a vital role in empirical social science. Moreover, they are key variables in the study of international and domestic conflict processes in particular. Despite the prominence of these variables, existing data used to operationalize them suffers from three problems: (1) measurement coverage (missingness), (2) measurement uncertainty, and (3) measurement bias. First, there is a lack of economic cross-country data coverage prior to 1950 (Gleditsch, 2002), or data are available with very limited temporal coverage even though many other datasets of interest cover variables beginning in the 1800s.[1] This leaves researchers unable to consistently estimate the relationship between key variables such as economic development and democratization, conflict, repression, or health prior to this date. Second, measurement error arises because of imprecision or disagreement in available data, which can mask real relationships or exacerbate false ones. Though scholars are aware that data suffers from measurement error, existing estimates of these variables provide only point estimates and offer no method for quantifying uncertainty (measurement error). Third, existing models for estimating GDP, population, and GDP per-capita offer no way to correct for any systematic bias in the data generating process. As previous research has demonstrated, failing to correct for measurement bias leads scholars to draw incorrect inferences about the relationships in their data. We develop a latent variable model of GDP, GDP per-capita, and population that provides remedies for each of these issues.[2]

First, the latent variable model generates estimates of cross-national data coverage back in time by several centuries to 1500. These estimates are important not only for scholars assessing existing explanations of outcomes in earlier periods of history, but also scholars interested in comparing inferences generated from data across time periods. Given that most of the major interstate wars occurred before 1950, these data will be particularly useful to scholars studying the causes of war, as well as the debate about its relative decline over time (e.g., Fazal, 2014; Lacina, Gleditsch and Russett, 2006). Until now, researchers have been unable to estimate these long-term relationships.

[1] For example, Bolt et al. (2018) and Maddison (2010) provide extended historical coverage only for specific years, e.g. 1800, 1820, and 1850.

[2] This measurement model builds on earlier data collection efforts and model versions in which the combined sources were used to estimate these three variables (e.g., Anders, Fariss and Markowitz, 2020; Fariss et al., 2017; Markowitz and Fariss, 2018). These earlier models did not account for the different modes of data generation for each of the datasets (i.e., PPP adjusted vs. exchange rate conversion), which our model accounts for.

For example, existing research argues that economic development reduces the risk of conflict both by contributing to the democratization process of states in the international system and by making war more costly (Hegre, 2000). As Souva and Prins state, "democratic regimes are about 37% less likely to initiate fatal militarized disputes given an average level of GDP per capita" (Souva and Prins, 2006, 194). However, other research argues that the relationship with economic development is curvilinear, for both democratization (Treisman, 2020) and conflict (Boehmer and Sobek, 2005; Gartzke, 2012). To date, the relationships between economic development and democratization, repression, or conflict have not been estimated with complete cross-national data prior to 1950. Scholars have instead relied on proxy-measures such as energy consumption per-capita (e.g., Markowitz, McMahon and Fariss, 2019), shipping and rail costs (e.g., Lake, 2009; Markowitz and Fariss, 2013), or simple linear interpolation of GDP per-capita (e.g., Treisman, 2020) to account for long-term economic variation. The new estimates we present in this article allow researchers to evaluate whether these empirical relationships are limited to the post-1950 period, or whether they generalize to earlier time periods (using more precise data with better coverage). We show that many correlational patterns between GDP per capita and these other variables vary considerably over time, which means that relationships for the post-1950 period do not necessarily generalize to periods prior to 1950. Critically, this also suggests that these relationships might change in the future.

Second, our new latent variable model provides estimates of the relative level of uncertainty for each country-year unit by accounting for variation in the level of coverage within and disagreement between component indicators. This is useful because it allows researchers to evaluate whether measurement error in the data (expressed as the level of uncertainty with which each country-year unit is accurately measured) is large enough to alter the size, or even the direction, of the effect of a given explanatory variable.

For illustration, we demonstrate that measurement uncertainty may be large enough to reduce the estimated magnitude of the effect of development (GDP per capita) on repression by as much as a third. If models do not incorporate uncertainty, then the researcher cannot rule out the possibility that the statistical associations of GDP, GDP per capita, or population with some other variable are not a false-negative result (Type 2 error or attenuation bias), or the related possibility that the relationship of these variables are a false-positive result (Type 1 error). In

models with multiple indicators (i.e., multiple-variable regression), bias due to measurement error is not always attenuating if un-modeled, higher-order interactions exist. Incorporating measurement uncertainty of variables addresses these difficult-to-model issues and can be further explored with non-parametric regression techniques. In sum, incorporating uncertainty into a regression model provides evidence that effect sizes are probabilistically distinguishable from zero even when we cannot measure a right-hand side variable perfectly (an assumption of standard regression models). Though some existing models that include these variables may have under-estimated an effect size, others may have over-estimated effect size or even incorrectly reversed the direction of the effect. The implications are potentially profound given the wide usage of these variables across the social sciences.

Third, the measurement model provides a framework for incorporating theoretical knowledge about the data generating process for information used in the measurement of each variable that can correct for potential bias in the existing data. We describe how the latent variable model we develop in this article can be further developed to address measurement bias and point to other areas of scholarship that have successfully built and extended latent variable models to address measurement bias.[3]

The new estimates provide ample opportunities for future scholarship to re-visit existing debates and investigate new research questions relating to war, peace, economic development and health and well-being, as well as the information necessary to reduce bias due to measurement uncertainty in GDP, GDP per capita, or population variables. The key output from the model is predicted intervals of the original source variables in the original unit-of-measurement, in addition to the relative level of uncertainty for each country-year estimate in the form of a standard deviation. The level of precision for a particular unit is based on the number of indicators available for that unit, its temporal proximity to other units with available information because the model is dynamic, and the agreement between the indicators. Missing values are inferred from the dynamic structure of the model based on the information contained in the variables associated with a given country year and the estimates from prior and future values. This directly addresses the issue of list-wise deletion that is well known in the conflict literature to potentially bias results (e.g., Boehmer,

[3]See examples from the measurement of human rights (Fariss, 2014, 2018*b*,*a*, 2019; Fariss and Dancy, 2017; Fariss, Kenwick and Reuning, 2020) and civilian control of the military (Kenwick, 2020).

Jungblut and Stoll, 2011; Gleditsch, 2002; King et al., 2001).

Fourth, scholars can use their preferred variable (e.g. the data series from the Penn World Tables, World Bank, or the Maddison Project) because our model generates a range of estimates for every country-year unit in terms of each and every one of the variables that we include in our model. The new estimates are the posterior predications generated from a dynamic latent variable model that approximates the process economic historians use when creating measures of these variables. Economic historians begin with population estimates, they then use historic data on prices and sector specific production to generate per-capita estimates of economic value. With the estimates of per-capita economic value and population, the calculation of overall GDP becomes possible. Because economic historians are using some data series that are designed to estimate change in productivity within a country over time and others are designed to capture cross-sectional variation, but usually not both, we incorporate this information into our measurement models of GDP and GDP per capita.[4]

Having demonstrated that solving these problems has important implications for all areas of social science research, we now turn to the task of demonstrating that we have used rigorous and appropriate methods to construct a new dataset to measure each of these concepts of interest. We close with several demonstrations of how our new estimates can be applied by scholars to improve their research.

Sources of Measurement Error

Existing measures of GDP, GDP per capita, and population attempt to measure three related concepts. Gross Domestic Product (and the related Gross National Product, GNP) are intended to measure the total amount of economic output created in a country in a given year (World Bank, N.d.). Population measures reflect the total number of people within the geographic unit in a given year (World Bank, N.d.). GDP per capita (and the related GNP per capita) is a measure of the total amount of economic output created in a country per person in a given year (World Bank, N.d.). The empirical constructs that are intended to map on to these theoretical concepts

[4]We do this by introducing appropriate random effects that make it so only that variation from the relevant dimension of comparison (i.e, across countries or over time) are used to inform the underlying latent trait that links all of the other parameters together. We discuss this in more detail below.

suffer from a fundamental measurement problem — they are incompletely observable and must be inferred from economic source material and census information.

To enumerate all individuals within most administrative units (e.g., municipalities, counties, provinces, states) and track every unit of wealth created in a country (GDP) or by the means of production owned by a country's residents (GNP) each year is labor intensive and cost prohibitive. Thus, these values can only be *estimated* each year and some degree of error will always be present in the resulting values available in a country's statistical material. While some measurement error is inevitable, recent research suggests that current measures contain considerable error and that this error is possibly systematically related to the estimation process itself (e.g., Jerven, 2010*c*,*b*,*a*, 2013*b*, 2014; Lee and Zhang, 2017; Wallace, 2016; UNU-IHDP and UNEP, 2014).

Sources of error in demographic data

One source of error is that states have incentives to release biased statistics (Jerven, 2013*b*; Wallace, 2016). States might have reason to over-report official statistics (Jerven, 2013*b*), because high population or GDP figures might enhance support domestically, or impress foreign governments or investors from other states. The Chinese government, for example, appears to exaggerate its official estimates to forestall opposition criticism and thereby improve regime support (Wallace, 2016). States might also under-report official statistics. This is because low population and GDP numbers might attract foreign aid (Neumayer, 2003; Burnside and Dollar, 2000) or other forms of international assistance (Büthe, Major and de Mello e Souza, 2012). These strategic incentives may lead to systematic reporting errors in the source material used to construct the data (e.g., Fariss, 2014; Scott, 1998). States may also simply not be able to collect accurate data. Even countries with a capable bureaucracy may face political issues when attempting to accurately count all individuals (Lee and Zhang, 2017; Prewitt, 2010). The situation is worse in some parts of the developing world, where population estimates are often outdated simply because many countries are unable to conduct a census on a regular basis (e.g., Discoll and Lidow, 2014; Lee and Zhang, 2017; Scott, 1998). The absence of a regular census might be because states are not interested in gathering these data (intentional misgovernance) or because they do not have the capacity to do so (unintentional misgovernance) (e.g., Banerjee, 1997; Slough and Fariss, 2020). Regardless of

the cause, the result is the same — outdated data for many country-year observations. Our model provides an open and transparent framework that begins to address these problems as new data and understanding arise about different countries and time periods, which is a point of discussion we return to in the conclusion.

Sources of error for GDP and GDP per capita

GDP and GDP per capita estimates are also subject to potential biases. For example, statisticians collect additional data on the structure of a country's economy in some years, which are called base years (Jerven, 2013*a*; Data Quality Institute, World Economics, 2016). The values from the base years are used to weight the relative contribution of each sector of the economy to the national income (Jerven, 2013*b*). The contribution of sectors can change considerably over time, as new industries emerge and old ones disappear (Jerven, 2013*a*, 141-144). To address the changes in the economy that affect GDP estimates, the United Nations and International Monetary Fund both recommend that states update their base year data collection every five years to capture changes in the influence of sectors within countries (Jerven, 2013*a*; Data Quality Institute, World Economics, 2016). However, many countries update this information less frequently than suggested. For example, a recent survey suggests that only 7 out of 48 countries in sub-Saharan Africa follow this recommendation (Jerven, 2013*a*, 146), which can lead to large changes in GDP estimates. For example Nigeria updated its base year in 2013, which led to an 89% increase of its GDP estimate for that year (Economist, 2014). As this case illustrates, using old base years can cause considerable error in reported GDP estimates.

In addition to these shared sources of error, there are many other factors that influence the accuracy of GDP estimates. One such factor relates to the United Nations System of National Accounts (SNA) (European Communities et al., 2009). The System of National Accounts (SNA) is a set of international recommendations about measuring economic activity, designed to make cross-national comparisons easier (Deaton and Heston, 2010; Data Quality Institute, World Economics, 2016). The recommendations are updated semi-regularly (e.g., United Nations 1953; European Communities et al. 2009). However, not all states use these standards, and many that do might not use the most up to date version (Data Quality Institute, World Economics, 2016). This leads to

considerable heterogeneity across countries in the degree to which reported GDP statistics capture underlying economic activity as reflected by a standardized reporting procedure.

A second factor that can lead to misreporting is the size of a country's informal, or shadow, economy. This sector of the economy comprises a mix of legal activities, such as self-employment and bartering, and illegal activities, such as drug-dealing and gambling (Schneider and Enste, 2000, 5). In developing countries, the informal market can represent a substantial share of the economy — perhaps more than 40 percent (International Labour Office, 2013). Even in OECD countries, it can represent about 15 percent of all activity (International Labour Office, 2013). While states usually try to incorporate informal work in their national statistics, this can be difficult since it requires accurately calculating the value of this work (Schneider and Enste, 2000).

A third factor that can lead to error in GDP estimates is whether and how states account for public sector outputs (Atkinson, 2005). There are two issues here. One is that the national accounts system in most states only tracks government inputs, such as salaries and fixed costs, but not government outputs, such as education and welfare services (Stiglitz, Sen and Fitoussi, 2009, 31). By not accounting for government outputs, states underestimate national economic activity (Stiglitz, Sen and Fitoussi, 2009). Another issue is that governments often subsidize the cost of public goods, leading to sub-market prices for these goods (Stiglitz, Sen and Fitoussi, 2009, 17). As a result, states that measure government outputs based on their market value also underestimate national economic activity.

Existing approaches for addressing sources of error

Despite these various sources of error, and the potential dangers they pose for empirical work, there have been few attempts to systematically account for them across all available data sources.[5] Most efforts to improve population and GDP measures have focused on extending their spatial or temporal coverage (which we are interested in as well). These projects typically 'fill in' gaps from country-year series that were the result of conflict or political instability (Gleditsch, 2002). Perhaps the most prominent example of this is the historic GDP and population dataset constructed by Angus Maddison and later the Maddison Project (Bolt and van Zanden, 2020; Maddison, 2010).

