# Bread before guns or butter:
# Introducing Surplus Domestic Product (SDP)

Therese Anders
University of Southern California

Christopher J. Fariss
University of Michigan

Jonathan N. Markowitz
University of Southern California

**Abstract**

Scholars systematically mismeasure power-resources and military burdens by using GDP (Gross Domestic Product) as a proxy for the income states can devote to arming. The core problem is that GDP confounds two conceptually distinct forms of income into one additive indicator. Subsistence income represents resources needed to provide the "bread" necessary to cover the basic subsistence needs of the population. Surplus income represents the remaining resources that could be allocated to "guns" or "butter." Our new measure of SDP (Surplus Domestic Product) corrects for this measurement error by decomposing subsistence income and surplus income from total GDP. Validation exercises demonstrate that SDP outperforms GDP at measuring the distribution of power-resources. Though theoretically, we expect states' decisions to arm is influenced by the distribution of power, empirical models using GDP find mixed support for this expectation. Strikingly, using SDP reveals strong support for this proposition.

*The physical product of hundreds of millions of peasants may dwarf that of five million factory workers, but since most of it is immediately consumed, it is far less likely to lead to surplus wealth or decisive military striking power.*

*—Paul Kennedy, The Rise and Fall of Great Powers.*

International relations scholarship systematically mis-measures both power-resources and military burdens because the operationalization of these variables depends on gross domestic product (GDP) as an indicator of the income states can devote to arming and projecting power. The core problem is that GDP confounds two conceptually distinct forms of income into one aggregate indicator. Subsistence resources are the income necessary to provide the minimal amount of "bread" that the population needs to survive. Surplus resources are the remaining economic income that can be invested in "guns" or "butter" (Garfinkel and Syropoulos 2019). As a consequence of conflating these two types of income into a single indicator, existing inferences about the distribution of power-resources, the capacity of states to build power projection capabilities, and the costs of arming are biased. In particular, the misuse of GDP as a proxy measure of power-resources systematically overestimates the power-resources of low-income states with large populations and underestimates the rate at which these resources increase when low-income countries begin to experience economic growth. We address this bias by developing a new measure called Surplus Domestic Product (SDP). SDP is created by decomposing total GDP into surplus income and subsistence income.

This paper yields three contributions for scholars of international relations, and political science more broadly. First, we introduce surplus income (SDP) and subsistence income, which are both part of total income (measured in GDP). We demonstrate that once we account for both types of income produced by a state, SDP is a more conceptually appropriate measure of the power resources available for a state to arm and project military force abroad. For illustration, scholars mis-measure military burdens by using military expenditures as share of GDP. When SDP is used in the denominator instead, we demonstrate that, for most of history, nearly all countries endured much higher military burdens than previously realized. Failing to take into account that only surplus resources can sustainably be invested in the military concealed just how rapid and steep the fall in military burdens has been after the Cold War. This drop is especially pronounced for developing states around the world, particularly in Asia. Our results suggest that scholars have underestimated the size of the peace dividend associated with the end of the Cold War and the gains from entering into hierarchical security relationships (e.g. Lake 2009). Our recommendation of using SDP while also accounting for subsistence income as a separate variable—instead of total GDP—is potentially critical to scholars who currently use GDP as a proxy for the capacity of states to devote resources to non-military purposes, for instance, in work evaluating states' decisions to invest in education, healthcare, or other quasi-public goods (e.g. Lake and Baum 2001; Tanzi and Schuknecht 2000).

Second, in the process of building comprehensive data on surplus and subsistence income we provide new data for some of the most widely used variables in political science and economics, with coverage from 1800-2018.[1] In particular, we improve existing cross-sectional and temporal data coverage for both GDP and population (the components of our new SDP indicator). Currently, cross-national data on military expenditure as a percentage of GDP go back to 1950 (Nordhaus, Oneal, and Russett 2012). Our new data allow us to extend the data coverage for military expenditure as a percentage of economic resources (SDP or GDP) back to 1800 for most of the

[1] See Coyle (2014) for a discussion of the history of the statistic known as GDP.

world's states.

Third, we use SDP to improve measurement of relative power-resources between states, allowing us to provide a more accurate identification of the most powerful and potentially threatening states in the world each year. Based on existing scholarship, which relies on GDP as a proxy for power-resources, China is ranked as the world's most powerful country in the early and mid-19$^{th}$ century. However, until the 1990s, nearly all of China's income was used to sustain its large, impoverished population and little surplus remained to invest in guns or butter. Using SDP to measure relative power-resources leads to a markedly different set of countries topping the rankings of powerful states. In the early 19$^{th}$ century, our estimates place the United Kingdom rather than China in the top spot, which comports much better with actual arming and power-projection behavior of countries during this time period. Decomposing GDP into surplus and subsistence income thus provides new insights for scholars working on a broad range of topics related to the distribution of power such as arming, alliances, power transition, peace, and great power politics.

While financial resources are not the only dimension of a state's power-resources, they represent a particularly important, if not the most critical, dimension of those resources (Beckley Forthcoming; Norloff and Wohlforth Forthcoming; Cappella-Zielinski 2016). The importance of financial resources has led prior scholarship to use GDP as the primary measure of a state's power-resources. Our claim is not that SDP measures every dimension of states' power-resources, or that other dimensions of power should not be considered, but only that SDP outperforms GDP as an approximation for the income a state can actually devote to pursuit of its various objectives, which may include not only arming, but also investment in industrialization. As we demonstrate in this paper, SDP is more closely related than GDP to alternative measures of power-resources, such as industrial capacity-related components of the Composite Index of National Capabilities (CINC). We also develop and examine solely population-based measures of power-resources. We show that GDP measures are more closely related to population measures than SDP. In fact, in the early to mid-19$^{th}$ century, variation in GDP is almost entirely explained by variation in population—a reflection of the Malthusian constraint faced by most countries at the time. Thus, SDP captures a concept of power-resources distinct from GDP or population.

Our new measure of SDP outperforms GDP in three validation exercises. First, SDP fares better than GDP when compared to alternative indicators that approximate states' relative power-resources.[2] Second, SDP more accurately ranks states with the greatest power-resources. We modify a recently developed operational strategy to incorporate SDP into measuring the level of potential threat in states' geopolitical environments (Markowitz and Fariss 2018). Specifically, we use SDP to measure relative power-resources between pairs of states and weight these ratios by distance and preference compatibility. Third, we show that weighted power-resource ratios between states based on SDP produce a more valid ranking of countries each state finds the most threatening, as compared to the same approach using GDP.

For our primary empirical application, we generate an aggregate country-year measure of the potential threat each state faces in its strategic environment. We illuminate how the potential threat states face influences the degree to which they invest scarce resources into arming and power projection capabilities. We demonstrate that when SDP is incorporated into this model, it outperforms the same model using GDP in predicting military investments. Strikingly, our model

---

[2] Following previous research by Beckley (2018), we define a state's power-resources as the pool of resources a state can potentially invest in generating influence.

reveals that key variables suggested by existing theories of arming, such as relative power, geographic proximity, and preference compatibility *cannot* explain the degree to which states arm when power-resources are measured using GDP. However, these same variables *can* explain the degree to which states arm, when power-resources are measured using SDP and military burdens are measured using military expenditures as share of SDP. It is only the misuse of GDP as a measure of available power-resources that makes existing theories of arming appear empirically unsupported. Our results are robust to using naval tonnage relative to SDP as an alternative measure of arms levels.

Decomposing GDP into surplus income and subsistence income allows us to examine patterns of SDP historically, which reveals that in the 19th and early 20th centuries, military burdens—the percentage of income devoted to arming—were, on average, higher than suggested by existing research. This difference exists because, until recently, most states were able to generate only small amounts of economic surplus. Those that did generate surplus, spent a large proportion of it on arming. As a result, many states in this earlier period had military burdens as high as 25%-50% of SDP. To put this in perspective, even during the Cold War, the United States spent only about 10% of its SDP on the military.

Finally, SDP reveals that military burdens have fallen faster than previously realized. Newly industrialized states like China are seeing large gains in economic surplus, which they could invest in arming. Yet, when measured as a percentage of SDP, the military burdens of these states came tumbling down over the last several decades. Today, most governments are prioritizing butter over guns and, as a result, military burdens as a percentage of SDP are far lower than in the past. However, for the least developed states, military burdens are still much higher than previously realized.

## Subsistence and Surplus

Scholars have long recognized the close relationship between a state's income and the resources available for military use. For example, Sandler and Hartley argue that "as GDP rises, a nation has both more resources to protect and greater means to provide protection" (Sandler and Hartley 1995, 60). However, scholars recognize that a state's ability to generate military power depends not only on the size and sustainability of its resource base, but also on the degree to which the state is constrained from extracting and mobilizing those resources (Sandler and Hartley 1995; Lamborn 1983; Milward 1977). Our argument and our measure of SDP build on these insights. While most previous research highlights how domestic political factors constrain the amount of resources a state can extract, we focus on how biological factors limit the amount of resources that are potentially available for extraction. Political constraints are important, but we cannot accurately estimate their impact without first creating a measure of the resources that could, given the political will, be extracted by the state.

We argue that when estimating the amount of resources that are potentially available for extraction and arming, scholars should account for a state's surplus and subsistence income separately, rather than adding them together. Surplus income (SDP) is calculated by removing from GDP the resources the population must consume to survive. Biology determines the number of calories required for survival and this lower caloric bound is largely stable across time and space. If citizens do not survive, they cannot use their labor to produce income.

It is possible for states to extract even subsistence income from their population, and states sometimes do so, particularly in times of crisis. However, this level of predation results in the population growing sicker and weaker, decreasing their ability to produce income and consequently reducing the resources available for the state to extract even in the very near future. Thus, the subsistence needs of the population place an upper-bound to the total amount of income states can sustainably extract. Any resources remaining after the subsistence needs of the population are met represent surplus income that can potentially be extracted by the government.

How do we estimate a population's subsistence needs? The World Bank monitors health and wellbeing associated with extreme poverty at several different thresholds. We argue that a $3-per-day threshold, which the World Bank calls "close to extreme poverty," is a conservative estimate of the subsistence resources required per capita.[3] This estimate is conservative because even at $3 per day, people tend to be chronically malnourished. As a result, they are more likely to succumb to disease and generate less surplus income. The World Bank estimates that in low-income countries (defined as countries having a per-capita gross national income of $1095 or just under $3 per day), 27% of the population were undernourished in 2015 (The World Bank 2017). Moving from a $3 to a $1.90-per-day threshold is associated with chronic malnourishment—causing approximately 10% to 40% of children under the age of five to be underweight (Ezzati 2004, pp. 1949 and 1985).

We use the $3-per-day threshold to calculate the SDP available for the state to extract and spend on public or private goods—in particular, arming and power projection capabilities—and the remaining subsistence income. While it is possible for states to extract subsistence income via taxation, pushing citizens below this threshold, we argue that the costs of doing so are very high and that this cannot be done sustainably. Taxing the population to below the subsistence threshold

[3] The thresholds are calculated in constant 2011 purchasing power parity dollars. The World Bank's poverty threshold has undergone adjustments over time—the threshold of $1.90 today is equivalent to $1.08 in 1993, and $1.00 in 1985 (Ezzati 2004). The statistic on malnourishment referenced here was originally calculated based on the $1.08 threshold and adjusted to reflect the World Bank's new poverty threshold of $1.90 (Ferreira 2015).

undermines economic productivity in even the very near future.

Of course, biological constraints are not the only limits states face in taxing their population (e.g. Lamborn 1983; Milward 1977). Most states historically have lacked either the political will or the capacity to extract this entire surplus. However, SDP estimates the upper bound on the resources a state *can* sustainably extract if it has the capacity and political will to do so.

To compute SDP for each state $i$ in year $t$, we first calculate the *minimum subsistence value* $v_{it}$ for each country-year, which is the level of income necessary to sustain the state's population, and then use this value to determine how much surplus income remains. We let $v_{it} = [(365 \times \tau) \times Population_{it}]$, where $\tau$ is the per-day, per-person subsistence threshold. Based on our discussion above, we use a subsistence threshold $\tau$ of \$3 per-day per-person. We also assess the sensitivity of our results to per-day subsistence thresholds at \$3, \$2, \$1, and \$0 (standard GDP). If $GDP_{it} > v_{it}$, then $SDP_{it} = GDP_{it} - v_{it}$ and $subsistence_{it} = v_{it}$. If $GDP_{it} \leq v_{it}$, then $SDP_{it} = 0$ and $subsistence_{it} = GDP_{it}$.

For a state to have surplus income, it must generate enough subsistence income to meet the needs of the population. If the state's income does not exceed this *minimum surplus value,* then it has an SDP of zero and only has subsistence income to work with. We have thus decomposed GDP into surplus income (SDP) and subsistence income (see Supplementary Appendix Section A for further details about this formalization). We account for both of these income values in the measurements and regression models we develop below.

***Converging and diverging trends in the international system: GDP vs. SDP***

Having decomposed GDP into surplus income (SDP) and subsistence income, we now ask: *is SDP a more valid measure of power-resources than GDP?* As a first concurrent validity test, we compare temporal trends in states' shares of global economic power for GDP versus SDP.[4] We demonstrate that SDP produces a better representation of historical trends and a more valid list of the global top ten powers than GDP for the past 200 years. Second, we demonstrate the benefit of using surplus as a measure of power-resources by showing that SDP correlates more strongly with alternative indicators of power-resources than GDP.

Figure 1 displays the historic evolution of states' shares of global power-resources for six countries, based both on SDP and GDP. When using SDP, wealthier states like the United States and the United Kingdom are estimated to have a much higher share of global power-resources than when using GDP.[5] The opposite is true for most poor states. For countries like China and South Korea, measuring power-resources via GDP overestimates their relative power-resources prior to 1980 and underestimates the rate of their rise since. The effect of economic development can be illustrated by the case of South Korea. Prior to its burst of economic growth, the county's share of global GDP was substantially higher than its share of global SDP. Beginning in 1985, this trend reversed and South Korea's share of global SDP started to become larger than its share of global GDP.

Figure 2 displays the top ten largest powers based on their average share of global SDP or

---

[4] Supplementary Appendix Section A presents formal descriptions of the validity tests.

[5] Figure 3 in the Supplementary Appendix provides scatterplots of SDP versus GDP of all countries for a select number of years. Figure 4 in the Supplementary Appendix illustrates the relationship between SDP and per capita GDP. These graphs demonstrate that, while economic development increases the correlation between SDP and per capita GDP, individual countries vary considerably regarding the strength of this relationship.