[5]There has been substantial work on understanding the error in government statistics for a subset of countries and years, particularly in countries in Africa (e.g., Jerven, 2010*a*,*b*,*c*, 2013*b*, 2014).

Several organizations, such as World Economics (Data Quality Institute, World Economics, 2016), have tried to address the concern we raise about these statistics by releasing datasets that measure their quality. There are two primary problems with these datasets, however. The first is that their temporal range is usually extremely limited. The second is that while these measures provide some information about the relative uncertainty of official GDP statistics, they do not incorporate this information into country-year estimates.

While these efforts have been extremely useful, as reflected in the thousands of citations they receive, they do not directly address our primary concern, which is the existence (1) measurement coverage (missingness), (2) measurement uncertainty, and (3) measurement bias in reported country-year estimates. We contribute to this area of research by both extending the coverage of the estimates temporally and spatially for missing cases, by providing researchers with a means of quantifying and analyzing the relative precision of the resulting estimates, and a principled measurement framework for addressing bias. Importantly, our model should be considered a useful starting point that can be extended as new information becomes available.

GDP, Population, and GDP per capita Component Datasets

Our latent variable model is estimated based on data for GDP,[6] GDP per capita,[7] and population.[8] Details on the sources, measurement choices, and coverage of the component variables are provided in Table 1. For each component dataset, we extract relevant indicators, attach unique country identifiers, and reshape the data into a common country-year format. Details on the underlying source materials for each component measure and coding decisions are provided in supplementary appendix A and are documented in the `R` code we use to merge the constituent datasets together.

When does a country enter our dataset? A country enters our dataset beginning in the earliest year covered by one of the datasets we draw on for our project.[9] For example, we generate estimates

[6]For observed data on GDP see Measuring Worth (2019); World Bank (N.d.); Feenstra, Inklaar and Timmer (2015); Broadberry and Klein (2012); Bairoch (1976).

[7]For observed data on GDP per capita see Measuring Worth (2019); Bolt and van Zanden (2020); Bolt et al. (2018); World Bank (N.d.); Broadberry (2015); Broadberry and Klein (2012); Bairoch (1976).

[8]For observed data on population see Bolt and van Zanden (2020); Measuring Worth (2019); World Bank (N.d.); Feenstra, Inklaar and Timmer (2015); Broadberry and Klein (2012); Singer, Bremer and Stuckey (1972); Deng (2004).

[9]We use the Gleditsch and Ward (1999) revised list of independent states as the base set of units. Conversion issues between country identifiers in alternative data sources are solved using a script from Graham and Tucker

for England from 1500 to 2019 A.D. and for Ghana from 1820 to 2019 A.D. because the Maddison Project dataset covers England beginning in 1500 and Ghana beginning in 1820. Note that if a country does not have any coverage across datasets (i.e, it has no observed values for any variable) it does not enter our dataset.[10]

As we discuss in each dataset description, different datasets sometimes use different spatial definitions for units and have different coverage over time (see online appendix A and appendix B). We have matched country-year units across datasets using the best match available. In some cases, units exist in the dataset that are not historically accurate such as a unified Germany prior to 1871. Maddison includes this unit in his historic data series, aggregating information across the various principalities and other administrative districts that existed until German states unified in 1871. As another example, Maddison also disaggregates information about North and South Korea backwards in time. Additional details about these unit specific issues are available in the original source material. Documentation about how we merged all of the data sources together are available in our code files, which are publicly accessible. Importantly, because many of these units are subsets of larger ones (e.g., North and South Korea), analysts can aggregate the estimates of these two units together if necessary for a specific empirical application (we will continue to address these issues in future updates of our GDP, GDP per capita, and population estimates).

(2019).

[10]See appendix G for more details on when a country enters our dataset.

Table 1: Component Measures of Latent Variable Model

GDP items	**Coverage**	**Source**
Expenditure-side real GDP in prices constant across countries (in millions of 2017 international dollars, PPP).	1950-2019	Penn World Tables 10.0 (Feenstra, Inklaar and Timmer, 2015)
Expenditure-side real GDP in prices constant across countries and over time (in millions of 2017 international dollars, PPP).	1950-2019	Penn World Tables 10.0 (Feenstra, Inklaar and Timmer, 2015)
National accounts based real GDP in prices constant within countries (in millions of 2017 international dollars, PPP).	1950-2019	Penn World Tables 10.0 (Feenstra, Inklaar and Timmer, 2015)
GDP (in constant 2017 international dollars, PPP).	1960–2019	World Development Indicators (World Bank, N.d.)
GDP (in millions of 1990 international dollars, PPP). National currencies are converted to international dollars using Angus Maddison's purchasing power parities. GDP data limited to European countries and the United States, accounting for changing country boundaries.	1870–2001	Broadberry and Klein (2012)
GNP (in constant 1960 US dollars and prices, PPP). GNP data limited to European countries, accounting for changing country boundaries.	1830–1973	Bairoch (1976)
GDP (in constant 2010 USD, market exchange rate).	1960–2019	World Development Indicators (World Bank, N.d.)
GDP (in constant 2010 USD, market exchange rate). Data are available for the United States, the United Kingdom, Australia and Spain.	1790–2015	Measuring Worth (2019)
GDP per capita items	**Coverage**	**Source**
Real GDP per capita in prices constant across countries (in 2011 international dollars, PPP).	1500–2016	Maddison Project Database (Bolt et al., 2018)
National accounts based real GDP per capita in prices constant within countries (in 2011 international dollars, PPP) .	1500–2016	Maddison Project Database (Bolt et al., 2018)
Real GDP per capita in prices constant across countries and over time (in 2011 international dollars, PPP).	1500–2018	Maddison Project Database (Bolt and van Zanden, 2020)
GDP per capita (in constant 2017 international dollars, PPP).	1960–2019	World Development Indicators (World Bank, N.d.)
GDP per capita (in 1990 international dollars, PPP). Data limited to European countries and the United States, after accounting for changing country boundaries.	1870–2001	Broadberry and Klein (2012).
GNP per capita (in constant 1960 US dollars and prices, PPP). The data are limited to European countries, after accounting for changing country boundaries.	1830–1973	Bairoch (1976)
GDP per capita (in 1990 international dollars, PPP). Data limited to England/Great Britain, Holland/Netherlands, Italy, Spain, Japan, China, India.	1500–1850	Broadberry (2015)
GDP per capita (in constant 2010 USD, market exchange rate).	1960–2019	World Development Indicators (World Bank, N.d.)
GDP per capita data (in constant 2010 USD, market exchange rate). Data are available for the United States, the United Kingdom, Australia and Spain.	1790–2015	Measuring Worth (2019)
Population items	**Coverage**	**Source**
Total population (in millions).	1950–2019	Penn World Tables version 10.0 (Feenstra, Inklaar and Timmer, 2015)
Total population (in thousands at mid-year).	1500–2018	Maddison Project Database (Bolt and van Zanden, 2020)
Total population.	1960–2019	World Development Indicators (World Bank, N.d.)
Total population (in millions). Population data limited to European countries and the United States.	1870–2001	Broadberry and Klein (2012)
Population (in thousands). Data are available for the United States, the United Kingdom, Australia and Spain. The original unit of measurement and data coverage varies by country.	1789–2015	Measuring Worth (2019)
Total population. Data only available for China.	1500-1901	Deng (2004)

Construction of GDP and GDP per capita indicators

In conceptualizing GDP as a latent variable, we assume that each country-year unit has a "true" value that is unobserved or at least difficult to observe. As highlighted above, this is because each attempt at measuring GDP is likely imperfect and subject to uncertainty. A number of judgement calls have to be made when constructing a GDP indicator — prioritizing some aspects, such as applicability to measurement of economic growth, over others, such as comparability across countries. Therefore, depending on their purpose, existing GDP constructs capture varying aspects of the "true" underlying GDP. Existing approaches to the construction of GDP indicators vary broadly with respect to (a) the underlying primary data sources used[11] and (b) the methodology employed to render the values comparable across space and/or over time. Knowledge about differences in the construction of these GDP indicators can be used to enhance the accuracy of our model estimates. Since our model exploits agreement in observed values across different data sources to increase precision, but yields predictions in the original unit-of-measurement, we focus on differences in indicators' comparability across space and time to inform the construction of our model.

Different methods are used to make GDP comparable across time and/or space. Based on their construction, indicators for GDP and GDP per capita can be grouped into (1) series that are comparable over time but not space (except for the base year) via market exchange rates and are expressed in a constant currency unit, typically U.S. dollars, (2) series that comparable over time but not countries (except the base year) via purchasing power parity (PPP) exchange rates, (3) indicators that are comparable across countries but not over time via PPP, and, (4) series that are comparable across countries and over time via PPP. Tables 2 and 3 show the grouping of the indicators from each of the datasets we have collected.

Constant GDP data are comparable over time by adjusting for the effects of inflation within a country. Series, such as the World Bank's (N.d.) GDP in constant US dollars, are constructed by applying the market exchange between local currency units and the common currency in a given base year, such as 2010. However, because such series preserve the growth rates of the constant series in local currency units,[12] they should not be used to make comparisons of living standards

[11] An example would be the distinction between expenditure-side versus output-side GDP in the most recent release of the PWT (10.0), see Feenstra, Inklaar and Timmer (2015).

[12] https://datahelpdesk.worldbank.org/knowledgebase/articles/114943-what-is-your-constant-u-s-dollar-methodology, accessed 11 September 2018.

across countries, except in the base year.[13] For constant GDP to be comparable across countries in the base year, the series rely on the assumption of the law of one price, which states that "once prices are converted to a common currency, the same good should sell for the same price in different countries" (Rogoff, 1996, 649).

However, due to trade barriers and transaction costs, the law of one price is unlikely to hold, and market exchange rates that are used to construct constant series in a common currency fail to account for price differences between countries (Pinkovskiy and Sala-i-Martin, 2016, 22). PPP exchange rates are an alternative to market exchange rates for converting local currency units into a common currency. For years since 1968, the International Comparison Program (ICP) constructs PPP exchange rates based on the difference in the number of local currency units required to purchase a basket of similar goods in different countries.[14] Until recently, most PPP series have been constructed by making local currency units comparable via a contemporary cross-country price comparison benchmark for a specific year from the ICP and then using National Account growth rates to extrapolate the data back in time. For example, Angus Maddison compiled his data "starting from a modern-day cross-country income comparison — for the year 1990 — and then using growth rates of GDP per capita from (reconstructed historical) National Accounts to make comparisons for earlier years" (Bolt et al., 2018, 4). The Penn World Tables (PWT) RGDP^{NA} indicator (NA stands for National Account growth rates), the World Bank PPP figures, the Maddison Project Database (MPD) RGDPpc^{NA}, and most historical PPP series like Bairoch (1976) or Broadberry and Klein (2012) are constructed using this methodology. These series are adequate for a comparison of growth rates across countries, but not income levels (Feenstra, Inklaar and Timmer, 2015) and are comparable within a given country over time, but not across countries. We therefore include a cross-sectional random effect into the estimation of these variables.

By relying on multiple ICP benchmarks, recently released versions of the PWT (CGDP^{e} or CGDP^{o}) and the MPD (CGDPpc) provide PPP-based GDP and GDP per capita series that are adequate for comparing relative living standards across countries at a single point in time.[15]

[13] This is similar to the construction of the PWT RGDP^{NA} variable — a PPP series that is based on national account growth rates and should not be used for cross-country comparisons, except in the base year, see (Feenstra, Inklaar and Timmer, 2015).

[14] To date, the ICP has released price comparisons for the benchmark years 1970, 1975, 1980, 1985, 1996, 2005, 2011, and 2017.

[15] The practice of expressing these series in a constant currency unit, for example 2011 US$ to adjust for inflation in the United States, makes magnitudes comparable over time. However, "relative prices used for the cross-country

GDP indicators			
Comparable over time but not across countries via exchange rates	*Comparable over time but not across countries via PPP*	*Comparable across countries but not over time via PPP*	*Comparable across countries and over time via PPP*
• World Bank (N.d.) GDP in constant 2010 US$. • Measuring Worth GDP.	• World Bank (N.d.) GDP PPP in 2017 int. $. • PWT 10.0 $RGDP^{NA}$ in 2017 US$ (Feenstra, Inklaar and Timmer, 2015). • Broadberry and Klein (2012) GDP in 1990 int. $. • Bairoch (1976) GNP in constant 1960 US$.	• PWT 10.0 $CGDP^{e}$ in 2017 US$ (Feenstra, Inklaar and Timmer, 2015).	• PWT 10.0 $RGDP^{e}$ in 2017 US$ (Feenstra, Inklaar and Timmer, 2015).
Cross-sectional random effect $\gamma^{*}_{gdp[i]}$	**Cross-sectional random effect $\gamma_{gdp[i]}$**	**Yearly random effect λ_t**	
Cross-sectional conversion $\beta_{gdp[i]}$ to PPP			

Table 2: Grouping of GDP indicators based on the methodologies used to make series comparable over time and/or across countries. PWT indicates data from Penn World Tables. The component data sets from the MeasuringWorth project have been converted to 2010 USD (see Hutchinson and Ploeckl 2018; Johnston and Williamson 2018; Prados-de-la-Escosura 2018; Thomas and Williamson 2018). We define the relationship between the cross-sectional random effects and conversion factor with respect to the other model parameters below.

These GDP figures are considered "current-price" PPP series (Bolt et al., 2018; Feenstra, Inklaar and Timmer, 2015). Chaining PPP benchmarks from multiple rounds of the ICP in the PWT 10.0 $RGDP^{e}$ and the MDP 2020 RGDPpc indicator allows for a comparison of living standards across countries and over time. For benchmark years, GDP is made comparable across space using PPP exchange rates. Comparability over time is achieved by interpolating between benchmark years—which results in GDP growth rates that are static between benchmark years and typically differ from National Account rates (Feenstra, Inklaar and Timmer, 2015; Bolt and van Zanden, 2020, 3156).