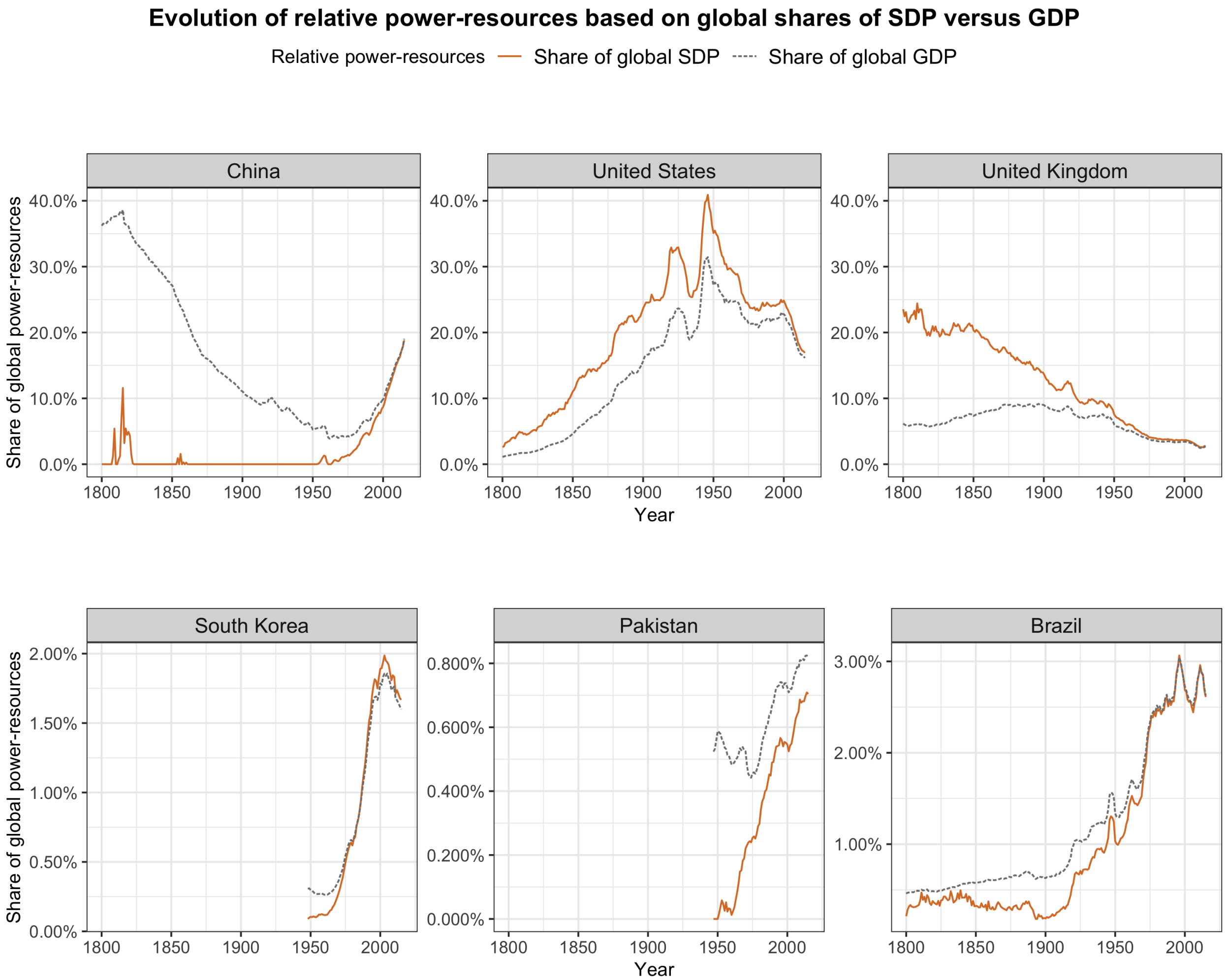



Figure 1: The plots illustrate the evolution of economic power-resources based on a state's share of global SDP versus GDP. SDP is computed based on a $3 per diem subsistence level. A state either has positive surplus or no surplus at all. If a state has no SDP then it would need to extract from the available subsistence resources ($3 x 365 x population), which, as we describe in the text, reduces the ability of individuals to produce economic surplus. The plot shows that when measuring economic income via SDP, developed states have a much higher share of relative power than previously believed. The opposite is true for developing states. For countries like contemporary China, but also Pakistan or South Korea in the past, measuring economic income via GDP leads to an overestimation of their relative power compared to other states. The effect of economic development of an individual country can best be observed in the case of South Korea. Prior to its burst in economic growth, the county's share of global GDP was mucg higher than its share of global SDP.

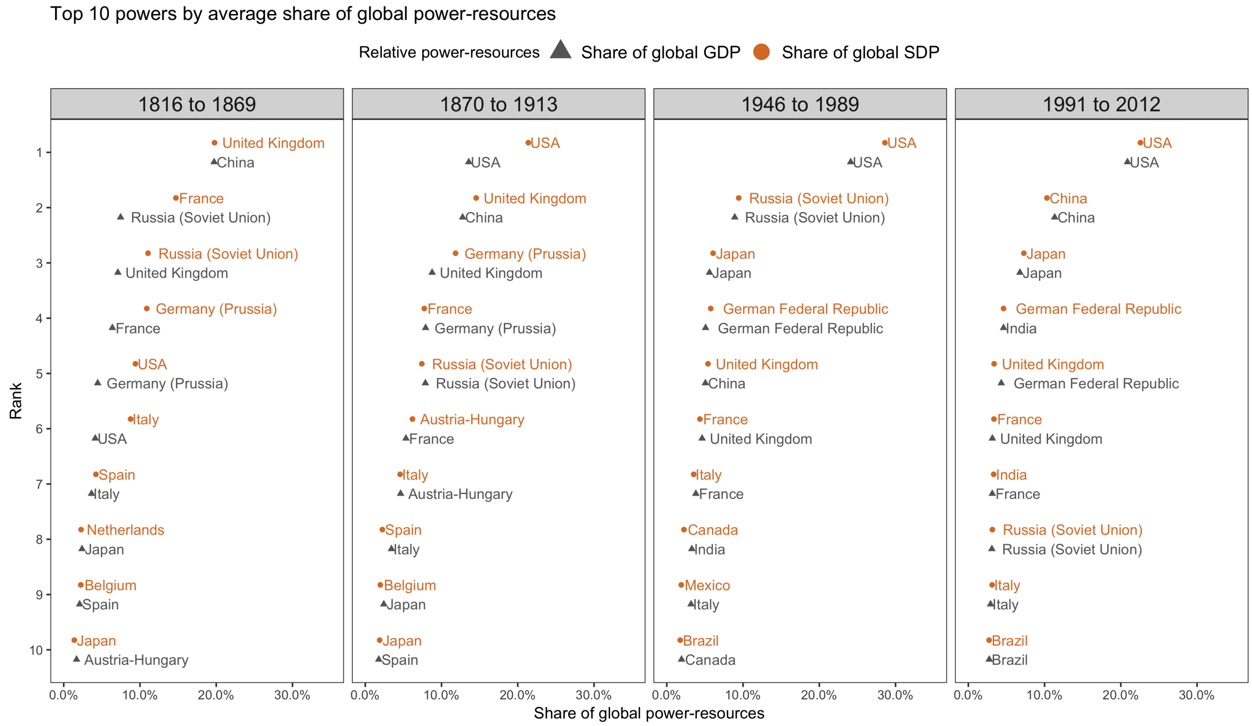



Figure 2: The plot displays the top 10 powers ranked by their average share of global SDP or GDP for four time periods. With the exception of China and a few other powers, the membership in the top 10 club is similar between the two measures of economic income. What does change is the rank-ordering of the countries. SDP produces a more historically valid ranking of the great powers than GDP.

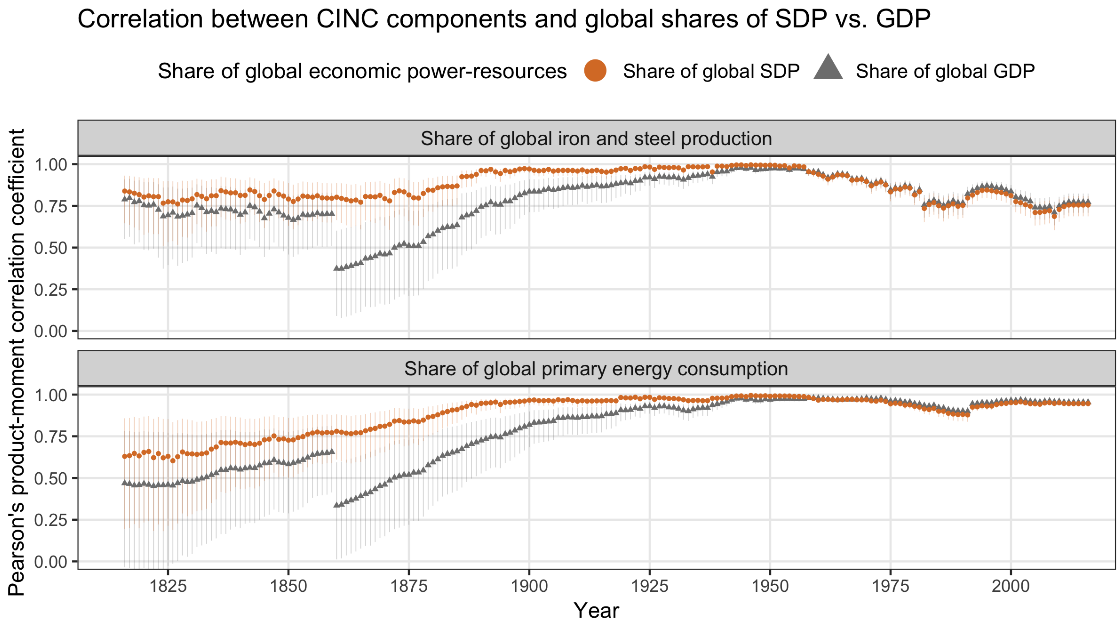
GDP for four time periods. The difference is most striking for China. In the 19th century, the economic income produced by hundreds of millions of peasant-farmers dwarfed the production of other states. Based on its share of global GDP, China would be considered the most powerful country in the period from 1816 to 1869. However, because most of this income was immediately consumed for subsistence needs, China had little surplus wealth to invest in arming or power projection. SDP takes this into account and therefore does not place China in the top ten powers in the early 19th century. These cases provide evidence that SDP produces a more valid ranking of great powers than GDP over several historic periods.

For a second set of validations, we assess convergent validity by comparing a country's share of global SDP to several component variables from CINC. While CINC has well-known drawbacks (Beckley Forthcoming; Kadera and Sorokin 2004),[6] its components are useful for comparison as an established, widely-used source of variables related to states' economic power. Because SDP and GDP are indicators of economic resources, we assess them in relation to four CINC variables that measure resources that could potentially be invested in military capabilities. The remaining CINC components, namely military expenditure and personnel, are related to actual military capabilities, which we use as a dependent variable to measure military investment below.

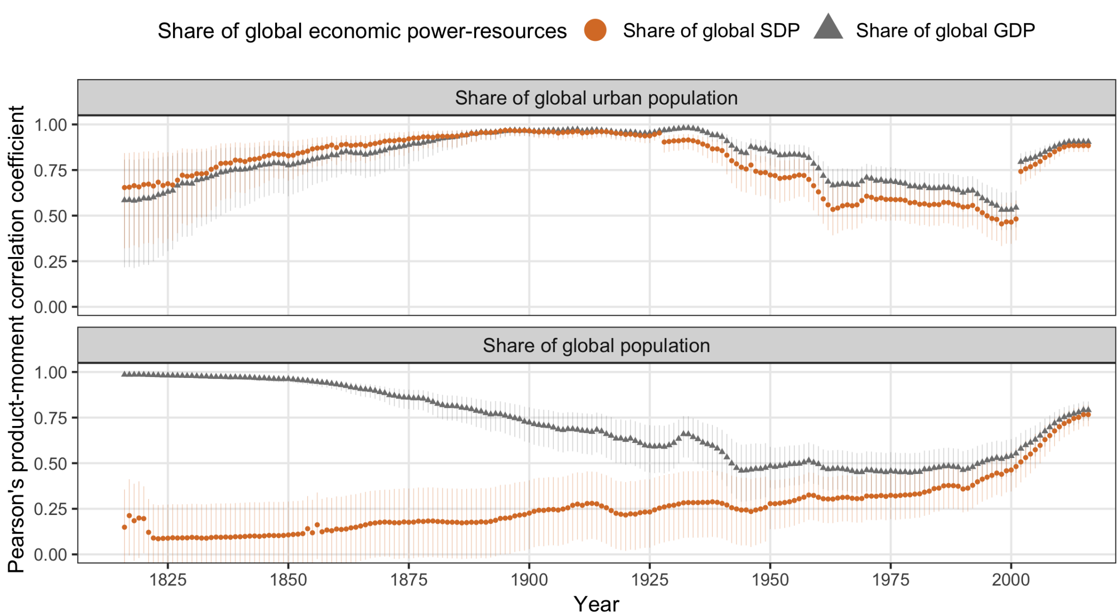
In the top panel of Figure 3, we display correlations between a country's share of global SDP or GDP and two components of CINC, which are yearly shares of global (1) iron and steel

---

[6] CINC's restrictive approach toward including countries as members of the international system leads to large disagreements in the estimates of power for some years (Supplementary Appendix Section C).




(a) CINC industrial capacity variables




(b) CINC population variables

Figure 3: The plots display yearly correlation coefficients with 95% confidence intervals. In each of the four panels, we assess the degree to which SDP (orange) and GDP (grey) correlate with each of four component variables of CINC. In the upper graph of panel (a) we look at a country's share of global iron or steel production (CINC). In the lower graph of panel (a), we assess correlations with a country's share of global primary energy consumption (CINC). In the upper graph of panel (b) we assess a country's share of global urban population (CINC). In the lower graph of panel (b), we look at total population (for more on the population variable, see the online appendix).

production and (2) primary energy consumption.[7] The year-by-year correlation coefficient with SDP ranges between 0.69 and 1 for iron and steel production and between 0.62 and 1 for primary energy consumption. Until WWII, a state's share of global SDP correlates more strongly with related measures of power-resources than a state's share of global GDP. Between 1860 and 1960, the 95% confidence intervals for the two series of correlation coefficients do not overlap—suggesting a statistically significant difference between GDP and SDP.

In the bottom panel of Figure 3, we present analogous graphs for the correlation of the yearly share of SDP or GDP with a country's share of global (3) urban and (4) total population.[8] The contrast with the industrial capacity variables is striking: GDP is more closely related to a country's share of global population than SDP. The discrepancy between the two series changes over time. In the early 19th century, population is perfectly predictive of GDP, while SDP is not at all correlated with a country's population. In pre-industrial years and extending into the WWII period for many states, national wealth is primarily a function of how many citizens a government has available for exploitation. The surplus resources most countries can extract from their population pre-1800 are nearly zero. Thus, the development of a force structure that goes beyond feeding soldiers and obtaining basic equipment is cost-prohibitive for most states. The Industrial Revolution changed this pattern because states began to produce economic wealth beyond the basic

---

[7] Note that the break in the series of the correlation between industrial capacity and GDP in 1860 is caused by China entering the CINC data set (with an approximate global population proportion of 47% according to CINC and 31% according to our computations in 1860).

[8] To alleviate concerns regarding the coverage of CINC's total population figures, we use our extended series of population estimates (see Supplementary Appendix Section C for more details).

subsistence needs of their populations. These resources were, in turn, invested into the equipment and technology necessary to project force over distance.

The historical patterns depicted in Figure 3 demonstrate that accounting for subsistence income is crucial for the study of global power relationships, particularly in historical international relations analyses. By shifting the conceptual focus from gross to surplus domestic product, we can more accurately identify which states possess the greatest power-resources and potential to generate the military capabilities to threaten other countries. States with low surplus economic resources are largely incapable of projecting power, even if their total GDP is quite large.

Next, we show how SDP, compared to GDP, represents a more valid measure of relative power-resources between pairs of states. This dyadic-level measure, along with additional information about the state-pair in question, allows us to assess the level of potential threat one state might expect from its interactions with other states.

## Measuring Relative Power-Resources and Potential Threat

To date, international relations scholarship had difficulty explaining patterns of arming. Even though theoretical expectations suggest that states arm in response to threats in the international system, empirical results are inconclusive (Nordhaus, Oneal, and Russett 2012; Cappella-Zielinski, Fordham, and Schilde 2017; Sandler and Hartley 1995, 46). Our argument suggests that these inconclusive results emerge because prior scholarship has mis-measured power-resources by relying on GDP. Our SDP-based approach addresses this systematic bias and reveals new patterns in line with theoretical expectations. State's effort to arm, i.e. their military expenditures, should be scaled by the potential surplus resources they have available to arm (i.e. their SDP). While previous scholars have scaled military expenditures by GDP, we are the first to argue that they should instead be scaled by SDP. Our findings demonstrate that a strong relationship between a state's threat environment and efforts to arm exists, but only when military expenditures are scaled by SDP.

Relative power-resources between two states is the key construct we use to understand patterns of arming and power projection. If military capabilities represent the latent power to hurt, then power-resources represent the latent power to arm, or a state's military potential (Kennedy 1987). To account for the fact that not all dyadic economic power relationships are created equal, we also consider how other dyad-level features, specifically shared preferences and the loss of strength gradient, mitigate the level of threat associated with differences in relative power-resources between states.