Tables 2 and 3 summarize the grouping of the GDP and GDP per capita items based on their comparability across space and time, as well as differences between constant and PPP series. We estimate the latent trait by including all GDP and GDP per capita series in Tables 2 and 3 and adjust for differences between constant and PPP series via a country-specific conversion factor

comparisons differ over time" (Bolt et al., 2018, 5).

GDP per capita indicators			
Comparable over time but not across countries time via exchange rates	*Comparable over time but not across countries via PPP*	*Comparable across countries but not over time via PPP*	*Comparable across countries and over time via PPP*
• World Bank (N.d.) GDP per capita in constant 2010 US$. • MeasuringWorth GDP per capita.	• World Bank (N.d.) GDP PPP per capita in 2017 int. $. • MDP 2018 RGDPpc^{NA} in 2011 US$ (Bolt et al., 2018). • Broadberry (2015) GDP per capita in 1990 int. $. • Broadberry and Klein (2012) GDP per capita in 1990 int. $. • Bairoch (1976) GNP per capita in constant 1960 US$.	• MDP 2018 CGDPpc in 2011 US$ (Bolt et al., 2018).	• MDP 2020 RGDPpc in 2011 US$ (Bolt and van Zanden, 2020).
Cross-sectional random effect $\gamma^*_{gdppc[i]}$	**Cross-sectional random effect** $\gamma_{gdppc[i]}$	**Yearly random effect** λ_t	
Cross-sectional conversion $\beta_{gdppc[i]}$ **to PPP**			

Table 3: Grouping of GDP per capita indicators based on the methodologies used to make series comparable over time and/or across countries. MDP indicates data derived from Maddison Project Database. The component data sets from the MeasuringWorth project have been converted to 2010 US$ (see Hutchinson and Ploeckl 2018; Johnston and Williamson 2018; Prados-de-la-Escosura 2018; Thomas and Williamson 2018). We define the relationship between the cross-sectional random effects and conversion factor with respect to the other model parameters below.

β_c, as well as discrepancies in the comparability across space and over time via the inclusion of random effects. The construction of the constant (column 1) and PPP series (column 2) that are comparable over time but not across countries is similar. Cross-sectional comparability is established at a single point in time — the base year — via market exchange rates (for the constant series) or via PPP exchange rates. The data is then projected back in time via growth rates from national accounts. To account for this limitation in the comparability of the series, we include a cross-sectional random effect $\gamma_{k[i]}$ when projecting from the latent trait to the estimated country-year value for each individual data source. To account for the limitation of the GDP and GDP per capita items that are comparable across countries via ICP benchmarks but not over time (column 3), we include a year-specific random effect λ_t when projecting from the latent trait to the estimated country-year value. No random effect is included when projecting from the latent trait to the GDP series that is comparable across countries and over time by construction via chaining of ICP benchmarks (column 4). What this all means is that we only use comparable variation to inform the estimation of the underlying latent trait, the rest of the variation for any particular observed variable is captured by the appropriate random effect parameter or conversion parameter.

Model Description, Specification, and Estimation

Estimation procedure

The empirical goal of this article is to specify a dynamic latent variable model that will allow us to leverage the available information across different contemporary and historical sources of data on GDP, population, and GDP per capita. Unlike other latent variable models, we are not directly interested in the estimated latent traits (or the random effect parameters or conversion parameters). Instead, we use these estimated traits to generate prediction intervals for each of the observed variables in their original unit-of-measurement. The model therefore allows us to leverage the information across all of the observed data and to generate predictions for each of the individual observed variables. In this way, interested users can select their preferred dataset for empirical applications in conjunction with the prediction intervals generated from our model. Importantly, the prediction intervals are available for every unit included in the dataset. As we

discuss later in the article, the size of the prediction interval is related to total coverage and agreement of the observed variables. Importantly, when using the latent intervals in terms of each of the original variables, they should still primarily be used only for the relevant type of comparison (e.g., cross-sectional, time-series).

Latent Variable Model

We estimate the latent trait by including all GDP and GDP per capita series in Tables 2 and 3 and adjust for differences between constant and PPP series, as well as discrepancies in the comparability across space and over time via the inclusion of random effects. Our model estimates derive precision from a combination of (a) agreement across different data sets and (b) extent of cross-sectional and temporal coverage within a given data source. The more parameters we introduce, the less variance can be exploited for estimation because it is subsumed into these model parameters. To specify the dynamic latent variable model, let $i = 1, \ldots, N$, index cross-sectional units and $t = 1, \ldots T$, index time periods. For each country-year unit, $j = 1, \ldots, J$ indexes the observed variables y_{itj}. For each set of observed variables, we use indicator k to denote variable type: *gdp*, *gdppc*, *pop*.

We use a standard dynamic setup to relate the estimate of the latent trait in year t to the value in the prior year $t-1$. $\theta_{k[it]} \sim \mathcal{N}(0,1)$ for all i when $t = 1$ (the first year a country enters the dataset). When $t > 1$, the standard normal prior is centered around the latent variable estimate from the previous year such that: $\theta_{k[it]} \sim \mathcal{N}(\theta_{k[it-1]}, \sigma_k)$.

Because the observed variables y_{itj} represent three different concepts — GDP, population, and GDP per capita — we estimate three latent variable parameters, where k indexes categories *gdp*, *gdppc*, *pop*. This allows us to define the set of y_{itj} that we observe for k latent traits, where $y_{itj} * 1\{y \in \pi_k\}$. Note that the sets of observed variables do not need to be mutually exclusive to estimate each of the three latent traits $\theta_{k[it]}$. For our model, we do not allow the observed variables to inform the estimation of multiple latent variables, but this is a possibility in other applications (for details see Gelman and Hill, 2007; Imai, Lo and Olmsted, 2017). Overall, this notation allows us to denote the set of observed variables used to estimate each of the three underlying latent variables such that $\pi_{gdp} = \{y_{it1}, ..., y_{it7}\}$, $\pi_{gdppc} = \{y_{it8}, ..., y_{it13}\}$, $\pi_{pop} = \{y_{it14}, ..., y_{it20}\}$.

For the latent GDP per capita items, we estimate $\theta_{gdppc[it]} \sim \mathcal{N}(0,1)$ for all i when $t = 1$ (the

first year a country enters the dataset). When $t > 1$, the standard normal prior is centered around the latent variable estimate from the previous year such that: $\theta_{gdppc[it]} \sim \mathcal{N}(\theta_{gdppc[it-1]}, \sigma_{gdppc})$. For the latent population items, we estimate $\theta_{pop[it]} \sim \mathcal{N}(0, 1)$ for all i when $t = 1$ (the first year a country enters the dataset). When $t > 1$, the standard normal prior is centered around the latent variable estimate from the previous year such that: $\theta_{pop[it]} \sim \mathcal{N}(\theta_{pop[it-1]}, \sigma_{pop})$. The latent variables themselves are estimated with uncertainty. The first year each country enters the model, the variances for these parameters are set to 1. For all years after $t = 1$, σ_{gdppc} and σ_{pop} are drawn from a half-normal distribution $\mathcal{HN}(0, 1)$.[16] Finally, because GDP is a function of the population and GDP per capita estimates, we do not estimate this latent variable independently but instead estimate it as a function of the other two latent traits. Specifically, we set the parameters for the GDP latent trait using the following deterministic transformation using the GDP per capita and population latent variables themselves: $\theta_{[gdp]it} \sim \mathcal{N}(\theta_{gdppc[it]} + \theta_{pop[it]},\ \sigma_{gdp})$. We set σ_{gdp} to $\sqrt{\sigma^2_{pop} * \sigma^2_{gdppc}}$.

The latent variables are estimated by linking each of these parameters to the sets of observed GDP, population, or GDP per capita variables. Since all of the GDP, population, and GDP per capita variables are continuous, we specify a Gaussian link function with a unique error term for each of the three types of variables τ_k: $\{\ \tau_{gdppc},\ \tau_{gdp},\ \tau_{pop}\ \}$. These τ_k parameters are estimates of model level uncertainty, which link each of the latent variables to the sets of observed items. For the intercept parameters α_j, we center the standard normal prior around the mean value of the observed data with a relatively large variance (low precision): $\alpha_j \sim \mathcal{N}(\bar{y}_j, 4)$. We choose the mean value of the observed variables because the latent traits themselves are centered around 0. The intercept parameter therefore transforms the latent trait into the unit-of-measurement of the original observed variable.

For the observed GDP and GDP per capita items that are comparable over time but not across countries, as described above, we add a country-specific random effect $\gamma_{k[i]} \sim \mathcal{N}(0, 1)$. For observed GDP and GDP per capita items that are comparable across countries but not over time, we add a year-specific random effect $\lambda_t \sim \mathcal{N}(0, 1)$.

For identification of the model we do not estimate a slope parameter. Instead, we assume

[16] We choose the half normal distribution because the standard deviation has strictly positive support. The choice of prior values for these parameters ensures that the underlying latent variables will be in roughly unit scale.

a one-unit change in the latent trait is equivalent to a one-unit change in the original observed variable. However, because the GDP and GDP per capita items estimated using exchange rates, which are not always one-to-one with items created using purchasing power parity dollars, we estimate a conversion parameter for each country: $\beta_{k[i]} \sim \mathcal{HN}(1, 0.5)$.[17]

The modeling innovations we have introduced in this article allow information from both the GDP per capita and GDP variables to inform the estimation of the underlying latent traits, because we organize the three types of observed variables in three sets such that $y_{itj} * 1\{y \in \pi_k\}$. We combine all of the parameters described above using a likelihood function that links the observed data to the estimated parameters: $\mathcal{L}(\alpha, \beta, \gamma, \lambda, \tau, \theta | y_{itj} * 1\{y \in \pi_k\})$, which we specify as:

$$\begin{aligned}\mathcal{L}(\alpha, \beta, \gamma, \lambda, \tau, \theta | y_{itj} * 1\{y \in \pi_k\}) = \prod_{i=1}^{N} \prod_{t=1}^{T} \prod_{j=1}^{J} \prod_{k \in K} & \underbrace{\mathcal{N}(\alpha_j + \theta_{k[it]}, \tau_k)^{(v_j)(r_j)}}_{\text{comparable units by PPP}} * \\ & \underbrace{\mathcal{N}(\alpha_j + \gamma_{k[i]} + \theta_{k[it]}, \tau_k)^{(1-v_j)(1-r_j)}}_{\text{comparable countries by PPP}} * \\ & \underbrace{\mathcal{N}(\alpha_j + \lambda_t + \theta_{k[it]}, \tau_k)^{(v_j)(1-r_j)}}_{\text{comparable years by PPP}} * \\ & \underbrace{\mathcal{N}(\alpha_j + \gamma^*_{k[i]} + \beta_{k[i]} * \theta_{k[it]}), \tau_k)^{(1-v_j)(r_j)}}_{\text{comparable countries by exchange rates}}\end{aligned}$$

v_j and r_j are indicator variables that determine which portion of the likelihood function a particular item should be passed through. For all population items, GDP and GDP per capita items that are comparable across all units by PPP, $v_j = 1$ and $r_j = 1$; for GDP and GDP per capita items that are comparable across countries by PPP, $v_j = 1$ and $r_j = 0$; for GDP and GDP per capita items that are comparable across years by PPP, $v_j = 0$ and $r_j = 1$; and for GDP and GDP per capita items that are comparable across countries by exchange rate $v_j = 0$ and $r_j = 1$. For all population items, $v_j = 1$ and $r_j = 0$. All of the prior distributions are summarized in Table 4.[18]

[17] These conversation factors tend to be close to 1 and practically only influence the values of the model estimates in the post 1950 period because all observed dataset values prior to 1950 are in PPP international dollars. We select a half-normal prior distribution because the GDP values in terms of PPP or exchange rate dollars are always positively correlated.

[18] The model is estimated with four MCMC chains to run for 2,000 iterations each. The first 1,000 iterations were thrown away as a warmup period. The 4,000 samples were thinned by a factor of 2 and are used to generate the posterior prediction intervals for the original observed variables. The sampler was implemented using the Stan

Table 4: Prior Distribution for Latent Variables and Model Level Parameter Estimates

Parameter	Prior		
Country i latent population estimate in first year t	$\theta_{pop[it=1]}$	$\sim$	$\mathcal{N}(0,1)$
Country i latent population estimate in all other years	$\theta_{pop[it]}$	$\sim$	$\mathcal{N}(\theta_{pop[t-1]}, \sigma_{pop})$
Latent population standard deviation	σ_{pop}	$\sim$	$\mathcal{HN}(0,1)$
Country i latent GDPPC estimate in first year t	$\theta_{gdppc[it=1]}$	$\sim$	$\mathcal{N}(0,1)$
Country i latent GDPPC estimate in all other years	$\theta_{gdppc[it]}$	$\sim$	$\mathcal{N}(\theta_{gdppc[t-1]}, \sigma_{gdppc})$
Latent GDP standard deviation	σ_{gdppc}	$\sim$	$\mathcal{HN}(0,1)$
Country i latent GDP estimate in all years	$\theta_{gdp[it]}$	$\sim$	$\mathcal{N}(\theta_{gdppc[it]} + \theta_{pop[it]},\ \sigma_{gdp})$
Latent GDP standard deviation	σ_{gdp}	$\leftarrow$	$\sqrt{\sigma^2_{pop} * \sigma^2_{gdppc}}$
Model j intercept "difficulty parameter"	α_j	$\sim$	$\mathcal{N}(\bar{y}_{itj}, 4)$
Item-type cross-sectional conversion parameter	$\beta_{k[i]}$	$\sim$	$\mathcal{HN}(1, 0.5)$
Item-type cross-sectional random effect (ppp)	$\gamma_{k[i]}$	$\sim$	$\mathcal{N}(0,1)$
Item-type cross-sectional random effect (exchange rate)	$\gamma^*_{k[i]}$	$\sim$	$\mathcal{N}(\gamma_{k[i]}, \sigma_\gamma)$
Uncertainty for exchange rate cross-sectional random effect	σ_γ	$\sim$	$\mathcal{HN}(0,1)$
Temporal random effect	λ_t	$\sim$	$\mathcal{N}(0,1)$
Model uncertainty for all population items	τ_{pop}	$\sim$	$\mathcal{HN}(0,1)$
Model uncertainty for all GDP per capita items	τ_{gdppc}	$\sim$	$\mathcal{HN}(0,1)$
Model uncertainty for all GDP items	τ_{gdp}	$\sim$	$\mathcal{HN}(0,1)$

software (Stan Development Team, 2021). Conventional diagnostics (i.e. trace plots, effective sample size, and R-hats) all suggest convergence (Gelman and Hill, 2007).