***Measuring the difference in relative power-resources between states using SDP***

For each country-year unit, $i = 1, \ldots, N$ indexes states and $t = 1, \ldots, T$ indexes years. For every country-year unit, we assess information for each dyadic relationship between state $i$ and all other states in the international system that year. Let $j = 1, \ldots, J$ index the other states in relationship with state $i$. For every state $i$, we measure $i$'s annual value of surplus economic resources, $SDP_{it}$, as well as the surplus power-resources of each opponent, $SDP_{jt}$. For each $ij$ pair, we then compute power-resource ratios $r_{ijt}$ as the proportion of the opponent state's SDP relative to the total SDP in each dyad, such that $r_{ijt} = \frac{SDP_{jt}}{SDP_{jt}+SDP_{it}}$.

The relative power-resource variable $r_{ijt}$ is bounded between 0 and 1. When $SDP_{jt}$ is large, the relative power-resources of $j$ compared to $i$ will be greater than .5 and represent a state that is potentially threatening to $i$ because $j$ has more resources to invest in arming and power projection. The least powerful state's power-resource ratio will be close to 1. The most powerful state's power-resource ratio will be close to 0. If two states have equal power-resources, they will each find the other equally threatening and the power-resource ratio is 0.5. These ratios allow us to compare the relative power-resources of one pair of states to the relative power-resources of another pair of states, thus relativizing these comparisons and facilitating further measurement aggregation.

***Relative power-resource relationships between one state and all other states***

The relative distribution of power-resources between two states is useful for dyadic-level analyses. However, we also want to understand how an individual state responds to the total level of potential threat it faces from all potential opponents in the international system. We define a country's level of potential threat as its expectations regarding other states' potential ability to harm it. All else equal, the more power-resources a state has, the more potentially threatening it is. States shift from potentially threatening to actually threatening when they engage in behavior that is perceived as harmful, such as arming, coercing, or attacking. Our goal in this paper is to measure the degree to which a given state is *potentially threatening* based on its relative power-resources, rather than *actually threatening* based on its actions.

To measure the level of potential threat, we follow an existing operational procedure to create an aggregated country-year measure using each of the dyadic power-resource ratios.[9] This measure is simply the sum of all power-resources ratios, weighted by the loss of strength gradient and preference compatibility, between each state $i$ and all opponent states $j$ in the international system. This generates a country-year measure that summarizes all dyadic relationships of each country in each year and represents the unique level of potential threat each state faces.

[9] We modify the formalization by Markowitz and Fariss (2018) to better capture the intuition that the global environment might be transitioning into an increasingly competitive space. See Supplementary Appendix Section D.

We call this new variable $potential\ threat_{it}$, which is the sum of relative power-resource ratios that relate state $i$ to every opponent state $j$ in year $t$.

$$potential\ threat_{it} = \sum_{j=1}^{J} \left(r_{ijt} \times p_{ijt} \times w_{ijt}\right)$$

$r_{ijt}$ denotes relative power-resources between two states. Following Markowitz and Fariss (2018), we also include preference compatibility between two states $p_{ijt}$ , and the loss of strength gradient between two states $w_{ijt}$.

The first weight $p_{ijt}$ is based on the preference compatibility between states. $p_{ijt} = 0$ if state $i$ and state $j$ have compatible preferences with each other in year $t$; otherwise, $p_{ijt} = 1$ when these states do *not* share compatible preferences. We operationalize joint democracy as our indicator of compatible preferences using the Polity IV data (Marshall, Gurr, and Jaggers 2016). If both states have a Polity2 value greater or equal to 6, they have compatible preferences and are not threatening to one another—and thus coded 0. If a given state has incompatible preferences with all countries, then all states, especially those with the greatest power-resources, are potentially threatening. We consider alternative measures of preference compatibility in the supplementary files, which include other measures of joint democracy (Boix, Miller, and Rosato 2013; Pemstein, Meserve, and Melton 2010), defense pacts and alliances (Gibler 2009; Leeds et al. 2002), United Nations General Assembly voting similarity (Bailey, Strezhnev, and Voeten 2017), rivalry (Klein, Goertz, and Diehl 2006), energy consumption (Markowitz, Fariss, and McMahon Forthcoming; Greig and Enterline 2017), bilateral trade (Barbieri and Keshk 2012; Barbieri, Keshk, and Pollins 2009), diplomatic exchange (Bayer 2006), and shared Intergovernmental Organization membership (Pevehouse, Nordstrom, and Warnke 2004).

The inverse logged distance between state capitals creates a second weight, $w_{ijt} = \frac{1}{\ln(d_{ijt})}$, which operationalizes the loss of strength of state $j$ if it were to attempt to project power into state $i$. Short distances between the capital cities of two states yield values closer to 1 and further distances yield values closer to 0.

Substantively, the operationalization captures the total power-resources of state $i$ relative to all of the opponent states $j$ in the international system, weighed by distance and preference compatibility. In the supplementary files, we provide additional detail, a visual guide, and examples for each of the operational steps that generate the country-year values for the $potential\ threat_{it}$ variable.[10]

## Potential Threat using SDP vs. GDP

In this section, we compare the performance of SDP-based and GDP-based measures of potential threat. First, we compare the ability of distance- and preference-weighted power-resource ratios to correctly categorize country-year units as potentially threatening (concurrent validity). Next, we use SDP-based and GDP-based potential threat measures to explain variation in dependent variables that measure arming and power projection. We show that SDP-based models of states' military investments conform better with existing theoretical expectations than models based on GDP.

### *Assessing the most potentially threatening states using SDP vs. GDP*

Recall that the potential threat measure is the summation the distance- and preference-weighted power-resource ratios for all pairwise relationships for each country-year unit. If SDP better measures relative power-resources than GDP, then the potential threat variable using SDP should more accurately rank countries of greatest concern to other countries. As a control variable,

[10] See Figures 5 and 6 in the Supplementary Appendix.

we also generate a potential threat variable that uses *population* to measure relative power-resources within each pair of states. We show that the results using the population-based measure are very similar to the GDP version, which both contrast with SDP (see Supplementary Appendix Section C). This provides further evidence that poor, populous countries appear potentially threatening only because their large population gives them a large GDP.

Figure 4 displays the top ten potentially threatening states within the strategic environment of the United States by decade.[11] The upper panel illustrates the ranking based on distance-weighted relative power-resources using SDP; the middle panel plots the ranking for an analogous measure using GDP; the lower panel shows this ranking for the same measure using population. States with the largest weighted power-resource ratios rank highest on the list as the adversaries that potentially threaten the United States. The upper panel's order of states differs substantially compared to the middle and lower panels of the plot. The distance-weighted relative power-resource measures that incorporate GDP or population produce similar rankings, and place countries that were unlikely to threaten the United States at the top. In particular, China is ranked as the most potentially threatening country for the United States during the entire 19th and early 20th centuries—a period when it was not possible for China to develop a military force structure capable of projecting force to the shores of the United States. The measure that uses SDP provides a more historically valid ranking of those states that posed a potential threat to the United States than the measures using GDP or population.

By using a lighter shading of tiles for states with compatible preferences, Figure 4 highlights the role of preference compatibility for the United States' assessment of potential

[11] For graphs for Japan and the United Kingdom, see Figures 1 and 2 in the supplementary files.

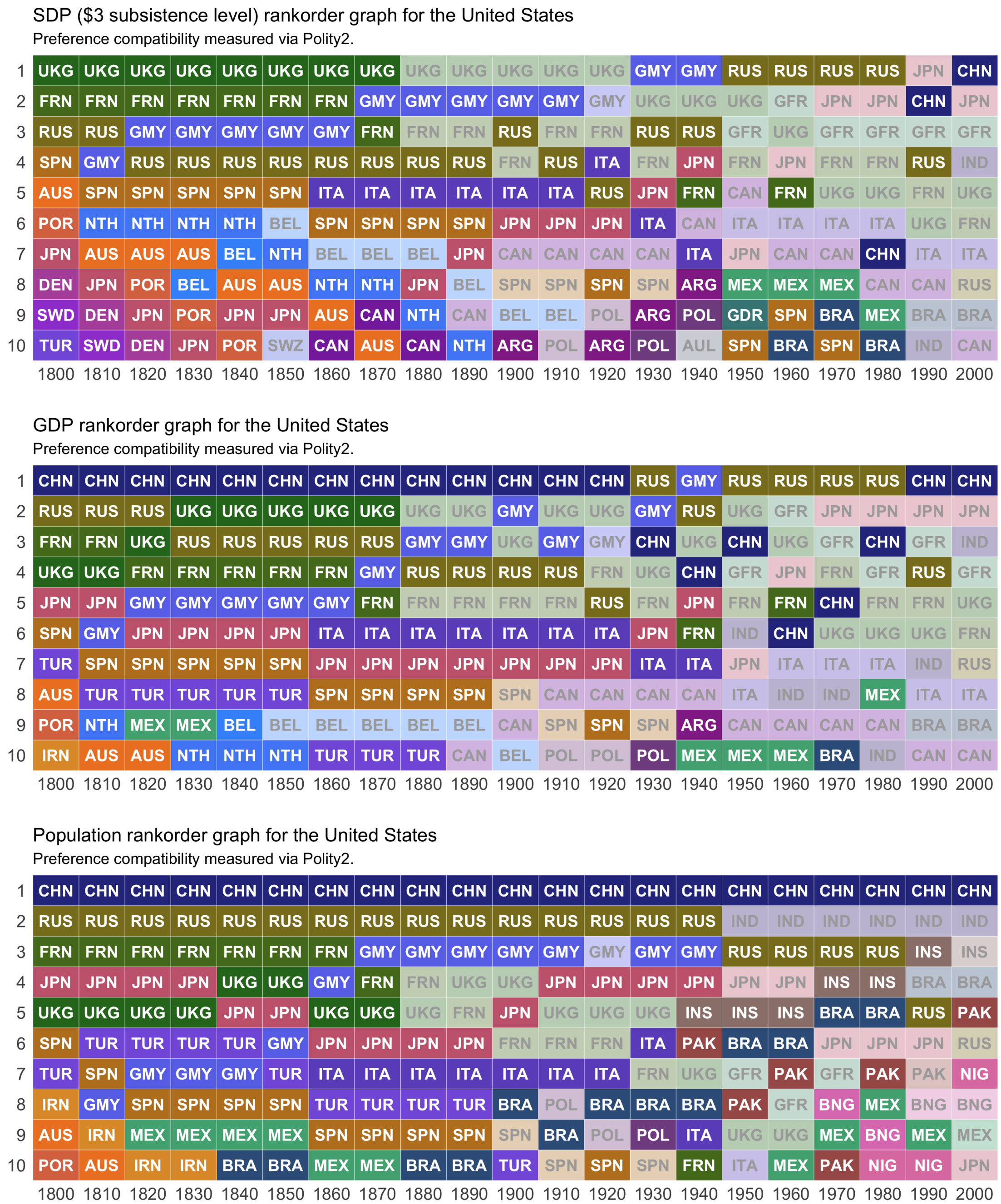



Figure 4: Top 10 potentially threatening states for the United States by decade for $3 per diem subsistence level relative SDP (upper panel), standard relative GDP (middle panel), and relative population (lower panel). All power-ratios are weighted by the inverse of the logged distance between Washington DC and the capital of the other country. Countries that are potentially threatening to the United States *are not* democratic, which is denoted through darker shading (not jointly democratic). Countries that are not potentially threatening to the US *are* democratic, which is denoted through lighter shading (jointly democratic). SDP better categorizes countries that are likely to be viewed as the most threatening by the United States. This is particularly clear when looking at the placement of China as a threatening country. China did not pose a threat to the US in the 19th century, even though it was very large in terms of its population. The measure incorporating SDP captures this intuition — it is therefore more valid with respect to its concurrent validity than the measure using standard GDP or population which both show China as the most threatening county to the United States for the entire period.

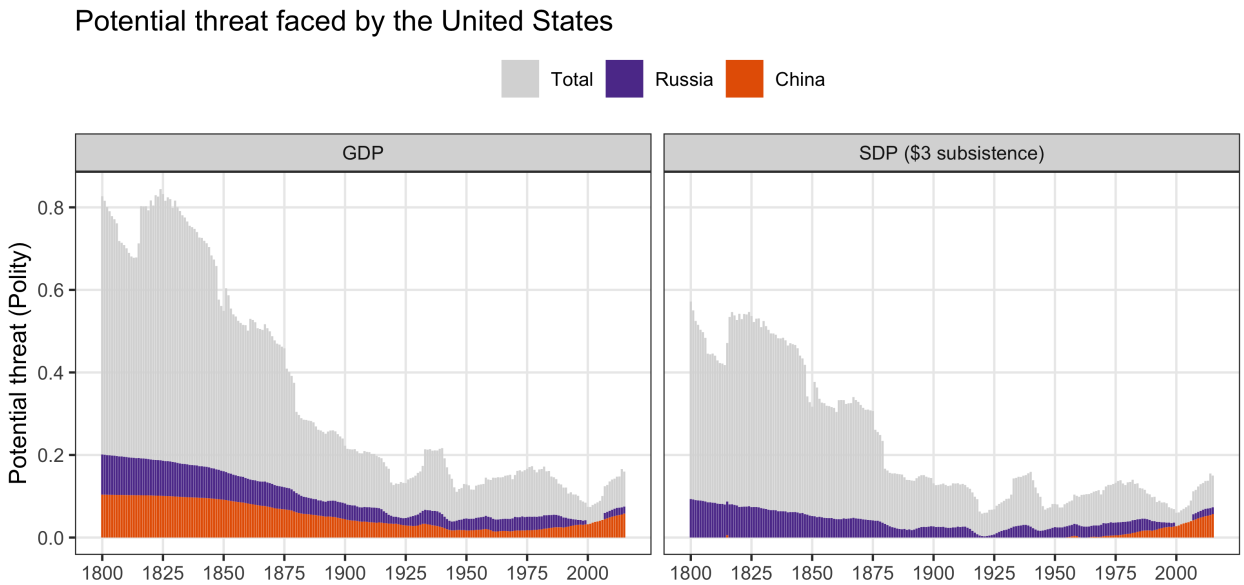



Figure 5: The graph illustrates how much China (in orange) and Russia (in purple) contribute to the total potential threat faced by the United States during the the 19th, 20th, and early 21st centuries. Preference compatibility is measured using Polity scores, and supplemented with data from Boix et al. [2013] to reduce the number of missing values. The left panel plots potential threat using GDP as an indicator for power-resources; the right panel plots the same measure when using SDP. The SDP measure presents a much more historically valid representation of how much countries contribute to the total potential threat in the United States' geopolitical environment. Based on GDP, China would be the largest contributor to the potential threat faced by the United States over the entire 19th century. Upon taking surplus resources into account, it becomes evident that China did not have the surplus power-resources to present a threat to the United States during this period. The total potential threat faced by the United States fell sharply over the course of the 19th century and remains much lower today than in the past. However, the rise of China as a major economic power leads to a visible upward trend in the total potential threat that the United States experiences in its geopolitical environment. In contrast, Russia's contribution to the total potential threat faced by the US today is at an all-time low. In relative terms, China is responsible for a larger share of the total potential threat faced by the United States today than Russia was over the entire course of the Cold War. This is both because the total level of potential threat for the United States today is much lower and recently, China has obtained a massive amount of surplus resources relative to the United States.

threats. Recall that the $p_{ijt}$ component of the potential threat variable down-weights the power-resource ratio between states if they are jointly democratic. Though the ranking corresponds to the top ten potentially threatening states for the United States based on their economic might and geographic proximity, this potential threat is mitigated if the country is democratic. Though they might have the economic capabilities to project power abroad, democratic states are not considered threatening by the United States because of the compatibility of their preferences. As former geopolitical rivals democratize, they stop contributing to the total potential threat faced by the United States. As a result, the strategic environment of the United States has become less threatening over time.

Figure 5 illustrates this downward trend. The height of each bar denotes the total level of potential threat faced by the United States each year. The total potential threat faced by the United States fell sharply over the course of the 19th century and remains much lower today than in the past. The colored values indicate how much China (in orange) and Russia (in purple) contribute to the total level of potential threat for the United States. The left panel plots potential threat using GDP as an indicator for economic resources; the right panel illustrates the same measure when SDP is used.[12]

The difference between measuring power-resources via SDP versus GDP for the United States' threat assessment is striking. Historically, when the United States considered potential threats, it paid careful attention to those states with the greatest power-resources. For over two hundred years, China was one of the largest *economies*, but only recently became one of the states

[12] Preference compatibility is measured using Polity2 scores, and supplemented with data from Boix, Miller, and Rosato (2013) to reduce the number of missing values.

with the greatest *power-resources* in the world. SDP yields a more historically valid representation of countries' contributions to the total level of potential threat in the United States' geopolitical environment. Today, China is the largest contributor to the total potential threat faced by the United States. In fact, contemporary China makes up a larger proportion of the total potential threat faced by the United States than Russia did at the height of the Cold War. The rise of China as a major economic power in the late 20$^{th}$ century dramatically increased the total potential threat that the United States experiences in its geopolitical environment.

***Modeling arming and power projection***

As previously discussed, while international relations theory suggests that the level of potential threat states face should explain their efforts to arm, existing research finds only mixed empirical support for this proposition (Cappella-Zielinski, Fordham, and Schilde 2017; Nordhaus, Oneal, and Russett 2012; Sandler and Hartley 1995, 46). We suggest a potential resolution for this puzzle by demonstrating that once we scale military expenditures by SDP, which corrects for systematic measurement error inherent in using GDP, we find a strong relationship between potential threat and military burdens.

SDP not only does a better job of measuring the distribution of relative power-resources—a core component in the level of potential threat a given state faces—it also does a better job of measuring the power-resources a given state could invest in arming. The example of China in 1990 illustrates this point. Military expenditures represented approximately 2.5 percent of China's GDP—a relatively modest military burden. However, even if China's 1.135 billion citizens at the time could survive on just $2 per day and the state could seize the entire remainder of economic income, SDP would be only half the value of its GDP. Hence, when military burden is measured using military expenditures as a percentage of SDP instead of GDP, China's military burden in 1990 was approximately twice as large as previously estimated.

We apply our potential threat measure to investigate the relationship between threat and military burden; comparing SDP-based and GDP-based approaches. We estimate several regression models in which we vary the measurement of SDP at $3, $2, and $1 per-day subsistence thresholds and compare those to GDP (equivalent to a $0 per-day threshold). We also control for the level of subsistence income available to countries in these regression models. We assess the relationship between the level of potential threat a state faces in its strategic environment (explanatory variable) and two measures of arming (dependent variables).

The first dependent variable is military burden, operationalized as military expenditure relative to income. Fearon (2018) argues that this is a reasonable proxy for a state's resources that could be dedicated to arming, and captures the magnitude of the social welfare costs of arming. In contrast to Fearon and others (Rasler and Thompson 1985; Khanna, Sandler, and Shimizu 1998), our preferred measure of military burden is a state's military expenditure as a percentage of the state's SDP rather than GDP or GNP. This operational choice better approximates the surplus resources available for arming.

We employ a revised and extended series of military expenditure as a percentage of income. This new indicator is created by first converting military expenditure values from the CINC data into constant monetary units (Singer 1987). We then use new GDP and population estimates to measure the proportion of a state's income (SDP or GDP) devoted to the military.[13] This allows us to extend data coverage to cover most countries in the world from 1816-2012. As a robustness check, we assess the relationship between potential threat and a second dependent variable that captures states' investments in arming: power projection capabilities. We

[13] The Supplementary Appendix describes in detail the new estimates of GDP and population.

operationalize power projection capabilities via a state's naval tonnage as a share of income. Data on naval tonnage come from Crisher and Souva (2014).[14] States with higher military spending as a percentage of income have higher military burdens, as do states that have more naval tonnage relative to their income.

For each dependent variable, we estimate a series of country-year fixed effect regression models. Right-hand side variables are lagged by one year. All models include controls for the natural log of income (SDP or GDP), the natural log of subsistence income, the natural log of population, Polity2 score, and a measure of potential threat based on population.[15]

Figure 6 displays coefficients and 95% confidence intervals of standardized potential threat variables with and without the subsistence income control variable.[16] Our preferred SDP-based approach to measuring power-resources using a $3 dollar-per-day threshold produces statistically significant relationships between the level of potential threat a state faces and both measures of arming, while a GDP-based approach does not. As we adjust our measure of SDP to use lower

---

[14] Figure 11 in the Supplementary Appendix compares temporal trends of all dependent variables.

[15] For additional discussion of control variables see Supplementary Appendix Section G.6

[16] Table 1 contains complete regression results based on SDP for the $3 per-day subsistence threshold. Table 2 contains analogous results using GDP. For most models, we do not observe statistically significant or substantively meaningful interaction effects between potential threat based on SDP (or GDP) and population. We therefore limit the results presented in Figure 6 to additive model specifications. Tables 4 and 5 in the Supplementary Appendix contain analogous models without control variables. Figure 15 in the Supplementary Appendix shows that the regression results are robust to limiting observations to the post-WWII period.

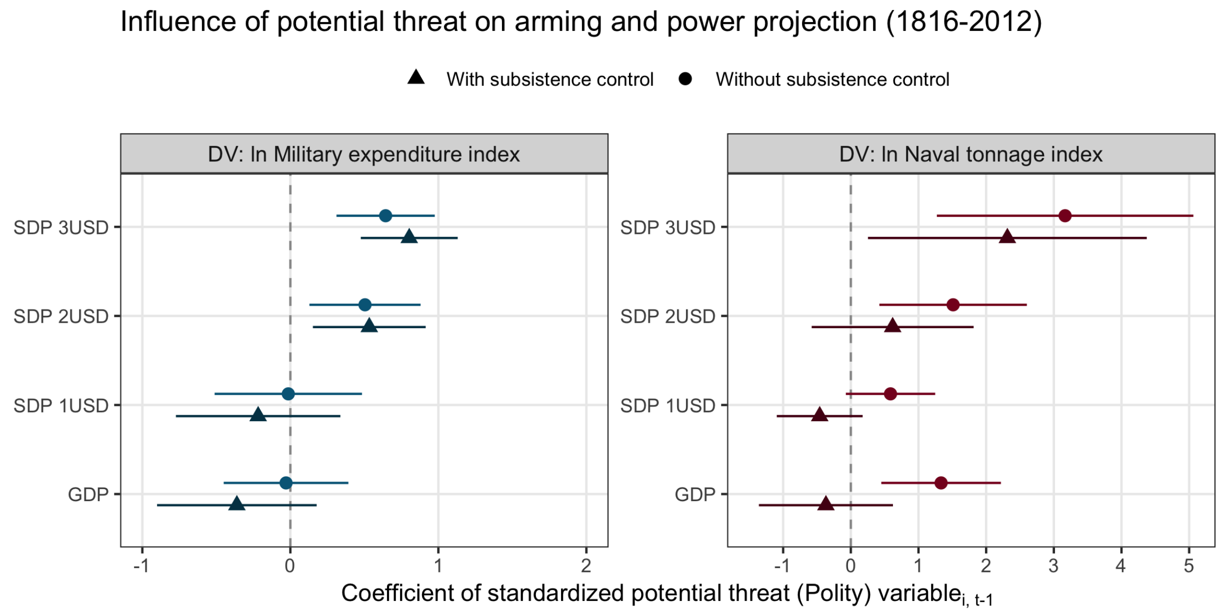



Figure 6: The graph plots the coefficients and 95% confidence intervals of standardized potential threat variables for regression models of two dependent variables — the military expenditure index and naval tonnage index — on potential threat and control variables. Preference compatibility is measured via joint democracy using Polity scores. The loss of strength gradient is measured as the inverse of logged distance. All models include controls for the natural log of income (SDP or GDP), a country's Polity2 score, and a measure of potential threat based on population. We distinguish between models that control for subsistence (or population for the GDP models) and those that do not. Standard errors are clustered by country; right-hand side variables are lagged by one year. Within each panel, rows distinguish between alternative measurements of economic power, that is SDP using a \$3, \$2, and \$1 per diem subsistence level, as well as standard GDP. Higher levels of potential threat are associated with higher investments in the military and naval capabilities — when measuring economic power using SDP. The plots show that when measuring power-resources using SDP, potential threat variables fare better in predicting arming and power projection, as compared to using GDP. As the subsistence level decreases, the coefficients become smaller and most cease to be statistically significant.

subsistence levels, the coefficients become smaller and cease to be statistically significant at conventionally accepted levels. For many models, decreasing the subsistence level to $1 or switching to GDP renders the effect of potential threat on military investments negative and insignificant for both arming and power projection. Controlling for population-based potential threat increases the size of the estimated effects.

Overall, higher levels of potential threat are associated with larger investments in military and naval capabilities when measuring economic power-resources using SDP. Crucially, all results depend on accurately measuring economic resources that states have at their disposal to invest in guns or butter. The conventionally used measure of GDP does not yield a statistically significant association between the level of potential threat and arming—a result that runs counter to existing theoretical expectations, but is consistent with the mixed empirical findings in the current literature. Only when measuring power-resources via SDP can we explain arming decisions based on the level of potential threat in states' geopolitical environment.

## Evaluating Military Burdens

In Figure 7, SDP measures states' military burdens, both at present and historically. Scaling military expenditures by SDP rather than GDP reveals that, historically, military burdens have been much higher than existing research suggests (Fearon 2018). In particular, many Asian states that were spending a relatively small percentage of their GDP on the military were, in fact, laboring under extraordinarily high military burdens—spending 25% to 50% of surplus.

Figure 8 illustrates these trends in specific countries. Scaling by SDP shows that the military burdens of poor states are much higher than the conventional measure of military expenditure as a percentage of GDP suggests. This divergence is particularly large for poor, populous countries like China. Over time, military burdens do fall for most states—especially for

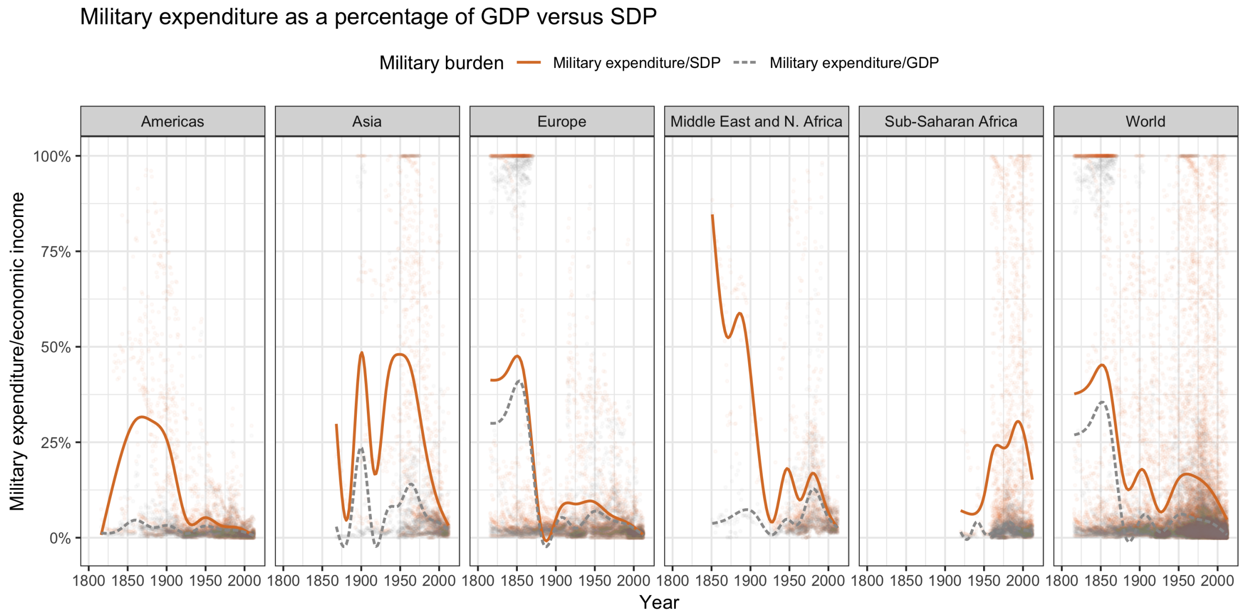



Figure 7: The graphs illustrates the change in military burden over time for five broadly defined regions: the Americas (including the US and Canada), Europe (including Russia), Asia, the Middle East and North Africa, and Sub-Saharan Africa. The lines represent the smoothed average over all countries in the region for two alternative indicators of military burden: military expenditure as a percentage of SDP (orange solid line) versus as a percentage of GDP (black dashed line). SDP is computed based on a $3 per diem subsistence level. Surplus is truncated to zero — a state either has positive surplus or it does not have surplus at all. Both the average size of the military burden as well as the margin of difference between military expenditure as a proportion of SDP versus GDP decrease as states develop and become wealthier. While the global average of military burdens as a percentage of SDP today still exceeds the average of military burdens as a percentage of GDP, this difference appears to be driven by countries in Sub-Saharan Africa shouldering massive burdens relative to their power-resources. As poorer countries continue to develop and increase their surplus income, the two alternative measures of military burdens are expected to converge. The graph shows that military expenditures as a percentage of SDP have fallen much faster than suggested by previous empirical analyses that measure military burden as military expenditure as a percentage of GDP.

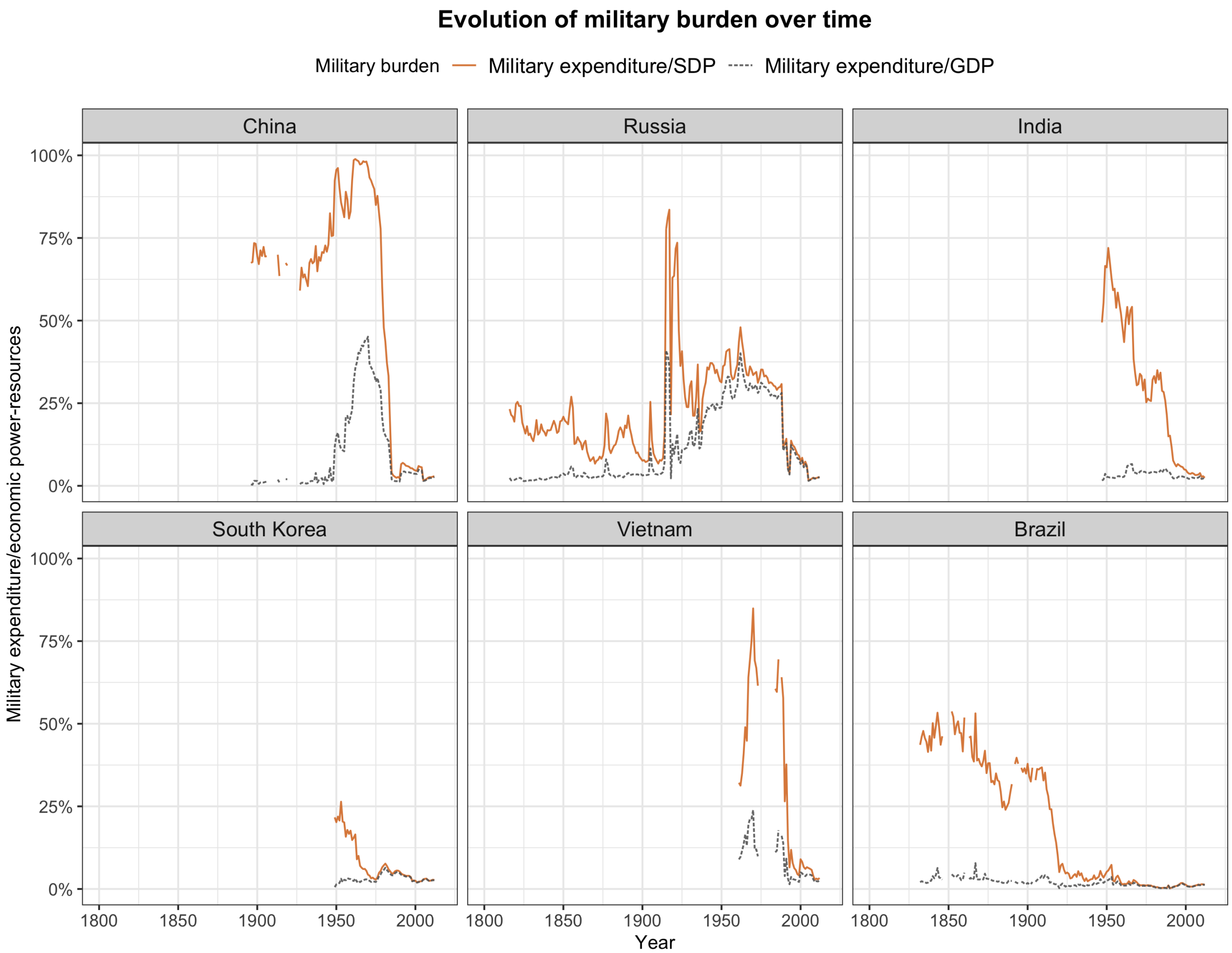



Figure 8: The graphs illustrate the change in military burden over time for select countries. The orange solid line shows military expenditure as a percentage of SDP; the grey dashed line shows military expenditure as a percentage of GDP. SDP is computed based on a $3 per diem subsistence level. Surplus is truncated to zero — a state either has positive surplus or it does not have surplus at all. The graph shows that when measuring a state's resources dedicated to the military after the subsistence needs of the population have been met, the military burden of poor states is much higher than the conventional measure of military expenditure as a percentage of GDP suggests.

major powers—but these costs remain high for poor states where most GDP is needed to cover basic subsistence and SDP is low.