Model Validation

If our latent variable model is doing a good job of estimating the observed dataset values, then most country-year dataset values should fall within the estimated intervals generated from our model. We present several validation assessments in this section which makes such comparisons.

Comparison of the Distribution of Observed Dataset Values and the Distribution of Posterior Predicted Intervals

For our first assessment, we compare the distribution of observed dataset values from each dataset to the distribution of country-year point estimates generated from the latent variable model. If our latent variable model is doing a good job at estimating GDP, GDP per capita, and population (in the unit-of-measurement for each variable), then the distributions will line up with one another.

Figure 1 and Figure 2 display the distributions of the original observed dataset values (y_{itj}) alongside the mean or expected value of the posterior prediction intervals ($E(\tilde{y}_{itj})$). Agreement between observed country-year-variable values (orange boxes) and the posterior predicted point estimates for which the observed value is not missing (light grey boxes). Dark grey boxes show the distribution for the full range of the posterior predicted point estimates (including estimates for which the original value is missing).

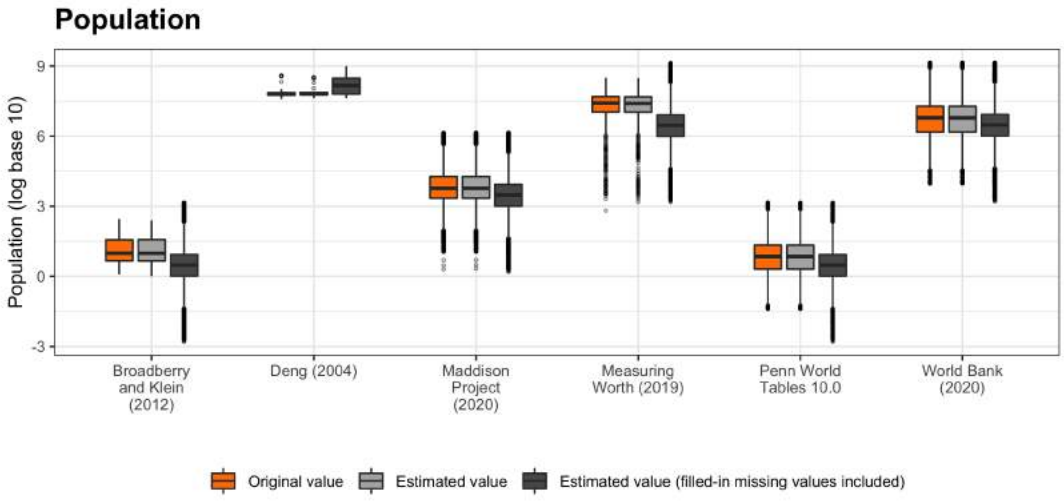

Figure 1: Agreement between observed country-year-variable values (orange boxes) and the posterior predicted point estimates for which the observed value is observed (light grey boxes). Dark grey boxes show the distribution for the full range of the posterior predicted point estimates (including estimates for which the original value is missing). Across all variables, these estimated values have a lower median value due to a bias in the missingness in the original data. We have more missing observations in earlier years, for which population levels are lower than in later years in the series. Conveniently for users, the posterior predicted values are estimated using the original unit-of-measurement. This means that the visual discrepancy, for example between the World Bank (N.d.) and the other population data, is not an empirical discrepancy. It is simply the difference in the unit-of-measurement used for the observed values (i.e., the World Bank (N.d.) publishes the absolute GDP and population values while others present these in per thousand units).

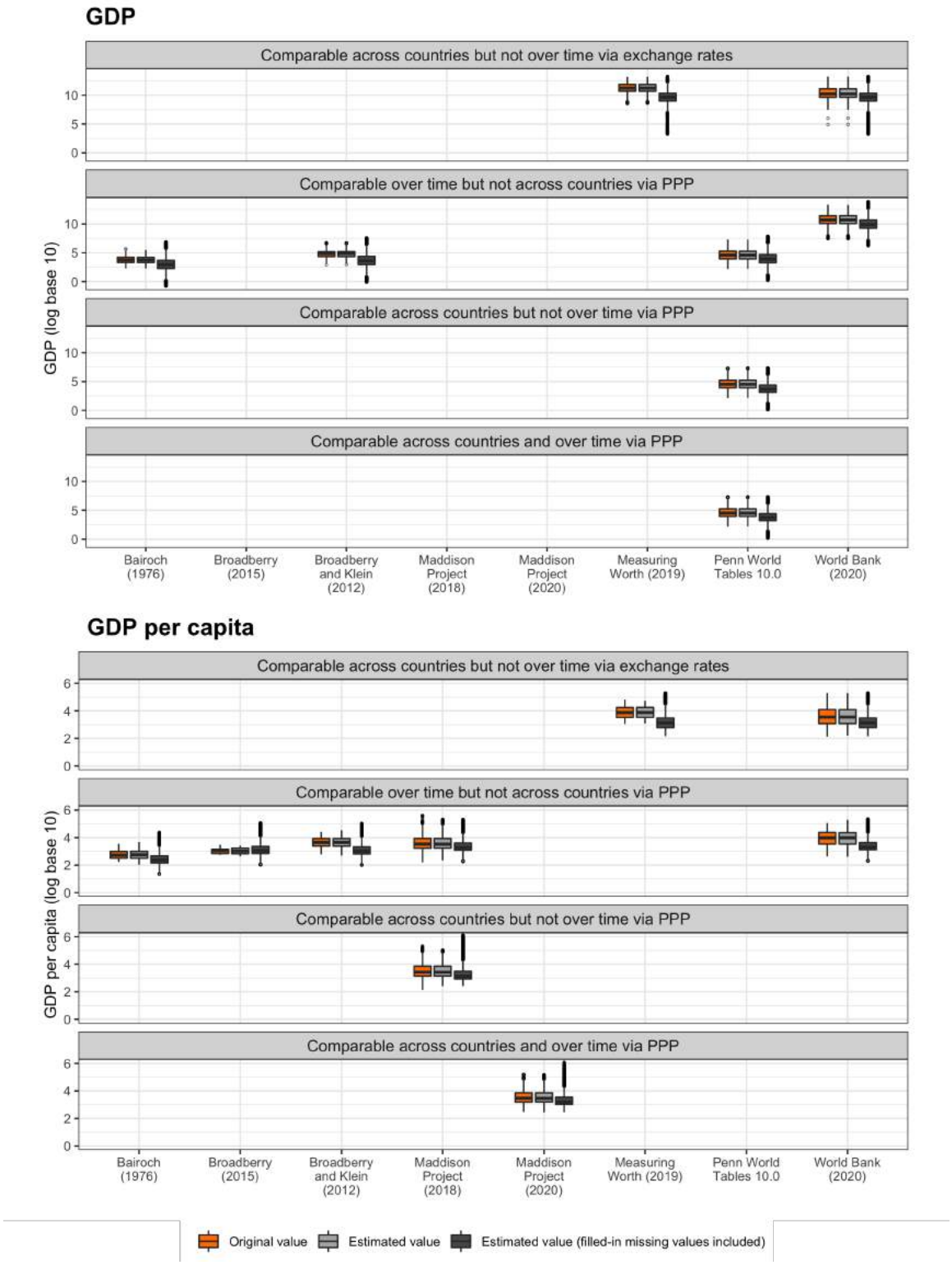

Figure 2: Agreement between observed country-year-variable values (orange boxes) and the posterior predicted point estimates for which the observed value is observed (light grey boxes). Dark grey boxes show the distribution for the full range of the posterior predicted point estimates (including estimates for which the original value is missing). Across all variables, these estimated values have a lower median value due to a bias in the missingness in the original data. We have more missing observations in earlier years, for which GDP, and GDP per capita levels are lower than in later years in the series.

Uncertainty of Estimates by Country

If our latent variable model is doing a good job of estimating the observed dataset values, then most country-year cases should fall within the estimated intervals generated from our model. Our model generates estimated intervals for every dataset included in our model.

To demonstrate that our latent variable model provides useful estimates for GDP per capita, GDP, and population, we compare our country-year estimates for country-year units with varying levels of dataset coverage. We begin with three prominent country examples that span the full timer period covered by our model: the United Kingdom, Japan, and China. Figure 3, 4, and 5 display country-year prediction intervals (grey lines) with observed dataset values (black points) for the three countries. Online appendix C displays a more diverse sample of country examples that vary in terms of size, geographic location, and date of entry into the international system (i.e., United States, Netherlands, France, Spain, Italy, Russia, Sweden, Albania, India, Colombia, Brazil, Democratic Republic of Congo/Zaire, Uganda, Iran, Afghanistan, North Korea, Pakistan, Kosovo, East Timor, and Eritrea). Online appendix C also displays yearly cross-sections of countries for early historic periods (1500, 1600, 1700, 1800).

These graphs display the estimated intervals (grey lines) generated from our model for each country-year unit. The graphs also show the location of the observed dataset values (black points) within the estimated intervals generated from our model. Finally, the graphs show the estimated intervals for years of missing coverage when those dataset values do not exist (no black dot). For country-year units that have one or more observed dataset values, the intervals are relatively small. Conversely, the further away a unit is from an observed dataset value, the greater the uncertainty and the estimate of the standard deviation. Importantly, the level of uncertainty is a country-year unit parameter. We can see this in the country examples below for the early time periods when dataset coverage is much more sparse. In general though, our estimated intervals capture the observed dataset value most of the time. Below we systematically summarize how well our model does for all country-years and variables.

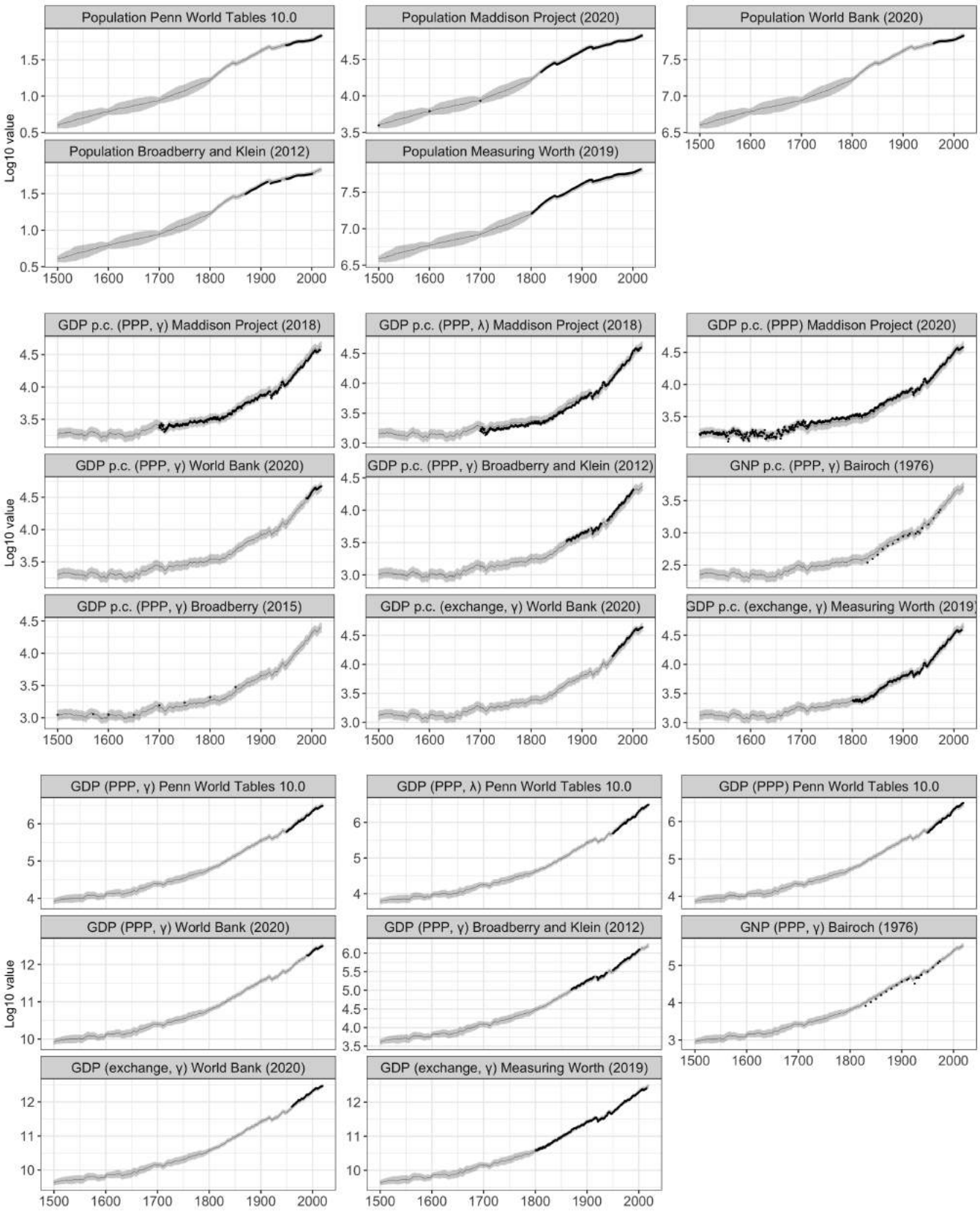

Figure 3: Posterior prediction intervals (grey lines) with ± 1 standard deviation confidence bands and observed variables (black points) for the United Kingdom.