The good news for states in the Western Hemisphere and Europe is that military burdens are much lower than in the past. Additionally, despite alarmist warnings of impending arms races and conflict, most states in Asia today face dramatically lower military burdens than they have in the past two hundred years. However, this decline only becomes apparent when using SDP instead of GDP. For illustration, as a share of GDP, South Korea's military spending decreases only slightly from 3.7% during the Cold War (1954-1991) to 3.0% after (1992-2012). However, as a share of SDP, military spending plunges from 9.9% during the Cold War to just 3.2% after. Scaling by SDP reveals a sharp decline in South Korea's willingness to prioritize guns over butter and implies that leaders in Seoul believe that the level of threat they face has fallen enough to justify lower military burdens (Lind 2011). While South Korea may choose to increase its military burden in the future, they will be doing so from a historically low baseline. This point informs the debate over the degree to which states are balancing China's rise; in general, the willingness of states in the region to bear a high military burden in recent years is lower than commonly recognized.

## Conclusion

GDP is a widely adopted measure of the financial resources that a state can potentially invest in guns or butter (Coyle 2014). We introduce the concept of SDP, which separates the subsistence income, or "bread", needed for the population to survive from the surplus income that can potentially be extracted and invested.

Using GDP as a measure of power-resources instead of SDP systematically overestimates the financial resources available to the governments of poor, populous countries and underestimates the speed with which these resources increase during the early stages of industrialization (i.e. during the stage when countries first begin to produce a significant surplus and the government could extract income without starving its citizens). Similarly, using military expenditures as a share of GDP to measure military burdens leads scholars to underestimate the size of the military burdens born by poor states. These conceptual errors are particularly problematic for historical-comparative work because, for most of human history, virtually all states had incomes at or near subsistence levels—at least on average—creating a large divergence between estimates of SDP and GDP. This is a major issue. For illustration, as recently as 2007, the gap between SDP and GDP was still large for over half of the world's states, which were classified by the World Bank as either low income (GNI per-capita below $995 or lower middle income (GNI per-capita between $996 and $3895) by the World Bank.[17] At the time, more than 70% of the World' population lived in such states.

Thus, using new data on SDP, we reveal that poor, populous states are far less powerful than generally assumed, and that low-income countries historically and today face more severe guns-butter tradeoffs and higher military burdens than GDP-based measures suggest. Thus, previous scholarship has dramatically underestimated the benefits from factors that allow poor states to lower their defense burdens, such as hierarchy and the liberal peace. Additionally, our results offer a potential solution to the puzzle of why previous scholarship has found only mixed support for one of the core propositions of international relations theory: *states arm against potential threats*. Once we apply SDP to correct for the systematic measurement error associated with GDP, we find strong empirical support for this proposition.

In addition to our theoretical contributions, we provide new data that extend cross-national

[17] (World Bank 2018)

coverage of GDP, SDP, and population from the present back to 1816 for nearly every country in the world. These new data allow scholars to apply our measure of SDP to reexamine a broad range of research questions in which GDP is frequently used as a proxy for potential or actual state capacity.

An area that is directly parallel to a state's capacity to arm is a state's capacity to repay its debts. For illustration, Malawi in 2015 had a plausibly manageable debt burden of 39.5% of GDP (The World Bank 2017; Reinhart and Rogoff 2010). However, Malawi's debt burden amounts to a crushing 226.5% of SDP. The annual payments on a debt of that magnitude consume a significant proportion of the country's surplus income, even if they constitute a seemingly manageable share of total economic income. This matters for understanding the ability of states to manage debt and engage with international institutions such as the World Bank and International Monetary Fund.

More broadly, SDP represents the resources a government may potentially draw upon to build physical infrastructure, establish the rule of law, or provide public services such as education and health care. SDP is not a direct measure of government capacity, but it measures the upper bound of the income states may sustainably extract from its citizens to develop that capacity.

When measuring the upper bound of a state's extractable income, SDP compared to GDP, showcases just how much more constrained state capacity is in low-income countries, relative to middle-income and upper-income countries. However, SDP also reveals that, in the early stages of economic development, government capacity expands much more rapidly than currently realized, increasing the compound returns to growth for countries near subsistence levels. SDP-based assessments are unlikely to lead to exclusively pessimistic or exclusively optimistic new conclusions with respect to the prospects for peace, prosperity, and democracy in the developing world. Thus, analysis of SDP has the potential to radically reshape our understanding of the extent to which different fiscal strategies are plausible or desirable in lower-income countries.

# A Paper Appendix

Table 1: Regression models relating different specifications of the potential threat variable to investments in arming and power projection. Power-resources are measured using SDP at a \$3 per diem subsistence level. The loss of strength gradient is conceptualized as curvilinear using the formula $\frac{1}{\log(distance)}$. Interest compatibility based joint democracy using Polity scores.

| | Dependent variables | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Military expenditure/SDP | | | | | Naval tonnage/SDP | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Potential threat (SDP)$_{i,t-1}$ | 0.62$^{***}$ | 0.26$^{**}$ | | 0.72$^{***}$ | 0.72$^{***}$ | 1.87$^{***}$ | 1.60$^{**}$ | | 2.53$^{*}$ | 2.87$^{*}$ |
| | (0.09) | (0.10) | | (0.16) | (0.15) | (0.49) | (0.55) | | (1.00) | (1.15) |
| Potential threat (Population)$_{i,t-1}$ | | | −0.05 | −0.74$^{***}$ | −0.76$^{**}$ | | | 0.82 | −1.64 | 0.19 |
| | | | (0.13) | (0.21) | (0.25) | | | (0.50) | (1.04) | (1.08) |
| ln SDP$_{i,t-1}$ | | −0.08$^{***}$ | −0.09$^{***}$ | −0.06$^{***}$ | −0.06$^{***}$ | | −0.04 | −0.09 | −0.01 | −0.01 |
| | | (0.01) | (0.01) | (0.01) | (0.01) | | (0.06) | (0.06) | (0.06) | (0.06) |
| ln Subsistence$_{i,t-1}$ | | −0.21 | −0.19 | −0.38$^{*}$ | −0.38$^{*}$ | | 1.95 | 2.45$^{**}$ | 1.54 | 2.03$^{*}$ |
| | | (0.18) | (0.17) | (0.18) | (0.17) | | (1.00) | (0.91) | (1.02) | (0.99) |
| Polity2$_{i,t-1}$ | | −0.03$^{**}$ | −0.04$^{***}$ | −0.04$^{***}$ | −0.04$^{**}$ | | 0.01 | −0.02 | −0.002 | 0.04 |
| | | (0.01) | (0.01) | (0.01) | (0.01) | | (0.04) | (0.05) | (0.05) | (0.04) |
| Interaction Potential threat$_{i,t-1}$ | | | | | 0.01 | | | | | −1.28$^{**}$ |
| | | | | | (0.08) | | | | | (0.43) |
| Fixed-effects | CY | CY | CY | CY | CY | CY | CY | CY | CY | CY |
| Observations | 11,616 | 11,616 | 11,616 | 11,616 | 11,616 | 12,033 | 12,033 | 12,033 | 12,033 | 12,033 |
| Adjusted R$^2$ | 0.02 | 0.09 | 0.09 | 0.11 | 0.11 | 0.01 | 0.02 | 0.01 | 0.03 | 0.05 |

*Note:* $^{*}$p<0.05; $^{**}$p<0.01; $^{***}$p<0.001
Clustered standard errors by country (Satterthwaite correction) in parentheses.
Potential threat variables are standardized.
CY denotes two-way fixed effects.
Period of observation: 1816-2012.

Table 2: Regression models relating different specifications of the potential threat variable to investments in arming and power projection, omitting population as a control. Power-resources are measured using GDP. The loss of strength gradient is conceptualized as curvilinear using the formula $\frac{1}{\log(distance)}$. Interest compatibility based joint democracy using Polity scores.

| | Dependent variables | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Military expenditure/GDP | | | | | Naval tonnage/GDP | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Potential threat (GDP)$_{i,t-1}$ | 0.23 | −0.01 | | −0.03 | −0.34 | 0.23 | 0.36$^{**}$ | | 1.33$^{**}$ | 1.42$^{**}$ |
| | (0.18) | (0.15) | | (0.21) | (0.28) | (0.14) | (0.14) | | (0.45) | (0.44) |
| Potential threat (Population)$_{i,t-1}$ | | | −0.003 | 0.02 | −0.09 | | | −0.04 | −1.31$^{*}$ | −1.26$^{*}$ |
| | | | (0.18) | (0.26) | (0.26) | | | (0.16) | (0.52) | (0.56) |
| ln GDP$_{i,t-1}$ | | 0.03 | 0.04 | 0.03 | −0.06 | | 0.69$^{**}$ | 0.56$^{*}$ | 0.88$^{**}$ | 0.91$^{**}$ |
| | | (0.20) | (0.21) | (0.20) | (0.23) | | (0.26) | (0.25) | (0.31) | (0.30) |
| Polity2$_{i,t-1}$ | | −0.04$^{***}$ | −0.04$^{***}$ | −0.04$^{***}$ | −0.05$^{***}$ | | −0.02 | −0.03$^{*}$ | −0.02 | −0.02 |
| | | (0.01) | (0.01) | (0.01) | (0.02) | | (0.02) | (0.02) | (0.02) | (0.01) |
| Interaction Potential threat$_{i,t-1}$ | | | | | 0.24 | | | | | −0.08 |
| | | | | | (0.19) | | | | | (0.11) |
| Fixed-effects | CY | CY | CY | CY | CY | CY | CY | CY | CY | CY |
| Observations | 11,616 | 11,616 | 11,616 | 11,616 | 11,616 | 12,033 | 12,033 | 12,033 | 12,033 | 12,033 |
| Adjusted R$^2$ | −0.03 | −0.01 | −0.01 | −0.01 | 0.0005 | −0.02 | 0.01 | 0.002 | 0.03 | 0.03 |

*Note:* $^{*}$p<0.05; $^{**}$p<0.01; $^{***}$p<0.001
Clustered standard errors by country (Satterthwaite correction) in parentheses.
Potential threat variables are standardized.
CY denotes two-way fixed effects.
Period of observation: 1816-2012.

## References


Bailey, Michael A., Anton Strezhnev, and Erik Voeten. 2017. "Estimating Dynamic State Preferences from United Nations Voting Data." *Journal of Conflict Resolution* 61 (2): 430–56. https://doi.org/10.1177/0022002715595700.

Barbieri, Katherine, and Omar Keshk. 2012. "Correlates of War Project Trade Data Set Codebook." http://correlatesofwar.org.

Barbieri, Katherine, Omar M.G. Keshk, and Brian M. Pollins. 2009. "Trading Data: Evaluating Our Assumptions and Coding Rules." Conflict Management and Peace Science. http://cmp.sagepub.com/content/26/5/471.abstract.

Bayer, Resat. 2006. "Diplomatic Exchange Data Set V2006.1." http://correlatesofwar.org.

Beckley, Michael. Forthcoming. "The Power of Nations: Measuring What Matters." *International Security* 43 (2): 7–44.

Boix, Carles, Michael Miller, and Sebastian Rosato. 2013. "A Complete Data Set of Political Regimes, 1800–2007." *Comparative Political Studies* 46 (12): 1523–54.

Cappella-Zielinski, Rosella. 2016. *How States Pay for Wars*. Cornell University Press.

Cappella-Zielinski, Rosella, Benjamin O. Fordham, and Kaija E. Schilde. 2017. "What Goes up, Must Come down? The Asymmetric Effects of Economic Growth and International Threat on Military Spending." *Journal of Peace Research* 54 (6): 791–805.

Coyle, Diane. 2014. *GDP A Breif But Affectionate History*. Princeton University Press.

Crisher, Brian, and Mark Souva. 2014. "Power at Sea: A Naval Power Dataset, 1865-2011." *International Interactions* 40 (4): 602–29.

Ezzati, Majid. 2004. "Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors." World Health Organization. http://apps.who.int/iris/bitstream/10665/42792/1/9241580348_eng_Volume1.pdf.

Fearon, James D. 2018. "Cooperation, Conflict, and the Costs of Anarchy." *International Organization* 72 (3).

Ferreira, Francisco. 2015. "The International Poverty Line Has Just Been Raised to $1.90 a Day, but Global Poverty Is Basically Unchanged. How Is That Even Possible?" Let's Talk Development. October 4, 2015. http://blogs.worldbank.org/developmenttalk/international-poverty-line-has-just-been-raised-190-day-global-poverty-basically-unchanged-how-even.

Garfinkel, Michelle R., and Constantinos Syropoulos. 2019. "Problems of Commitment in Arming and War: How Insecurity and Destruction Matter." *Public Choice* 178 (3–4): 349–69. https://doi.org/10.1007/s11127-018-0601-x.

Gibler, Douglas M. 2009. *International Military Alliances, 1648-2008*. Washington, D.C.: CQ Press.

Greig, J. Michael, and Andrew J. Enterline. 2017. "National Material Capabilities (NMC) Data Documentation Version 5.0." Correlates of War Project.

Kadera, Kelly M., and Gerald L. Sorokin. 2004. "Measuring National Power." *International Interactions* 30 (3): 211--230.

Kennedy, Paul. 1987. *The Rise and Fall of the Great Powers: Economic Change and Military Power from 1500 to 2000*. New York: Random House.

Khanna, Jyoti, Todd Sandler, and Hirofumi Shimizu. 1998. "Sharing the Financial Burden for U.N. and NATO Peacekeeping, 1976-1996." *Journal of Conflict Resolution* 42 (2): 176–95. https://doi.org/10.1177/0022002798042002003.

Klein, James P., Gary Goertz, and Paul F. Diehl. 2006. "The New Rivalry Dataset: Procedures and Patterns." *Journal of Peace Research* 43 (3): 331--348.

Lake, David A. 2009. *Hierarchy in International Relations*. Cornell University Press.

Lake, David A., and Matthew A. Baum. 2001. "The Invisible Hand of Democracy: Political Control and the Provision of Public Services." *Comparative Political Studies* 34 (6): 587–621.

Lamborn, Alan C. 1983. "Power and the Politics of Extraction." *International Studies Quarterly* 27 (2): 125–46. https://doi.org/10.2307/2600540.

Leeds, Brett, Jeffrey Ritter, Sara Mitchell, and Andrew Long. 2002. "Alliance Treaty Obligations and Provisions, 1815-1944." *International Interactions* 28 (3): 237–60. https://doi.org/10.1080/03050620213653.

Lind, Jennifer. 2011. "Democratization and Stability in East Asia." *International Studies Quarterly* 55 (2).

Markowitz, Jonathan N, and Christopher J Fariss. 2018. "Power, Proximity, and Democracy: Geopolitical Competition in the International System." *Journal of Peace Research* 55 (1): 78–93. https://doi.org/10.1177/0022343317727328.

Markowitz, Jonathan N., Christopher J. Fariss, and R. Blake McMahon. Forthcoming. "Producing Goods and Projecting Power: How What You Make Influences What You Take." *Journal of Conflict Resolution*, 0022002718789735. https://doi.org/10.1177/0022002718789735.

Marshall, Monty G., Ted R. Gurr, and Keith Jaggers. 2016. "Polity IV Project. Political Regime Charateristics and Transitions, 1800-2015, Dataset Users' Manual." Vienna, VA: Center for Systemic Peace.

Milward, Alan S. 1977. *War, Economy and Society, 1939-1945*. History of the World Economy in the Twentieth Century. 5. Berkeley: University of California Press.

Nordhaus, William, John R. Oneal, and Bruce Russett. 2012. "The Effects of the International Security Environment on National Military Expenditures: A Multicountry Study." *International Organization* 66 (3): 491–513.

Norloff, Carla, and William C Wohlforth. Forthcoming. "Raison de l'Hégémonie (The Hegemon's Interest): Theory of the Costs and Benefits of Hegemony." *Security Studies*.

Pemstein, Daniel, Stephen A. Meserve, and James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18 (4): 426–49. https://doi.org/10.1093/pan/mpq020.

Pevehouse, Jon, Timothy W. Nordstrom, and Kevin Warnke. 2004. "The Correlates of War 2 International Governmental Organizations Data Version 2.0." *Conflict Management and Peace Science* 21 (2): 101–19.

Rasler, Karen A., and William R. Thompson. 1985. "War Making and State Making: Governmental Expenditures, Tax Revenues, and Global Wars." *American Political Science Review* 79 (2).

Reinhart, Carmen M., and Kenneth S. Rogoff. 2010. "Growth in a Time of Debt." *American Economic Review: Papers and Proceedings*.

Sandler, Todd, and Keith Hartley. 1995. *The Economics of Defense*. New York: Cambridge University Press.

Singer, J. David. 1987. "Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816--1985." *International Interactions* 14: 115–32.

Tanzi, Vito, and Ludger Schuknecht. 2000. *Public Spending in the 20th Century. A Global Perspective.* Cambridge: Cambridge University Press.

The World Bank. 2017. "World Development Indicators." Washington, D.C.: World Bank. http://data.worldbank.org/data-catalog/world-development-indicators.

World Bank. 2018. "Classifying Countries by Income." October 4, 2018. http://datatopics.worldbank.org/world-development-indicators/stories/the-classification-of-countries-by-income.html.

# Supplementary Files

### Bread before guns or butter:

### Introducing Surplus Domestic Product (SDP)

June 2019

## Contents

# Introduction to the Online Appendix

The supplementary material presented in this document provides additional graphs and details about both the construction of the surplus domestic product (SDP) and subsistence measures, the potential threat measure, as well as the the latent variable model developed in the article "Bread before guns or butter: Introducing Surplus Domestic Product (SDP)." The main manuscript makes reference to the materials contained here. The estimates presented in this appendix along with the code necessary to implement the models in `R` will be made publicly available.

# A GDP = surplus + subsistence

Gross Domestic Product is technically an accounting identity, made up of four component parts, such that $GDP = consumption + investments + (imports - exports)$. Our decomposition of GDP into surplus and subsistence is also an accounting identity. Subsistence is technically part of the consumption component of the GDP accounting identity. Surplus is also part of the consumption component in addition to the other component parts of the identity. For country $i \in \{1, ..., N\}$ in year $t \in \{1800, ..., 2018\}$, the equation for Gross Domestic Product as an additive identity of two income components is:

$$GDP_{it} = surplus_{it} + subsistence_{it} \tag{1}$$

Before we define $surplus_{it}$ income and $subsistence_{it}$ income, we first have to define the *minimum surplus value*: $v_{it}$ for country $i$ in year $t$, which is calculated as:

$$v_{it} = \tau * 365 * population_{it} \tag{2}$$

where $\tau \in \{\$0, \$1, \$2, \$3\}$ is the daily surplus threshold. Conceptually, $\tau$ represents the minimal amount of income necessary for an individual to meet her caloric needs. As we describe in detail in the main manuscript, in the contemporary period, it is at least \$2 in constant US dollars and the World Bank recommends \$3 in constant US dollars.[1]

The variable $subsistence_{it}$ takes on positive dollar values that are less than or equal to the surplus value $v_{it}$ such that:

$$subsistence_{it} = \begin{cases} v_{it} & \text{if } GDP_{it} > v_{it} \\ GDP_{it} & \text{if } GDP_{it} \leq v_{it} \end{cases} \tag{3}$$

The variable $surplus_{it}$ takes on positive dollar values only if $GDP_{it}$ is greater than the value of the surplus value $v_{it}$ such that:

$$surplus_{it} = \begin{cases} GDP_{it} - v_{it} & \text{if } GDP_{it} > v_{it} \\ 0 & \text{if } GDP_{it} \leq v_{it} \end{cases} \tag{4}$$

In the paper we refer to $surplus_{it}$ as surplus domestic product $SDP_{it}$.[2]

Next we define the level of investment in military expenditures that a state makes each year. This is

[1]The value of $\tau$ has likely changed over time. In a future project, we plan to try to estimate this value based on historic information about the subsistence behaviors of individuals living and working in different periods of time and different countries. Such a measurement project is outside the scope of the current paper. Thus, we opted to set $\tau$ to one of four different constant values that we use in our statistical models.

[2]It is likely the case that many of the states without surplus income are still importing and exporting some goods and making some investments. However, to do this, the state must extract from the basic subsistence income of the citizens.

an important quantity that international relations theorists demonstrate is related to $milex\_ratio_{it}$. This is our main dependent variable. We calculate this in two ways. Both are ratios of the dollars spent out of all the available income to be spent by the state; given the political ability and willingness to extract it. In both cases, we place the total amount of military dollars spent as a ratio of either surplus domestic product $milex\_ratio_{it} = \frac{milex_{it}}{SDP_{it}}$, or gross domestic product $milex\_ratio_{it} = \frac{milex_{it}}{GDP_{it}}$.

With these two alternative versions of the dependent variable, we then specify a regression model to analyze the correlation between this quantity and several important covariates. We specify the following primary estimating equation:

$$milex\_ratio_{it} = \beta_1 * surplus_{it} + \beta_2 * subsistence_{it} + \mathbf{X}_{it}\beta + a_i + u_t + \epsilon_{it}, \tag{5}$$

where $\mathbf{X}_{it}$ is a matrix of additional covariates. We specify two-way fixed effects; using $a_i$, the country fixed effect, and $u_t$, the time-period fixed effect.

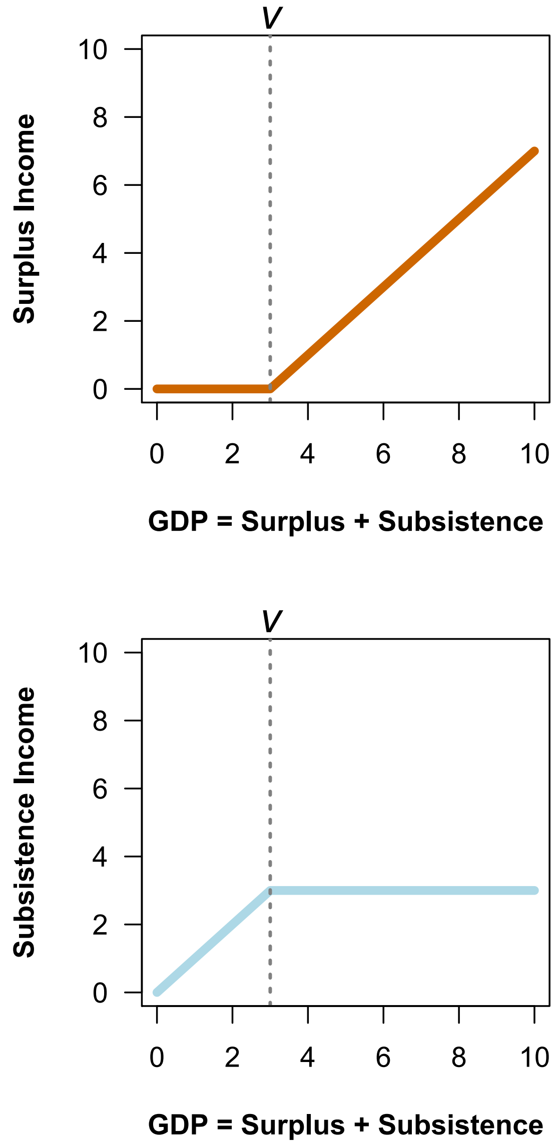



Figure 1: The dollar values displayed on the x-axes and y-axes in the panels above are in billions of $US. Suppose a country with a population of 2,739,726 people. Such a country needs to generate 3 billion $US dollars (365 days * $3 per-day * 2,739,726 people) per year to healthfully sustain each member of the population over the long term, which is $v$, the minimum surplus value. Such a country is consuming all of its income for subsistence up until it generates income surpassing this minimum surplus value $v$. Once such a country generates income greater than $v$, the country is generating positive surplus income which it can invest in items other than “bread” (e.g., “butter” or “guns”). Poor and under-developed countries do exist today and in earlier periods of history with income levels at and below this threshold. Indeed, some state governments have worked diligently to develop extractive institutions to take even the subsistence income of the population. However, these states do not maintain the levels of healthy adults necessary for other state-making tasks (e.g., conscription) to sustain such a strategy over the long run.

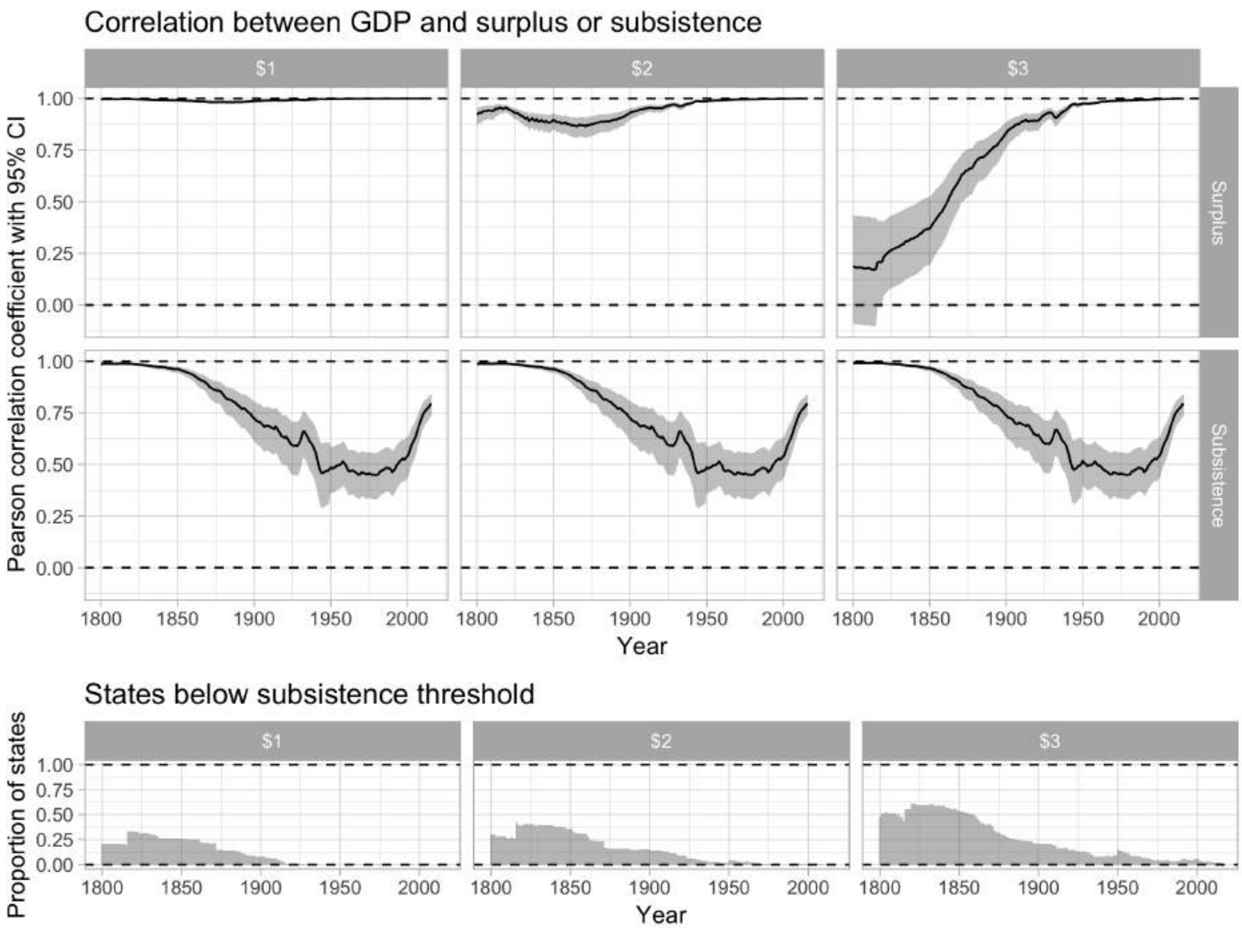



Figure 2: The top row of panels shows the yearly correlation between GDP and surplus income (SDP). The middle row of the panels shows the yearly correlation between GDP and subsistence income. The bottom row of panels shows the yearly proportion of countries that generate enough income to pass above the subsistence threshold at $1, $2, or $3 per person per day. The columns indicate these subsistence thresholds for each set of panels.

## B Rankorder Graphs

The figures below display the top ten potentially threatening states within the strategic environment of Japan and the United Kingdom — analogous to the graph representing the strategic environment of the United States presented in the main manuscript. The highest panel illustrates the rank order of the top ten potentially threatening states when using the distance-weighted relative power ratio that incorporates SDP; the middle panel plots an analogous ranking for the same measure using GDP; the lowest panel shows weighted relative power ratios based on a distance-weighted relative population measure. Opponent states with a large SDP that are geographically proximate to each of these states should have higher weighted relative power ratios than states with either low levels of SDP, or that are geographically distant, or both. The Loss of Strength Gradient is conceptualized as the inverse of the logged distance between capital cities.

We use concurrent validity to make these assessments. Concurrent validity is an assessment of the ability of an empirical measure to distinguish between cases that are distinct based on some prior theoretical knowledge about the status of those cases (Trochim and Donelly, 2008, 60). To have concurrent validity, the potential threat measure should be able to accurately categorize the opponent states that are the most threatening to any individual state in any historic period. The definition for concurrent validity is analogous to face validity, except that face validity assesses the link between the theory and the operational protocol, while concurrent validity assesses the link between the operational protocol and data. It should also be able to categorize states that face highly threatening strategic environments and those that do not.