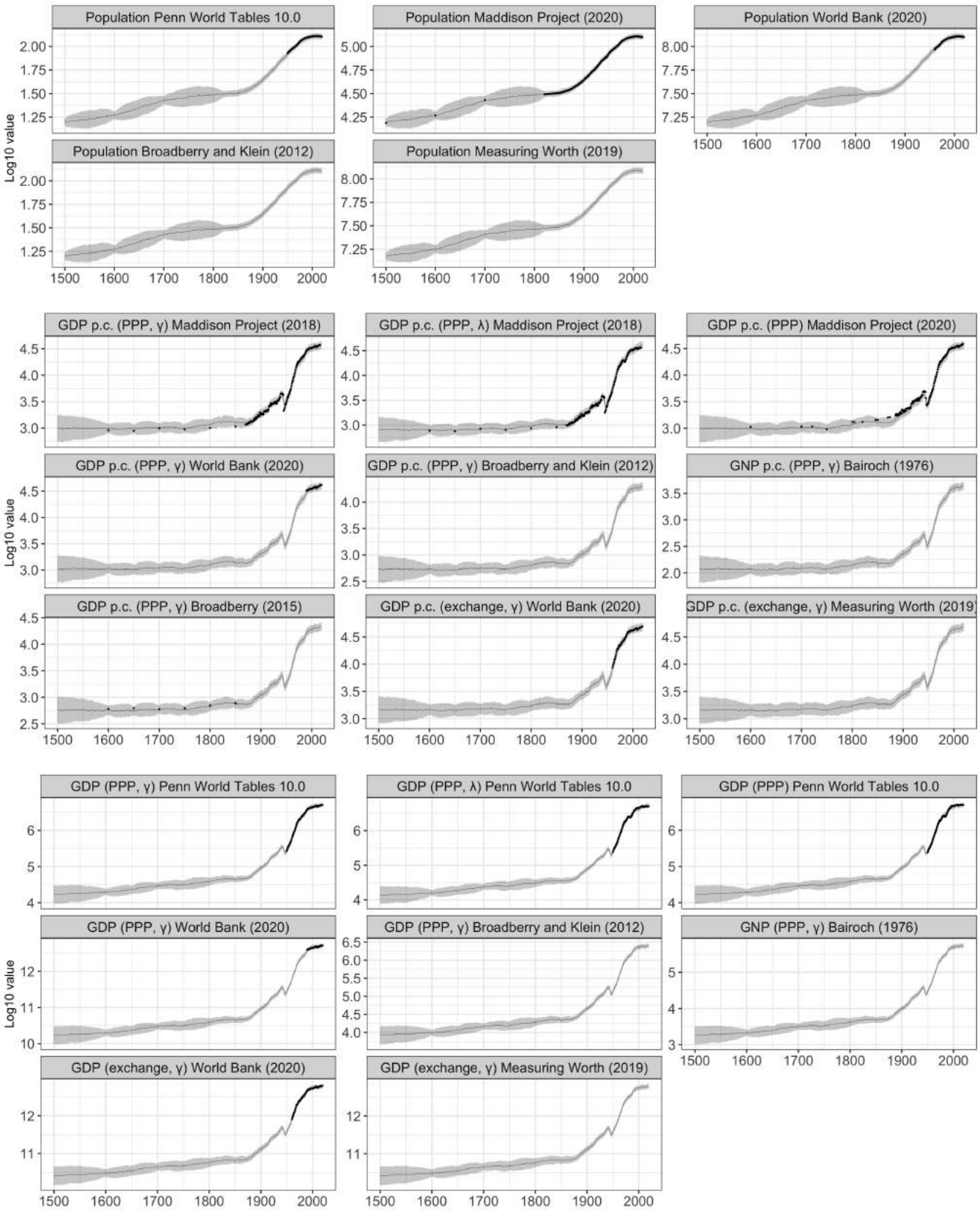

Figure 4: Posterior prediction intervals (grey lines) with ± 1 standard deviation confidence bands and observed variables (black points) for Japan.

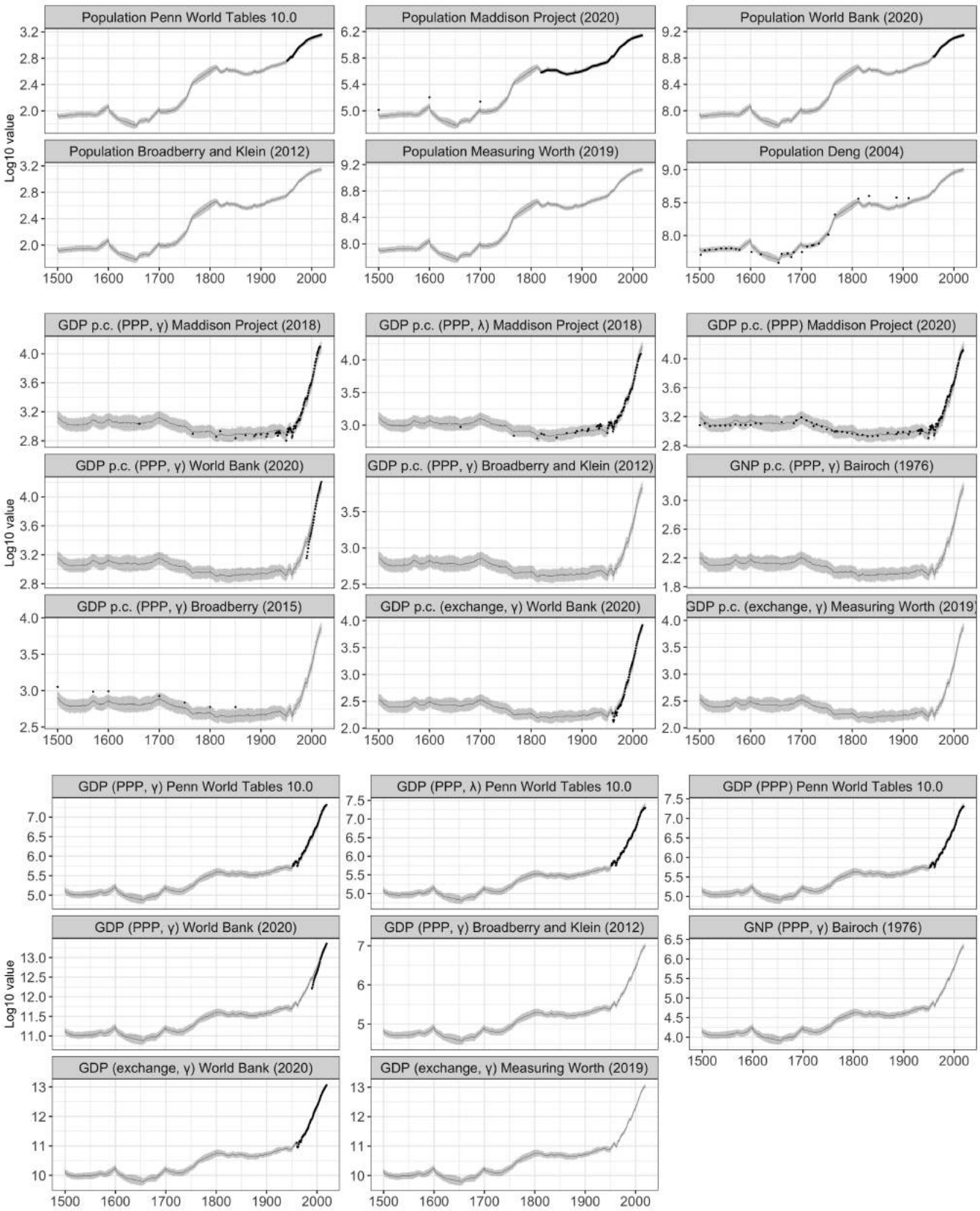

Figure 5: Posterior prediction intervals (grey lines) with ± 1 standard deviation confidence bands and observed variables (black points) for China.

Uncertainty of Estimates by Dataset Coverage

To demonstrate that our latent variable model provides useful estimates for GDP per capita, GDP, and population, we compare our country-year estimates for country-year units with varying levels of dataset coverage. First we consider the relative level of uncertainty for estimates with no dataset coverage relative to units with coverage from 1 or more datasets. In general our country-year intervals are larger for country-year units without any data coverage (the largest intervals are for country-year units that are farthest away from a unit with an observed dataset value). Conversely, these intervals decrease in size for country-year units as additional observed dataset values become available (i.e., more coverage), with the caveat that the intervals will increase if the observed dataset values disagree. If our model is doing a good job of capturing the observed dataset values, then our estimated intervals will contain the observed dataset value unless there is a high degree of disagreement. First we consider the relative level of uncertainty for each of our country-year estimates (see Figure 6) then we use this information to understand how well our estimates capture the observed dataset values (Figure 7). We do this with country-year standard deviations statistics in the first assessment and country-year Z-scores in the second assessment.

How do we measure the uncertainty of each county-year? As the number of observed pieces of information increases — the number of observed variables per country-year unit — the country-year standard deviation (i.e., the level of uncertainty) decreases, which is visualized in Figure 6. Because the latent variable model is dynamic, even country-year units with 0 observed dataset values have some information — the latent variable estimate from the prior year with a dataset value — with which to estimate a value of the latent variable. The country-year cases with 0 observed dataset values have the largest standard deviations as seen in the left most boxes for each panel of Figure 6.

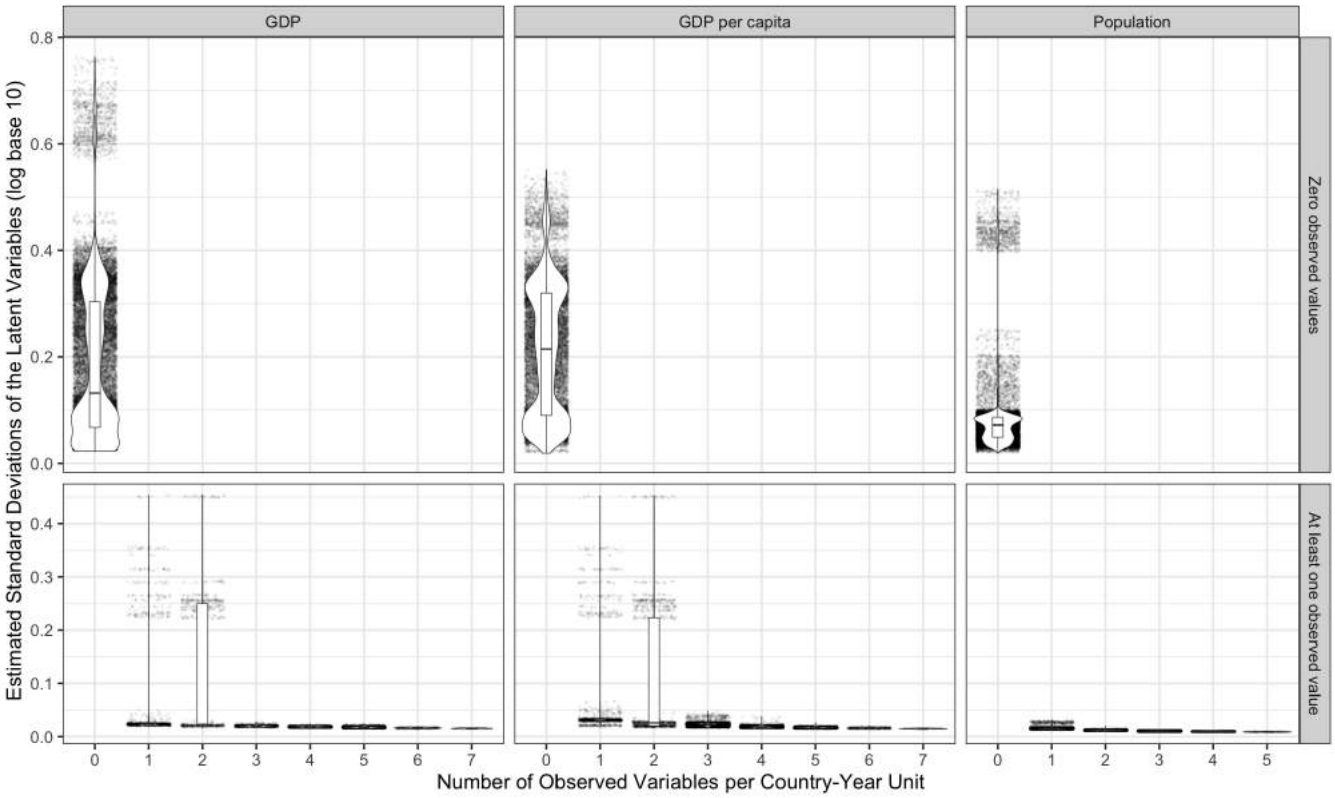

Figure 6: The plot illustrates the relationship between the amount of information — the number of observed variables per country-year unit — and the level of uncertainty for each country-year estimate of the latent variable. As the number of observed pieces of information increases, the level of uncertainty decreases. Because the latent variable model is dynamic, even country-units with 0 observed variables have some information — the latent variable estimate from the prior year — with which to estimate a value of the latent variable. For this reason, the amount of variation of the latent standard deviation is greatest for those units without information. The further away a unit is from an observed variable, the greater the uncertainty and the estimate of the standard deviation. Importantly, the level of uncertainty is a country-year unit parameter. This parameter can be used to incorporate uncertainty in statistical models that include the new estimates of GDP, GDP per capita, and population.