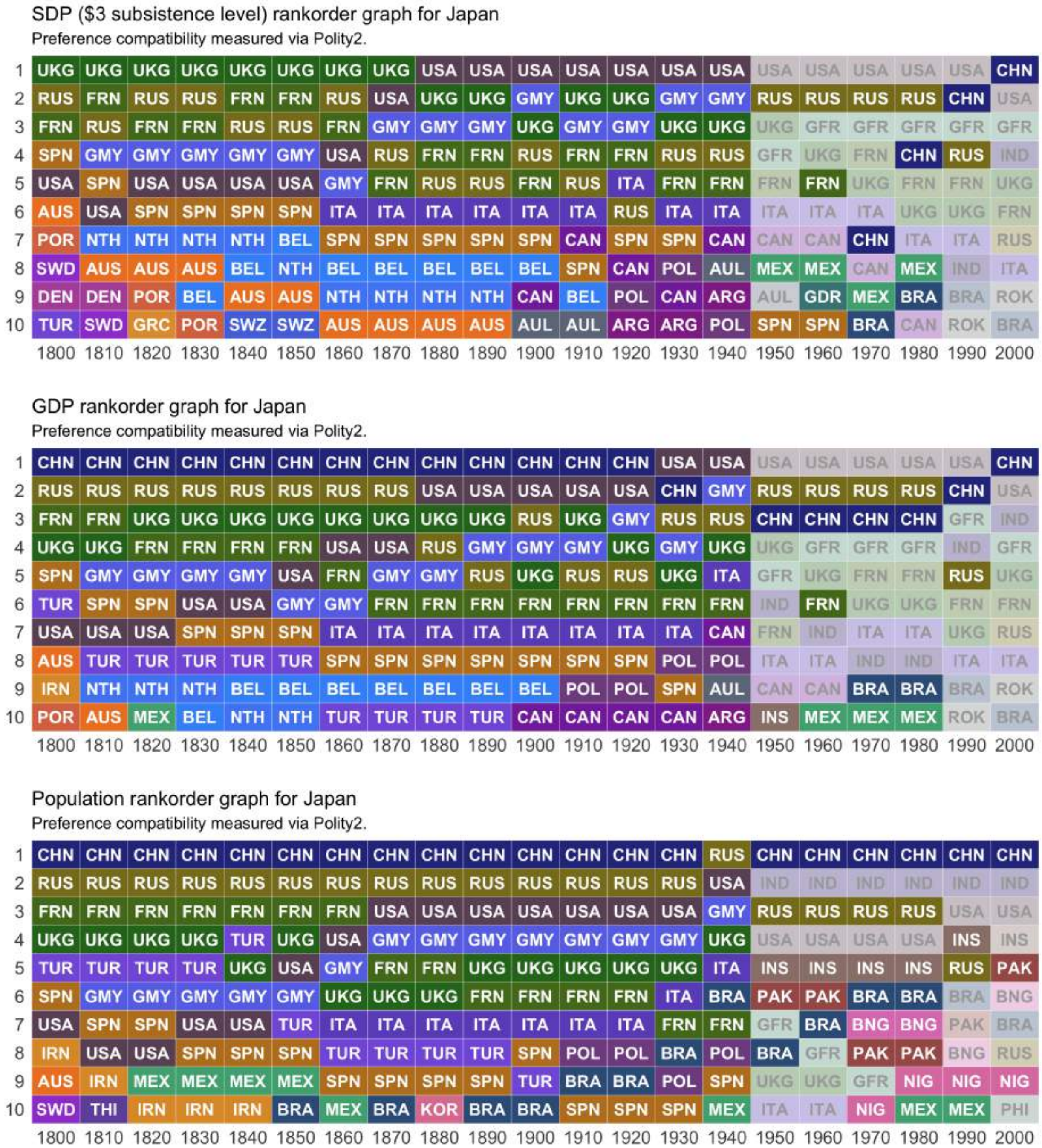



Figure 3: The Figure displays the top 10 potentially threatening states for Japan by decade for $3 per diem subsistence level relative SDP on the upper, standard relative GDP in the middle, and relative population on the lower panel. Dyads that *are not* jointly democratic are potentially threatening and denoted through opaque shading. Dyads that *are* jointly democratic are not potentially threatening and denoted through brighter shading. Dyads are coded as jointly democratic if both states have a Polity score greater or equal to six. All power-ratios are weighted by the inverse of the logged dyadic distance.

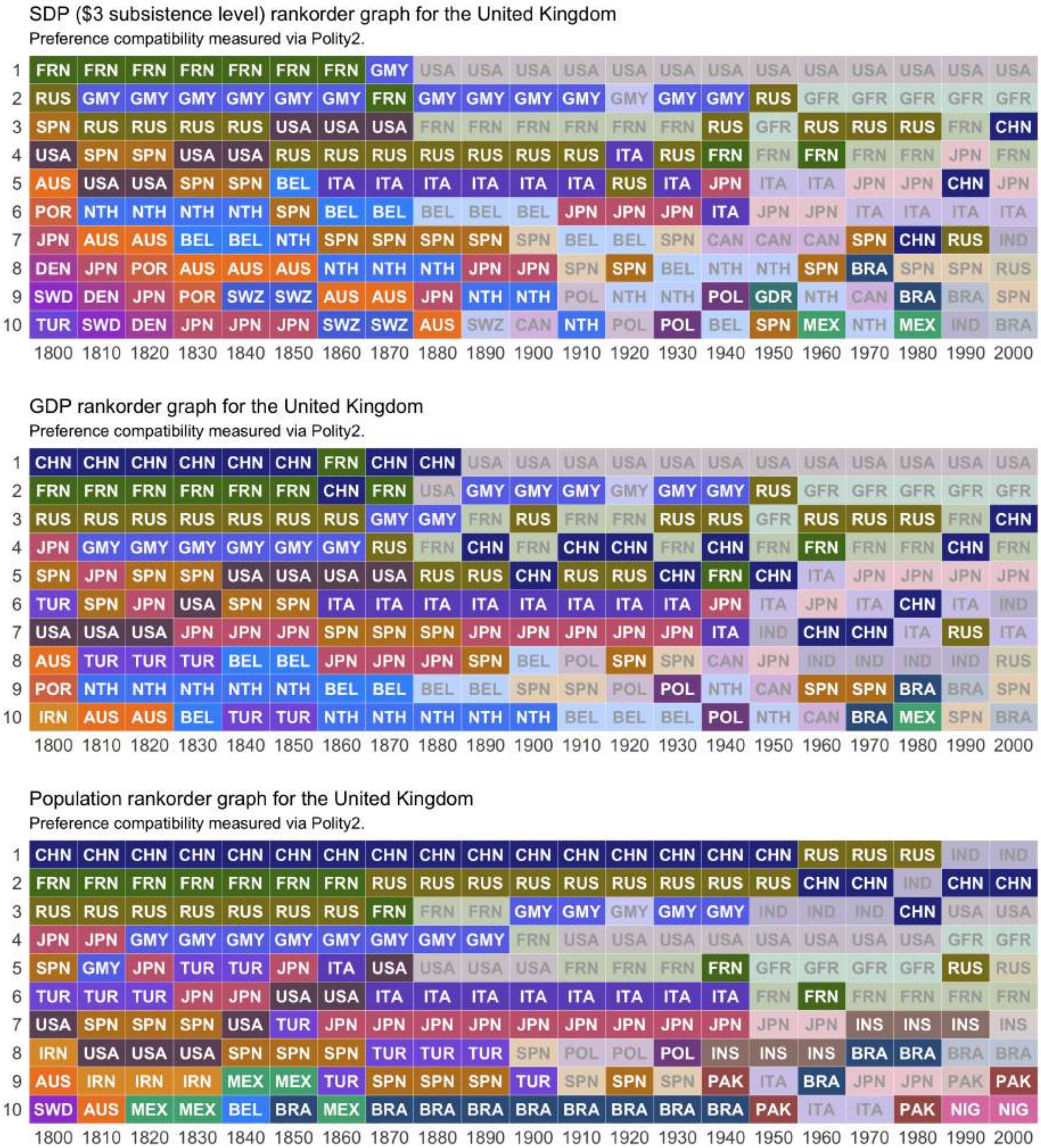



Figure 4: The Figure displays the top 10 potentially threatening states for the United Kingdom by decade for $3 per diem subsistence level relative SDP on the upper, standard relative GDP in the middle, and relative population on the lower panel. Dyads that *are not* jointly democratic are potentially threatening and denoted through opaque shading. Dyads that *are* jointly democratic are not potentially threatening and denoted through brighter shading. Dyads are coded as jointly democratic if both states have a Polity score greater or equal to six. All power-ratios are weighted by the inverse of the logged dyadic distance.

# C Coverage of CINC variables

In the main manuscript, we assess convergent validity by comparing a country's share of global SDP to several component variables from CINC. Convergent validity is defined as"the degree to which the operationalization is similar to (converges on) other operationalizations that it theoretically should be similar to." (Trochim and Donelly, 2008)

CINC's restrictive approach toward including countries as members of the international system leads to distortions in the estimates of power. For example, based on the Correlates of War (COW) classification, China does not become a member of the international system of states until 1860, while Gleditsch and Ward code it as a system member since 1816. As a result, CINC population totals are likely undercounting global population and therefore inflating other countries' relative population figures prior to 1860. Figure 5 illustrates the effect that China has on the total CINC score. Plotted in Figure 5 is the annual correlation between the original CINC score and a re-computed CINC score that drops China from the global sums of the component variables. Before China enters the COW system of states (and CINC), the two correlate perfectly, because China is included in neither of the series between 1816 and 1859. When China enters the National Military Capabilities data (Greig and Enterline, 2017) in 1960, the correlation drops to approximately 0.9976, mostly because China has such a large total population relative to other countries (see below).

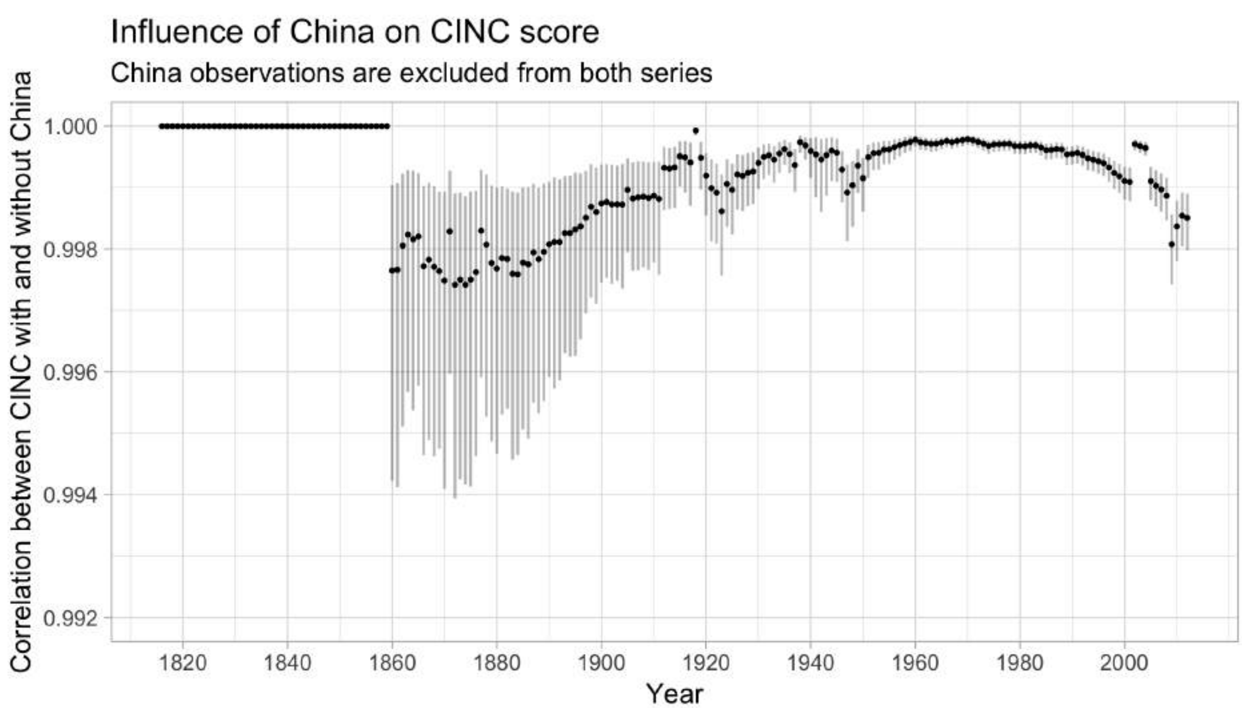
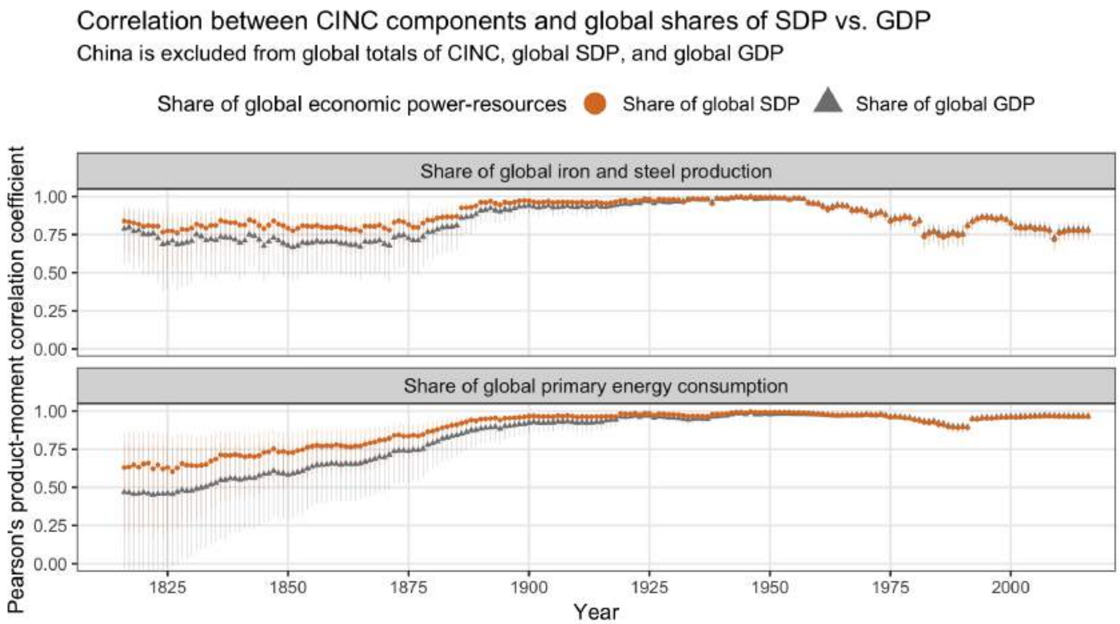



Figure 5: The plot shows the annual correlation with 95% confidence intervals between the original CINC score and a re-computation of CINC that drops China from the global sums of iron and steel production, primary energy consumption, total population, urban population, military expenditure, and military personnel. The annual observation for China is dropped from both series.

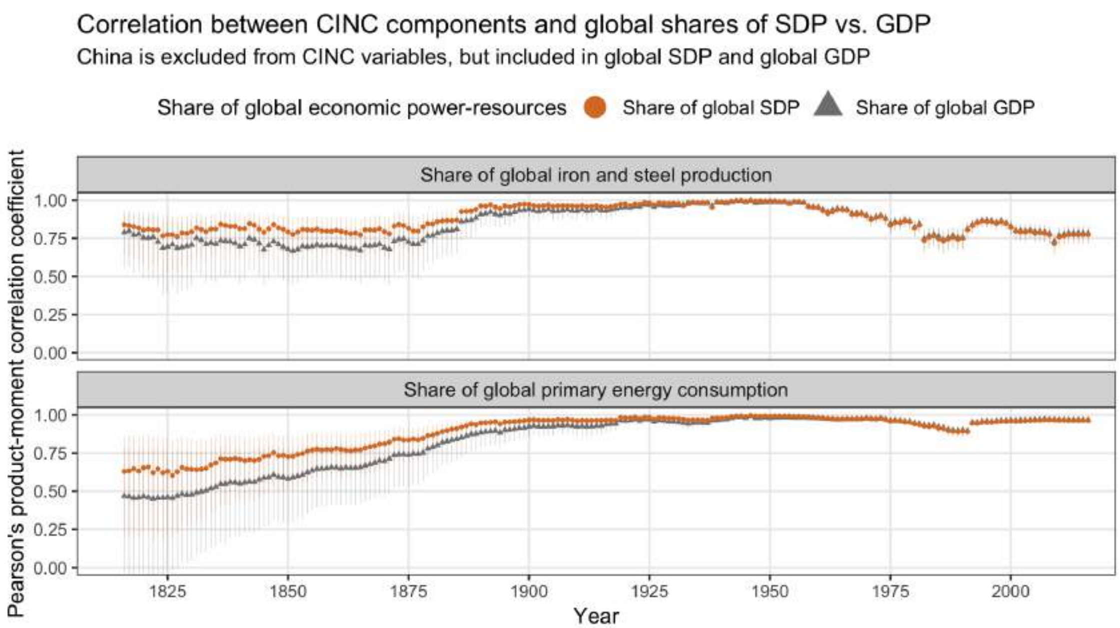
Our new measurement approach below makes population estimates available for a larger set of countries in the pre-industrial period and correct part of the bias resulting from the exclusion of units in CINC. For example, while CINC codes China as having 47% of the global population in 1860, our data code the population share to be 31%. The exclusion of China in the CINC scores from 1816 to 1859 also affects the population shares of other countries. The United States drops from having 8.3% of global population in 1859 to 3.9% in 1960 in CINC; our estimates are 2.6% and 2.6%, respectively. We this use our revised series of population data to compute a country's share of global population (Figure 3 in the main manuscript). These population estimates are available for a larger set of country-year units. They contain data on China and Japan (that are missing in CINC) prior to 1860.