Validation of Predicted Intervals and Dataset Values

Next, we use these country-year standard deviations to estimate the precision of the new posterior predictions relative to the original observed variable, which are a type of posterior predictive check (Gelman and Hill, 2007; Reuning, Kenwick and Fariss, 2019). That is, how closely do our country-year intervals correspond to the observed dataset values and what does it mean when they do not?

Specifically, to assess the correspondence between the new country-year intervals relative to the original observed variable we estimate a country-year unit Z-score, which is a comparable measure of the distance between the observed dataset value and the center of the estimated country-year distribution from our model. Because the original variables and the predicted variables are both continuous, interval-level, and normally distributed, we are able to compare the original observed country-year unit variable's position relative to the country-year unit distribution of the posterior predicted interval. The Z-score's value represents the positions of the observed variable's value relative to the center of the posterior predicted interval for each of the country-year units and for every dataset variable. A value of 0 indicates that the observed variable's value falls directly at the center of the posterior predicted interval. Z-score units above and below 0 represent standard deviation differences from the center value of the posterior predicted interval. We can therefore use these values to get a sense of how close the observed value for each country-year variable is to its corresponding posterior predicted interval. The country-year-variable Z-score values take the form: $z_{itj} = \frac{y_{itj} - E(\tilde{y}_{itj})}{\sigma_{\tilde{y}_{itj}}}$, where y_{itj} is the observed value for the country-year-variable, $E(\tilde{y}_{itj})$ is the expected value or mean for the posterior predicted interval, and $\sigma_{\tilde{y}_{itj}}$ is the standard deviation for the posterior predicted interval.

Figure 7 illustrates the average number of observed values that fall within ± 1 and ± 2 standard deviation(s) of the posterior estimate for each item (i.e., $z_{itj} \leq 1$ or $z_{itj} \leq 2$). 80.5% of the observed GDP per capita dataset values fall within ± 1 standard deviations of the posterior estimates. 81.9% of the observed GDP dataset values fall within ± 1 standard deviations of the posterior estimates. And finally, 95% of the observed population data set values falling within ± 1 standard deviation of the posterior estimates. Nearly 99% of the country-years for which we have GDP and GDP per-capita values during this period are within ± 2 standard deviations of the posterior mean. Our

model therefore produces estimates that are very close to the observed dataset values in most cases.

To put these proportions into context, recall that each estimated country-year distribution is normally distributed. As the number of observed data-set values increase for any given country-year unit, the standard deviation of the country-year distribution decreases. So our model does an excellent job of approximating the observed dataset value much more often than we would expect by chance. This is because on average, we would expect the observed dataset values to within ± 1 standard deviation only 68.27% of the time and within ± 2 standard deviations only 95.45% of the time.[19]

There is some variation across historical periods. For the period from 1500-1799, our estimate for GDP per capita are based almost exclusively on data from the Maddison project. Data from Broadberry (2015) cover the following countries: the UK, Holland, Spain, Italy, and China from 1500, and Japan and India from 1600. We also have estimates for the U.S. from Measuring Worth (2019): U.S. from 1790-1800 and we have estimates from Deng (2004) from 1500, but only for population, not GDP or GDP per-capita. For all other countries, during this period our estimates are based solely on data from the Maddison Project. Therefore there is very little disagreement between the original Maddison values and our estimates of the Maddison values because our estimated intervals are centered over the observed data point from the Maddison.

As more observed data becomes available in the later periods in Figure 7 (1800-1949 and 1950-2018), we begin to see the consequence of more disagreement between our estimated intervals and the original values for many of the datasets. We can see this in Figure 7 that the proportion of observed values that fall within ± 1 standard deviation of the posterior mean decreases but that nearly all the observed values still fall within ± 2 standard deviation of the posterior mean (i.e., $-2 \leq z_{itj} \leq 2$). The shift away from the observed values is caused by the disagreement between indicators, which is a useful feature of our model because it has allowed us to quantify how much disagreement exists between our latent estimates and the original observed value.

In sum, Figure 7 illustrates the high level of agreement between the original observed variables

[19]Note further that these proportions are generated only for country-years with an observed dataset value, which means the standard deviations are quite small so the the observed dataset value is very close to the center for the posterior predicted intervals. Our model achieves this high level of coverage because it is dynamic, which means that observed dataset information from one year, informs subsequent years in time, which both reduces variation (the country-year standard deviations) and improves the fit of the model.

and the new posterior prediction intervals. These intervals are larger for country-year units without any data coverage (the largest intervals are for country-year units that are farthest away from a unit with an observed dataset value). In general, these intervals decrease in size for country-year units as additional observed dataset values become available (i.e., more coverage), with the caveat that the intervals will increase if the observed dataset values disagree. Overall, the new estimates closely approximate the original data, but do not always agree. The cases of disagreement between original observed variable and posterior prediction represent useful deviant cases that can inform future research.[20]

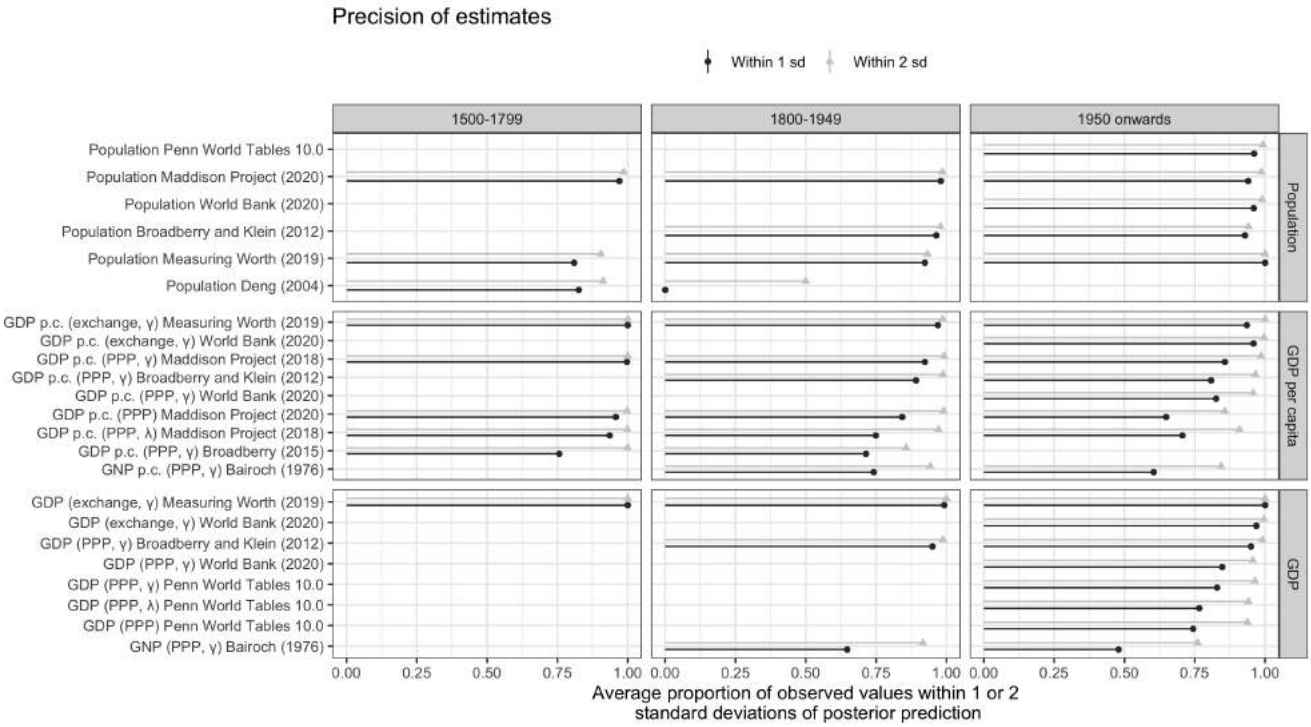

Figure 7: Average number of observed values that fall within ±1 or ±2 standard deviation(s) of the posterior estimate for each item of the latent variable model, separated into 3 time periods.

[20]See advice and examples for using deviant case studies in Eck and Fariss (2018); Fariss (2019); Seawright (2016).

Discussion: Coverage, Uncertainty, and Bias

Our latent variable model and the new estimates it produces are important new measurement tools for researchers interested in the study of politics and economics broadly, and conflict specifically. The modeling strategy in this project addresses problems in existing indicators regarding (1) measurement coverage, (2) measurement uncertainty, and (3) measurement bias. In the next three subsections, we discuss how our measurement model and new estimates help to address each of these problems in more detail. In particular, we provide guidance on how to use the new estimates in empirical models of conflict, how to incorporate uncertainty from the estimates into such models, and how to make theoretically meaningful extensions to the measurement model itself.

Measurement coverage

Our latent variable model estimates significantly increase the temporal coverage of GDP, GDP per capita, and population indicators. This allows conflict scholars to study relationships between economic development and key variables of interest, such as democracy, the incidence of violent conflict, or the use of repressive tactics like politically motivated killings, over a much longer time period than previously possible. We illustrate the potential of the new economic variables for studying these relationships over the long term by tracking their correlation over windows of time from the early 1800s (and for some plots even earlier) through the present.

Figure 8 plot Spearman rank-order correlation coefficients for the relationship between economic development and life expectancy (Max Roser and Ritchie, 2013), literacy rates (Roser and Ortiz-Ospina, 2016), democracy (Coppedge et al., 2019), repression (Coppedge et al., 2019), power projection capabilities (Modelski and Thompson, 1988), and conflict (Sarkees and Wayman, 2010), respectively, for ten-year rolling temporal windows (similar figures for 1-year and 50-year periods are available in online appendix D). We measure economic development by taking $m = 100$ draws from the posterior distribution of our latent GDP per capita variable and then correlating these draws with another observed variable. This procedure allows us to incorporate uncertainty from the distribution of GDP per capita when estimating the correlation coefficient for each time-period (with 95%-intervals generated from the distribution of 100 correlation coefficients).

Each of the plots illustrates considerable temporal variation in the magnitude of the correlation

between these variables. In short, there are not uniform relationships between these important variables over time. For example, the correlation between life expectancy and GDP per capita has dramatically increased since governments first started to track life expectancy, while the relationship between literacy rates and GDP is strong in every year for which we have data other than in the early 1800s. Interestingly, the build up of the industrial revolution appears to temporarily weaken these relationship for several decades in the 19th century.

The relationship between economic development and democracy is strong but increases after 1850 before decreasing again around 1950. A similar pattern emerges for repression in 1850 but it changes less in the post 1945 era.

Finally, the relationship between GDP and a measure of power projection is quite stable over time. Contrast this pattern with the correlations over time between economic development and conflict, where the patterns are much more stable over the full period. In early periods, countries with higher levels of GDP per capita were engaged in more conflict (MIDs or Wars) than countries with less GDP per capita. Today, the correlation of 0 suggests that both rich and poor countries are engaging in this type of interstate interaction, though the rate of this type of interaction has also been in decline.

While the full exploration of these new patterns is beyond the scope of this article, we hope that future research will investigate the historic processes related to this variation using our newly developed estimates of GDP, GDP per capita, and population, and evaluate the extent to which previously studied relationships between these variables are generalizable to earlier time periods. In brief, we expect that as countries transition through the industrial revolution, the relationship between wealth and the different variables we consider changes. Many of the relationships are much more volatile in the 19th century but stabilize in the early to mid 20th century (online appendix D for additional visualization and discussion of these relationships).

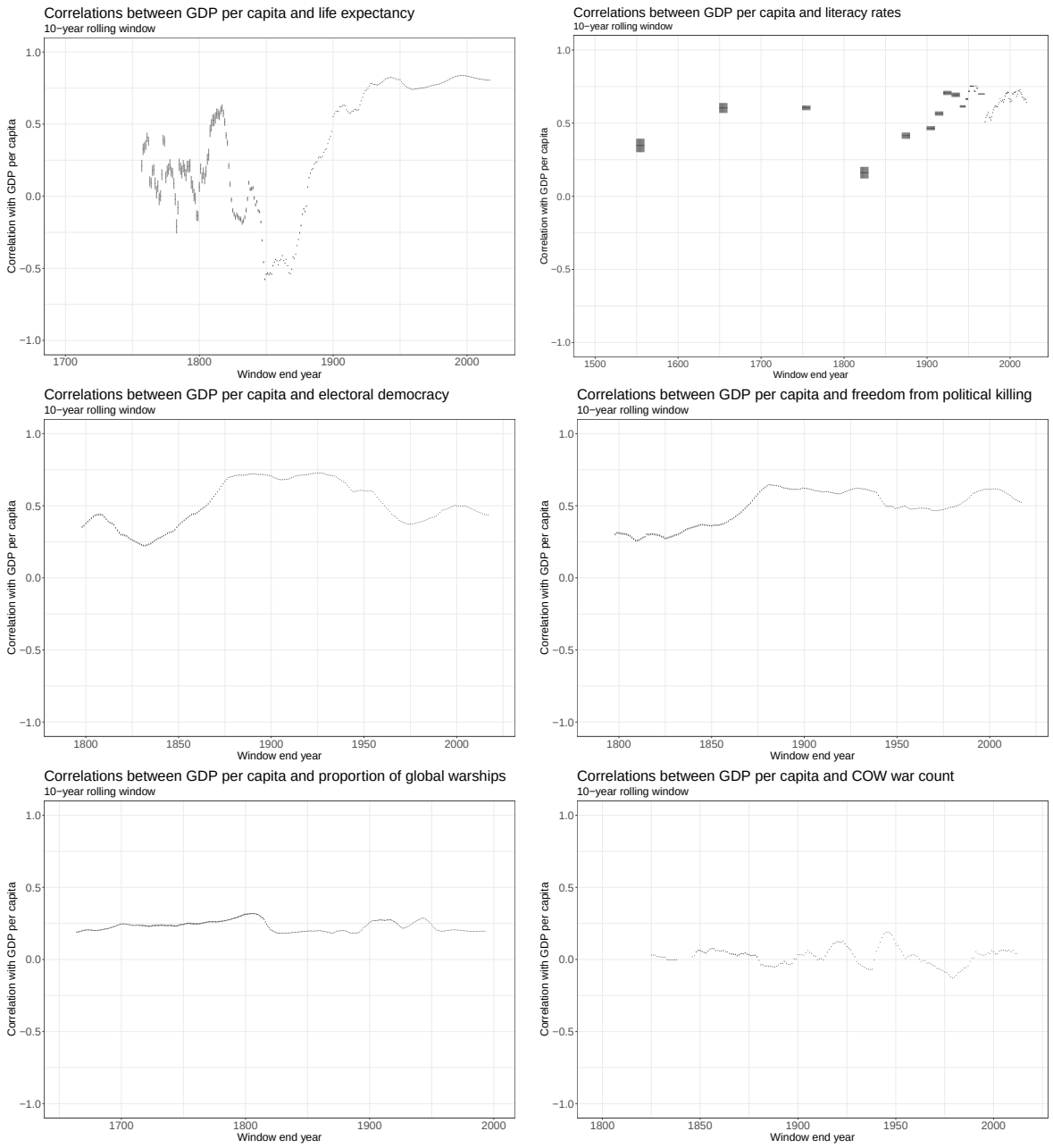

Figure 8: Distribution of Spearman rank-order correlation between the latent GDP per capita variable and other measures. Correlations are calculated for each 10-year period (1-year period and 50-year period graphs are available in online appendix D). Note that we only estimate a correlation coefficient for year-periods with at least 10 cases. The six variables include life expectancy (Max Roser and Ritchie, 2013), literacy rates (Roser and Ortiz-Ospina, 2016), a latent level of electoral democracy based on the polyarchy variable from VDEM (Coppedge et al., 2019), VDEM's freedom from political killing index (Coppedge et al., 2019), a measure of measure of naval arming (Modelski and Thompson, 1988), and the annual count of interstate wars that a country is involved in from the COW War data v4.0 (Sarkees and Wayman, 2010).

Incorporating measurement uncertainty into analysis

As we discussed in detail above, our latent variable model estimates the level of agreement between indicators for all observed values of the component indicators, which produces country-year estimates of the relative level of uncertainty for each variable. This uncertainty, or measurement error, is a well known problem in standard regression specifications. The new data provided in this article offer quantified estimates of the relative agreement between the component indicators of each variable. This measured uncertainty can easily be incorporated into regression models following guidance described in Bolck, Croon and Hagenaars (2004), Mislevy (1991), Schnakenberg and Fariss (2014), or Fariss (2019).

If models do not incorporate uncertainty, then the researcher cannot rule out the possibility that the statistical associations of GDP, GDP per capita, or population with some other variable are not a false-negative result (Type 2 error or attenuation bias), or the related possibility that the relationship of these variables are a false-positive result (Type 1 error). In models with multiple indicators (i.e., multiple-variable regression), bias due to measurement error is not always attenuating if un-modeled, higher-order interactions exist. Incorporating measurement uncertainty of variables addresses these difficult-to-model issues and can be further explored with non-parametric regression techniques. In sum, incorporating uncertainty into a regression model provides evidence that effect sizes are probabilistically distinguishable from zero even when we cannot measure a right-hand side variable perfectly (an assumption of standard regression models).

To illustrate the magnitude of the effect that measurement uncertainty has on estimated effect sizes, we compare regression models in which uncertainty in predictors is incorporated with specifications in which uncertainty estimates are omitted. We study the statistical association of two variables — GDP per capita and democracy — on the relative level of political killings (Coppedge et al., 2019), repression (Fariss, 2014, 2019), and, in online appendix E, battle deaths (Lacina and Gleditsch, 2005; Pettersson and Öberg, 2020). To incorporate measurement uncertainty in the regression models, we take $m = 1000$ random draws from the posterior distribution of latent variables, estimate $m = 1000$ regression models, and combine the results.[21]

In Table 5 we measure repression using draws from the distribution of VDEM's latent freedom

[21]See online appendix E for additional details about this approach and appendix F for other approaches to missingness.

from political killings scores (Coppedge et al., 2019); Table 6 leverages draws from the distribution of latent human rights protection scores from Fariss (2014, 2019).[22] We use our new extended series of GDP per capita in 2017 PPP dollars based on data from the World Bank (N.d.) because this indicator is estimated with the most precision among the GDP per capita components of our latent variable model that offer broad cross-sectional coverage in the original data (see Figure 7).[23] The baseline model that omits uncertainty in the measurement of democracy and GDP per capita overestimates the effect size of GDP per capita by approximately 26% to 32% depending on the model specification. The comparison with models in which we alternate the measurement of democracy and GDP per capita between point estimates and draws from the underlying variable distribution demonstrates that the bias in the estimates is driven predominantly by measurement uncertainty in the GDP per capita estimates.

Dependent variable: Draws from freedom from political killings scores (VDEM, Coppedge et al. 2019)

	Point Democracy, Point GDPPC	Draw Democracy, Point GDPPC	Point Democracy, Draw GDPPC	Draw Democracy, Draw GDPPC
GDPPC	0.082 (0.01)	0.085 (0.01)	0.08 (0.01)	0.083 (0.01)
		+ **3.49**	**-2.81**	+ **0.59**
Democracy	0.99 (0.048)	0.953 (0.047)	0.994 (0.048)	0.957 (0.047)
		-3.71	+ **0.40**	**-3.32**
LDV	0.727 (0.008)	0.731 (0.008)	0.728 (0.008)	0.731 (0.008)
		+ **0.45**	+ **0.05**	+ **0.51**
Intercept	-0.363 (0.018)	-0.355 (0.018)	-0.361 (0.018)	-0.353 (0.018)
		-2.23	**-0.48**	**-2.73**

Table 5: Comparison of coefficients between point estimates and draws (m = 1000) of democracy and GDP per capita variables 1816–2017. Ordinary Least Squares regression coefficients with standard errors in parentheses. Percentage change in coefficient size compared to baseline model (point estimates for both democracy and GDP per capita) in bold.

[22]The most recent version of these data are based on an updated latent variable model presented in Fariss, Kenwick and Reuning (2020).

[23]Data on electoral democracy (*v2x_polyarchy*) come from VDEM (Coppedge et al., 2019).

Dependent variable: Draws from latent human rights protection scores (Fariss, 2019)

	Point Democracy, Point GDPPC	Draw Democracy, Point GDPPC	Point Democracy, Draw GDPPC	Draw Democracy, Draw GDPPC
GDPPC	0.045 (0.008)	0.045 (0.008)	0.044 (0.008)	0.044 (0.008)
		+ 1.73	**-2.32**	**-0.62**
Democracy	0.288 (0.033)	0.279 (0.032)	0.289 (0.033)	0.28 (0.032)
		-3.08	**+ 0.54**	**-2.55**
LDV	0.891 (0.008)	0.891 (0.008)	0.891 (0.008)	0.892 (0.008)
		+ 0.07	**+ 0.03**	**+ 0.10**
Intercept	-0.174 (0.018)	-0.171 (0.018)	-0.173 (0.018)	-0.17 (0.018)
		-1.37	**-0.62**	**-2.01**

Table 6: Comparison of coefficients between point estimates and draws (m = 1000) of democracy and GDP per capita variables 1946–2017. Ordinary Least Squares regression coefficients with standard errors in parentheses. Percentage change in coefficient size compared to baseline model (point estimates for both democracy and GDP per capita) in bold.

Measurement bias

We demonstrated above that not accounting for measurement error can lead to unaccounted for attenuation bias in regression coefficients. We can also directly account for other sources of bias in the measurement model itself. Our measurement model provides a framework for addressing bias in the measurement of GDP, GDP per capita, and population in future research. As new information about the reporting practices of economic and population datasets emerges, the measurement model developed in this article can be adjusted, following recommendations by Fariss (2014, 2019).[24] As an example, future adaptations could incorporate information as to whether GDP, GDP per capita, and population figures are interpolated or extrapolated from observed data points to explicitly model measurement bias.[25] Once adjusted, the new estimates can be incorporated into standard regression models as described above.

[24]See Imai, Lo and Olmsted (2017), Kenwick (2020) and Reuning, Kenwick and Fariss (2019) for potentially fruitful modeling suggestions.

[25]For source data on GDP and GDP per capita, dropping interpolated or extrapolated data would result in a reduction of approximately 93% of the data from Bolt et al. (2018) and 87% of the data from Feenstra, Inklaar and Timmer (2015). We therefore keep all observations on GDP and GDP per capita in the current iteration of the project.

Conclusion

In this article, we identified and discussed the limitations of existing GDP, GDP per capita, and population variables. We then introduced and described a new latent variable model that estimates posterior predictions for these important variables. These new estimates provide several primary advantages over existing variables: (1) they extend the temporal and spatial coverage of existing data, (2) they include multiple manifest indicators of each concept and quantify the uncertainty for each of the resulting country-year estimates, and (3) the modeling framework itself can be expanded to account for new data or knowledge about systemic error in existing data. With the expanded temporal coverage of our estimates, we showcase several new empirical insights about the relationship between economic development and democracy, repression, conflict, power projection capabilities, literacy rates, and life expectancy, respectively. We also demonstrate how to incorporate uncertainty in commonly used observational models. These new results show that the relationship between repression and development is weaker than models that do not incorporate uncertainty suggest. Future extensions of the latent variable model can address other forms of systematic measurement error with new data, new theory, or both. To encourage the extension of our model, we make publicly available the data and code used to construct it. The code in our replication archive has been carefully annotated to encourage easy replication, modification, and — we hope — theoretically informed extensions.

The primary feature of our measurement model is that it provides a principled means to incorporate information for comparing measures of the economy from different datasets (across time, across country, or both). To inform our estimates, the measurement model only includes the dimension of information that is comparable. For example, using dataset values that are designed for comparison within a country over time, we include a country fixed effect (γ_c). When we produce the estimates of the country-year distributions, these additional parameters are also included. To be clear, we only include the relevant dimension of comparison from each of the datasets. The distributions of plausible estimates for each country-year unit are converted back into the relevant unit of measure for each of these datasets. Future users who are interested in within-country change should use one of the PPP estimates that are designed for this dimension of comparison (e.g., MDP 2018 RGDPpc^{NA} in 2011 $US). Users interested in making between-country comparisons should

use the estimates generated for the MDP 2018 CGDPpc in 2011 $US variable. This is because the model includes parameters that preserve the built in methodologies used to generate the original dataset.[26] All this to say, we take seriously the data production process of each variable included in our latent variable model, which should help guide scholars to the appropriate estimates generated from our model and alleviate concerns about combining different variables together in a unified latent variable framework.

To close, we wish to reiterate that the uncertainty estimates associated with our data should always be incorporated into any statistical analysis and that, though these uncertainty estimate are useful, they do not eliminate concerns about systematic measurement error that may arise because of potential bias in the information sources used to estimates these data (such as lying or low state capacity). Moving forward, we plan to update our new data each year and incorporate new information as it becomes available. We will maintain both a full historic data set (1500-present) and a smaller dataset from 1800-present. The goal in maintaining two datasets is to keep users informed about the relative level of data coverage in the early historic period relative to the historic data coverage available over the last 200 years. The smaller dataset will work well with other data sets like the Varieties of Democracy and Correlates, which also begins in 1790. Our new estimates are already helping scholars expand the scope of their inquiries (e.g. Anders, Fariss and Markowitz, 2020; Gerring and Veenendaal, 2020; Markowitz et al., 2020). We are excited to introduce these new estimates to the broader community of scholars and to see what new studies discover.

[26]We want to reiterate that we include a parameter for some of the data (β_c) that is used to convert between constant and PPP series. For data series that are comparable over time but not countries, we include country-random effects (γ_c). For data series that are comparable across countries but not over time, we include year-random effects (λ_t). These parameters ensure that the variation from a given dataset only informs the estimation of the latent parameter in the relevant dimension (i.e., time or space).

References

Anders, Therese, Christopher J. Fariss and Jonathan N. Markowitz. 2020. "Bread before guns or butter: Introducing Surplus Domestic Product (SDP)." *International Studies Quarterly* 64(2):392–405.

Atkinson, Tony. 2005. *Atkinson Review: Final Report.* Houndsmills, UK and New York: Palgrave MacMillanillan.

Bairoch, Paul. 1976. "Europe's gross national product, 1800-1975." *Journal of European Economic History* 5(2):273–340.

Banerjee, Abhijit. 1997. "A Theory of Misgovernance." *Quarterly Journal of Economics* 112(3):1289–1332.

Boehmer, Charles, Bernadette Jungblut and Richard Stoll. 2011. "Tradeoffs in Trade Data: Do Our Assumptions Affect Our Results?" *Conflict Management and Peace Science* 28(2):145–167.

Boehmer, Charles R. and David Sobek. 2005. "Violent Adolescence: State Development and the Propensity for Militarized Interstate Conflict." *Journal of Peace Research* 42(1):5–26.

Bolck, Annabel, Marcel Croon and Jacques Hagenaars. 2004. "Estimating Latent Structure Models with Categorical Variables: One-Step Versus Three-Step Estimators." *Political Analysis* 12(1):3–27.

Bolt, Jutta and Jan Luiten van Zanden. 2020. "Maddison style estimates of the evolution of the world economy. A new 2020 update." *Maddison-Project Working Paper WP-15, University of Groningen, Groningen, The Netherlands* .

Bolt, Jutta, Robert Inklaar, Herman de Jong and Jan Luiten van Zanden. 2018. Rebasing 'Maddison': new income comparisons and the shape of long-run economic development. GGDC Research Memorandum 174 University of Groningen. Groningen growth and development centre.

Broadberry, Stephen. 2015. "Accounting for the Great Divergence." Online. https://www.nuffield.ox.ac.uk/users/Broadberry/AccountingGreatDivergence6.pdf, accessed 23 November 2016.

Broadberry, Stephen and Alexander Klein. 2012. "Aggregate and per capita GDP in Europe, 1870–2000: Continental, regional and national data with changing boundaries." *Scandinavian Economic History Review* 60(1):79–107.

Burnside, Craig and David Dollar. 2000. "Aid, Policies, and Growth." *The American Economic Review* 90(4):847–868.

Büthe, Tim, Solomon Major and André de Mello e Souza. 2012. "The Politics of Private Foreign Aid: Humanitarian Principles, Economic Development Objectives, and Organizational Interests in NGO Private Aid Allocation." *International Organization* 66(4):571–607.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Anna Lührmann, Kyle L. Marquardt, Kelly McMann, Pamela Paxton, Daniel Pemstein, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Steven Wilson, Agnes Cornell, Lisa Gastaldi, Haakon Gjerløw, Nina Ilchenko, Joshua Krusell, Laura Maxwell, Valeriya Mechkova, Juraj Medzihorsky,

Josefine Pernes, Johannes von Römer, Natalia Stepanova, Aksel Sundström, Eitan Tzelgov, Yi ting Wang, Tore Wig and Daniel Ziblatt. 2019. "V-Dem [Country-Year/Country-Date] Dataset v9" Varieties of Democracy (V-Dem) Project." https://doi.org/10.23696/vdemcy19.

Data Quality Institute, World Economics. 2016. "Which Country's GDP data you can trust? (note: most you cannot)." Online. http://www.worldeconomics.com/pages/Data-Quality-Index.aspx, accessed 29 December 2016.

Deaton, Angus and Alan Heston. 2010. "Understanding PPPs and PPP-based National Accounts." *American Economic Journal: Macroeconomics* 2(4):1–35.

Deng, Kent G. 2004. "Unveiling China's True Population Statistics for the Pre-Modern Era with Official Census Data." *Population Review* 43(2):32–69.

Discoll, Jesse and Nicholai Lidow. 2014. "Representative Surveys in Insecure Environments: A Case Study of Mogadishu, Somalia." *Journal of Survey Statistics and Methodology* 2(1):78–95.

Eck, Kristine and Christopher J. Fariss. 2018. "Ill Treatment and Torture in Sweden: A Critique of Cross-Case Comparisons." *Human Rights Quarterly* 40(3):591–604.

Economist, The. 2014. "Step change; Nigeria's GDP." *The Economist* 411(8882):71.

European Communities, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations and World Bank. 2009. System of National Accounts, 2008. Technical report.
URL: *http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf*

Fariss, Christopher J. 2014. "Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability in Human Rights Documents." *American Political Science Review* 108(2):297–318.

Fariss, Christopher J. 2018*a*. "Are Things Really Getting Better?: How To Validate Latent Variable Models of Human Rights." *British Journal of Political Science* 48(1):275–282.

Fariss, Christopher J. 2018*b*. "Human Rights Treaty Compliance and the Changing Standard of Accountability." *British Journal of Political Science* 48(1):239–272.

Fariss, Christopher J. 2019. "Yes, Human Rights Practices Are Improving Over Time." *American Political Science Review* 113(3):868–881.

Fariss, Christopher J. and Geoff Dancy. 2017. "Measuring the Impact of Human Rights: Conceptual and Methodological Debates." *Annual of Law and Social Science* 13:273–294.

Fariss, Christopher J., Michael R. Kenwick and Keving Reuning. 2020. "Estimating one-sided-killings from a Robust Measurement Model of Human Rights." *Journal of Peace Research* 57(6):801–814.

Fariss, Christopher J, Therese Anders Charles D Crabtree, Zachary M Jones, Fridolin J Linder and Jonathan N Markowitz. 2017. "Latent Estimation of GDP, GDP per capita, and Population from Historic and Contemporary Sources." *arxiv* arXiv:1706.01099.

Fazal, Tanisha M. 2014. "Dead Wrong?: Battle Deaths, Military Medicine, and Exaggerated Reports of War's Demise." *International Security* 39(1):95–125.

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105(10):3150–82.

Gartzke, Erik. 2012. "Could climate change precipitate peace?" *Journal of Peace Research* 49(1):177–192.

Gelman, Andrew and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge, MA: Cambridge University Press.

Gerring, John and Wouter Veenendaal. 2020. *Population and Politics: The Impact of Scale*. Cambridge: Cambridge University Press.

Gleditsch, Kristian S. and Michael D. Ward. 1999. "A revised list of independent states since the congress of Vienna." *International Interactions* 25(4):393–413.

Gleditsch, Kristian Skrede. 2002. "Expanded Trade and GDP Data." *Journal of Conflict Resolution* 46(5):712–724.

Graham, Benjamin A.T. and Jacob R. Tucker. 2019. "The International Political Economy Data Resource." *Review of International Organizations* 14(1):149–161.

Hegre, Havard. 2000. "Development and the Liberal Peace: What Does it Take to be a Trading State?" *Journal of Peace Research* 37(1):5–30.

Hutchinson, Diane and Florian Ploeckl. 2018. "What Was the Australian GDP or CPI Then?" MeasuringWorth. Online. http://www.measuringworth.com/australiadata/, accessed 13 July 2018.

Imai, Kosuke, James Lo and Jonathan Olmsted. 2017. "Fast Estimation of Ideal Points with Massive Data." *American Political Science Review* 110(4):631–656.

International Labour Office. 2013. "Women and men in the informal economy: A statistical picture." Online. http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/publication/wcms_234413.pdf, accessed 12 December 2016.

Jerven, Morten. 2010*a*. "African growth recurring: An economic history perspective on African growth episodes, 1690–2010." *Economic history of developing regions* 25(2):127–154.

Jerven, Morten. 2010*b*. "Random growth in Africa? Lessons from an evaluation of the growth evidence on Botswana, Kenya, Tanzania and Zambia, 1965–1995." *The Journal of Development Studies* 46(2):274–294.

Jerven, Morten. 2010*c*. "The relativity of poverty and income: How reliable are African economic statistics?" *African Affairs* 109(434):77–96.

Jerven, Morten. 2013*a*. "For Richer, for Poorer: GDP Revisions and Africs's Statistical Tragedy." *African Affairs* 112(446):138–147.

Jerven, Morten. 2013*b*. *Poor numbers: how we are misled by African development statistics and what to do about it*. Ithaca: Cornell University Press.

Jerven, Morten. 2014. "A West African experiment: constructing a GDP series for colonial Ghana, 1891–1950." *The Economic History Review* 67(4):964–992.

Johnston, Louis and Samuel H. Williamson. 2018. "What Was the U.S. GDP Then?" MeasuringWorth. Online. http://www.measuringworth.org/usgdp/, accessed 13 July 2018.

Kenwick, Michael R. 2020. "Self-Reinforcing Civilian Control: A Measurement-Based Analysis of Civil-Military Relations." *International Studies Quarterly* 64(1):71–84.

King, Gary, James Honaker, Anne Joseph and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95(1):49–69.

Lacina, Bethany and Nils Petter Gleditsch. 2005. "Monitoring Trends in Global Combat: A New Dataset of Battle Deaths." *European Journal of Population* 21(2-3):145–166.

Lacina, Bethany, Nils Petter Gleditsch and Bruce M. Russett. 2006. "The Declining Risk of Death in Battle." *International Studies Quarterly* 50(3):673–680.

Lake, David A. 2009. *Hierarchy In International Relations*. Ithaca, NY: Cornell University Press.

Lee, Melissa M. and Nan Zhang. 2017. "Legibility and the Informational Foundations of State Capacity." *Journal of Politics* 79(1):118–132.

Maddison, Angus. 2010. "Statistics on World Population, GDP and Per Capita GDP, 1-2008 AD." Online. http://www.ggdc.net/maddison/oriindex.htm, accessed 20 August 2016.

Markowitz, Jonathan and Christopher J. Fariss. 2018. "Power, proximity, and democracy: Geopolitical competition in the international system." *Journal of Peace Research* 55(1):78–93.

Markowitz, Jonathan N. and Christopher J. Fariss. 2013. "Going the Distance: The Price of Projecting Power." *International Interactions* 39(2):119–143.

Markowitz, Jonathan N., R. Blake McMahon and Christopher J. Fariss. 2019. "Producing Goods and Projecting Power: How What States Make Influences What They Take." *Journal of Conflict Resolution* 63(6):1368–1402.

Markowitz, Jonathan N., Suzie Mulesky, Benjamin A.T. Graham and Christopher J. Fariss. 2020. "Productive Pacifists: The Rise of Production-Oriented States and Decline of Territorial Conquest." *International Studies Quarterly* 64(3):558–572.

Max Roser, Esteban Ortiz-Ospina and Hannah Ritchie. 2013. "Life Expectancy." *Our World in Data* . https://ourworldindata.org/life-expectancy.

Measuring Worth. 2019. Online. https://www.measuringworth.com/glossary/index.php, accessed 1 February 2019.

Mislevy, Robert. 1991. "Randomization-based inference about latent variables from complex samples." *Psychometrika* 56(2):177–196.

Modelski, George and William R. Thompson. 1988. *Sea Power in Global Politics, 1494-1993*. Seattle, WA: University of Washington Press.

Neumayer, Eric. 2003. "The Determinants of Aid Allocation by Regional Multilateral Development Banks and United Nations Agencies." *International Studies Quarterly* 47(1):101–122.

Pettersson, Therese and Magnus Öberg. 2020. "Organized violence, 1989-2019." *Journal of Peace Research* 57(4):597–613.

Pinkovskiy, Maxim and Xavier Sala-i-Martin. 2016. Newer Need Not be Better: Evaluating the Penn World Tables and the World Development Indicators Using Nighttime Lights. Working Paper 22216 National Bureau of Economic Research.

Prados-de-la-Escosura, Leandro. 2018. "What Was Spain's GDP Then?" MeasuringWorth. Online. http://www.measuringworth.org/spaingdp/, accessed 13 July 2018.

Prewitt, Kenneth. 2010. "The U.S. Decennial Census: Politics and Political Science." *Annual Review of Political Science* 13:237–254.

Reuning, Kevin, Michael R. Kenwick and Christopher J. Fariss. 2019. "Exploring the Dynamics of Latent Variable Models." *Political Analysis* 27(4):503–517.

Rogoff, Kenneth. 1996. "The Purchasing Power Parity Puzzle." *Journal of Economic Literature* 34(2):647–668.

Roser, Max and Esteban Ortiz-Ospina. 2016. "Literacy." *Our World in Data* . https://ourworldindata.org/literacy.

Sarkees, Meredith Reid and Frank Wayman. 2010. *Resort to War: 1816 - 2007*. Washington DC: CQ Press.

Schnakenberg, Keith E. and Christopher J. Fariss. 2014. "Dynamic Patterns of Human Rights Practices." *Political Science Research and Methods* 2(1):1–31.

Schneider, Friedrich and Dominik Enste. 2000. Shadow Economies around the World Size, Causes, and Consequences. International Monetary Foundation. IMF Working Paper.

Scott, James C. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven: Yale University Press.

Seawright, Jason. 2016. "The Case for Selecting Cases That Are Deviant or Extreme on the Independent Variable." *Sociological Methods & Research* 45(3):493–525.

Singer, David J., Stuart Bremer and John Stuckey. 1972. Capability Distribution, Uncertainty, and Major Power War, 1820-1965. In *Peace, War, and Numbers*, ed. Bruce Russett. Beverly Hills: Sage pp. 19–48.

Slough, Tara and Christopher J. Fariss. 2020. "Misgovernance and Human Rights: Experimental Evidence of Illegal Detention without Intent." *American Journal of Political Science* .

Souva, Mark and Brandon Prins. 2006. "The Liberal Peace Revisited: The Role of Democracy, Dependence, and Development in Militarized Interstate Dispute Initiation, 1950–1999." *International Interactions* 32(2):183–200.

Stan Development Team. 2021. "Stan Modeling Language Users Guide and Reference Manual, 2.25. https://mc-stan.org.".

Stiglitz, Joseph, Amartya K. Sen and Jean-Paul Fitoussi. 2009. The measurement of economic performance and social progress revisited: Reflections and Overview. Working Paper 33 OFCE. Centre de recherche en économie de Sciences Po Paris Cedex: .

Thomas, Ryland and Samuel H. Williamson. 2018. "What Was the U.K. GDP Then?" MeasuringWorth. Online. http://www.measuringworth.com/ukgdp/, accessed 13 July 2018.

Treisman, Daniel. 2020. "Economic Development and Democracy: Predispositions and Triggers." *Annual Review of Political Science* 23(1):null.

United Nations. 1953. A System of National Accounts and Supporting Tables. Studies in Methods 2 Department of Economic Affairs Statistical Office New York: .

UNU-IHDP and UNEP. 2014. *Inclusive Wealth Report 2014. Measuring progress toward sustainability.* Cambridge: Cambridge University Press.

Wallace, Jeremy L. 2016. "Juking the stats? Authoritarian information problems in China." *British Journal of Political Science* 46(01):11–29.

World Bank. N.d.