Figure 6 below re-plots the upper panel of Figure 3 from the main manuscript. In panel (a), we exclude China observations from the global sums of iron and steel production, primary energy consumption, SDP, and GDP, respectively. In panel (b), we exclude China observations from the global sums of iron and steel production and primary energy consumption, but keep China observations in the computation of global SDP and GDP. The graphs are virtually identical. The plots demonstrate that the drastic drop in the correlation between the share of global GDP and the CINC component variables is caused by the exclusion of China from CINC prior to 1860, not by an in issue our GDP or SDP estimates.

Note that in Figure 3 in the main manuscript, the drastic drop in the annual correlation between CINC component variables and a country's share of global GDP is not replicated in the correlation with a country's share of global GDP because China does not contribute much to global surplus until the post-WWII period. Based on our estimates, China starts to consistently have GDP income that exceeds subsistence needs in 1964.




(a) Excluding China from all series




(b) Excluding China from CINC, but not SDP or GDP series

Figure 6: The plots display yearly correlation coefficients with 95% confidence intervals. In each of the panels, we assess the degree to which SDP (orange) and GDP (grey) correlate with the iron and steel production and primary energy consumption variables of CINC.

# D Comparing SDP with GDP and GDP per capita

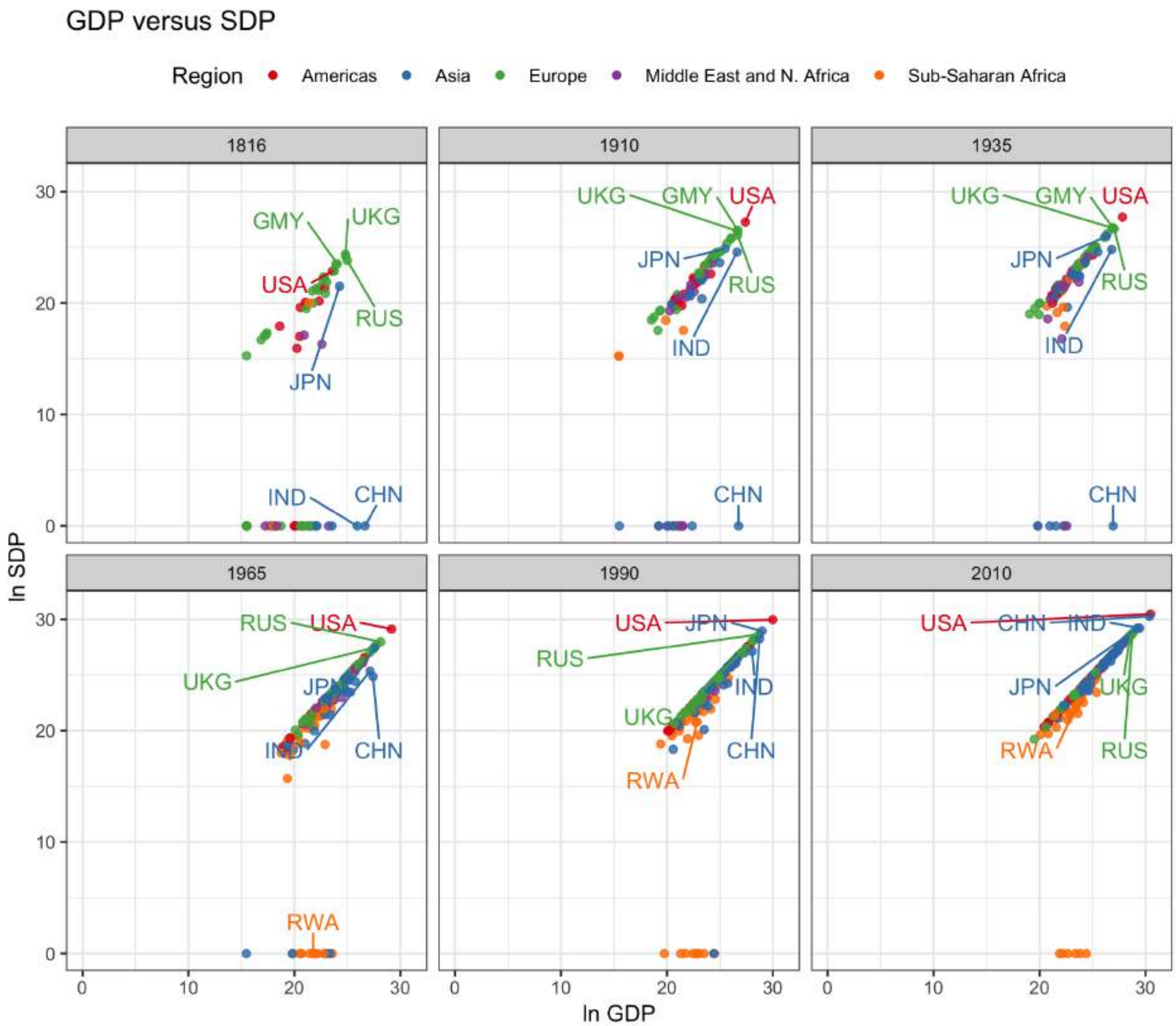



Figure 7: The graph plots the natural logarithm of GDP against the natural logarithm of SDP for select years. The plot shows that as time progresses and countries develop, SDP and GDP correlate highly. An exception are least developed countries, mostly in Sub-Saharan counties, who do not have a positive surplus in 2010. The SDP measure is based on a $3 per day subsistence threshold and is truncated to 1 for countries with no surplus resources in order to allow for a transformation via the natural logarithm.

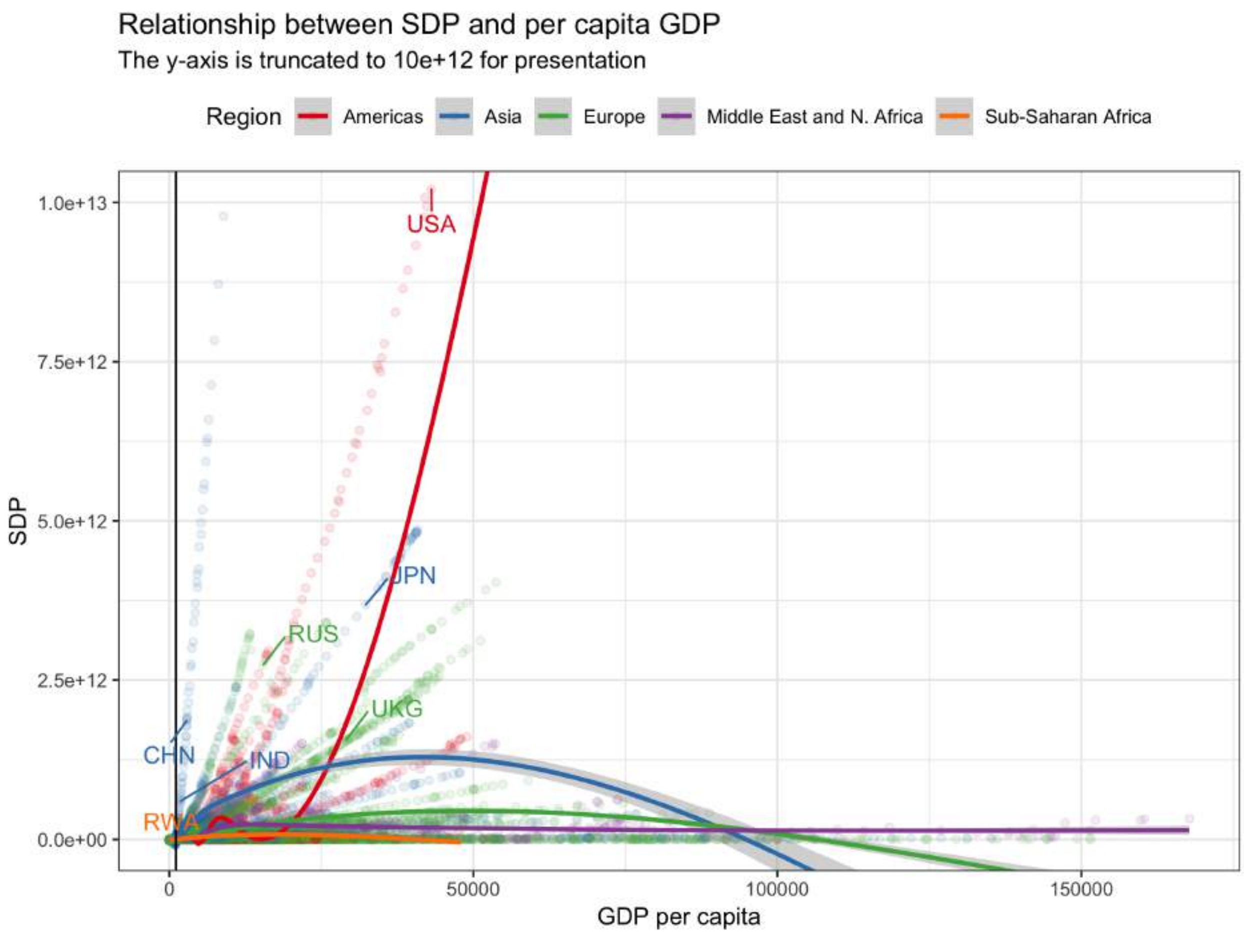



Figure 8: The graph plots GDP per capita against SDP across all country-years in the sample. The direction of the relationship between the two variables varies by world region. It is strong and positive for the most developed countries in the Americas (which includes the US and Canada). For most other world regions it is curvilinear or even flat. The linear patterns of dots show individual countries' trajectories over time. As countries develop, higher levels of GDP per capita are associated with higher levels of SDP. The slopes of the trajectory show that the strength of the relationship between GDP per capita and SDP varies considerably between countries. Labeled are observations for select countries in 1990. The relationship is very strong for large and rapidly developing countries like India and China, and weaker for developed countries like the United Kingdom. The SDP measure is based on a $3 per day subsistence threshold.

# E Measuring potential threat in the strategic environment

In the main manuscript, we defined three component variables of potential threat that capture information about dyads, which we review here. The variables exists for each unit, $i = 1, \ldots, N$ across each time period $t = 1, \ldots, T$. For each country-year variable, we make use of information about each of the dyadic relationships between state $i$ and the other $j$ states in the international system each year, which $j = 1, \ldots, J$ indexes the other states in relationship with state $i$. We consider three types of relationships between state $i$ and state $j$ all of which are bounded between 0 and 1:

1. Relative power ratio in terms of the difference in power-resources between state $i$ and state $j$ in year $t$ (i.e., is the opponent state $j$ is relatively larger or smaller than state $i$).
2. Loss of Strength Gradient over geographic distance between state $i$ and state $j$ in year $t$.
3. Preference compatibility between state $i$ and state $j$ in year $t$.

**Relative power ratio** for state $i$ with opponent state $j$ is defined based on the ratio of the power-resources as measured by the SDP of the opponent state $j$ as a proportion of the sum of the SDP values for both state $i$ and the opponent state $j$. SDP is measured as a function of each state's $GDP_{it}$, $Population_{it}$, and the subsistence level threshold $\tau$ which we set to either \$3, \$2, or \$1 dollars per day as defined in equation 4 above.[3]

The relative power ratio between two states is measured using the estimate of surplus domestic product $SDP_{it}$ for state $i$ and the $SDP_{jt}$ for the opponent state $j$ as

$$r_{ijt} = \frac{SDP_{jt}}{(SDP_{jt} + SDP_{it})}$$

This quantity falls on the unit interval $[0, 1]$ such that

$$r_{ijt} = \begin{cases} (0.5, 1] & if\ SDP_{it} < SDP_{jt} \\ 0.5 & if\ SDP_{it} = SDP_{jt} \\ [0, 0.5) & if\ SDP_{it} > SDP_{jt} \end{cases}$$

These relative power ratios capture the intuition that powerful states will find less powerful countries less threatening because they are the weaker state in the $ij$ pairing. The most powerful state in the system will fear all other countries less than those countries fear it. The most powerful state's relative power ratio will be close to 0. The least powerful state's relative power ratios will be close to 1. If two states have equal power, they will each find the other equally threatening and the relative power ratio for two equal states is 0.5.

[3] We are currently working on collecting additional data that will help us model this threshold parameter as a latent variable.

We weight these relative power ratios using two additional relational features between pairs of states: the preference relationship between states (preference compatibility) and the relative position of a state within the geographic arena (loss of strength gradient).

**Preference Compatibility:** Only certain powerful states are potentially threatening to others and observable indicators of shared preferences can help to identify these relationships. When preferences between pairs of states are compatible, the probability of conflict between the two is reduced and, as such, should minimize the importance of power-resource differences between the two states.

Though we consider many alternative indicators of preference compatibility in a related project,[4] in this paper, we focus on insights from the democratic peace literature to assess degree to which two state have compatible preferences. We assume that all states are potentially threatening to one another, unless they are both democracies. States with democratic institutions have more compatible preferences and are therefore not as threatening to one another. We make no claims regarding whether it is democratic institutions themselves or some other variable that co-varies with democracy that causes states to have more compatible preferences. Thus, we make a descriptive, rather than causal claim, when arguing that democratic states, and only democratic states, do not find each other democracies threatening. Thus, we assume that democracies find autocracies threatening, and autocracies find all states threatening regardless of their regime type.

We use utilize this assumption regarding which states will find each other threatening, to define a preference compatibility measure that we use to down-weight each power-resource ratio $r_{ijt}$. Preference compatibility is defined as $p_{ijt}$, which is a measure of the shared preferences of state $i$ and state $j$ in year $t$. This quantity falls on the unit interval $[0, 1]$. For some of the preference indicators we consider, this variable takes only integer values $\{0, 1\}$. Specifically, the value is 0 if state $i$ and state $j$ both have compatible interests in year $t$ based on the Polity2 or Boix et al. democracy variables. $p_{ijt}$ is otherwise coded as 1 when this is not the case. A coding of 1 captures incompatible relationships, which could potentially be threatening depending on the value of $r_{ijt}$. Using one of two binary democracy variables, we define the preference compatibility between two states as

$$p_{ijt} = \begin{cases} 0 & \text{if } i \text{ and } j \text{ jointly democratic} \\ 1 & \text{otherwise} \end{cases}$$

For the continuous measure of preference compatibility, we use the Unified Democracy Scale for each state to define preferences as

$$p_{ijt} = \Phi(-UDS_{it}) \times \Phi(-UDS_{jt}).$$

Thus, if a pair of states does not have compatible preferences, then the relative power-resource measure

[4]Further below, Figures 26 and 27 in this online appendix illustrate that our results are largely robust to indicators of preference compatibility that do not rely on joint democracy, such as rivalry, alliances, bilateral trade relationships, or United Nations General Assembly voting.

is not changed. If a pair of states has compatible preferences, then the power-resource ratio is reduced to 0—effectively making states both non-threatening to one another.

**Loss of Strength Gradient over geographic distance:** The costs associated with conflict and arming are increasing in the distance over which power must projected to state $i$ by an opponent state $j$. We therefore assume that the Loss of Strength Gradient, which increases over distance, reduces the level of threat between two states. Contiguous or geographically proximate states should be more influential or potentially threatening than states that are far away because the loss of strength gradient results in power-resources dissipating over distance (Markowitz and Fariss, 2013; Gleditsch and Ward, 2001; Boulding, 1962).

Loss of strength gradient over geographic distance is defined as $d_{ijt}$, which is the distance between the capital city of country $i$ and the capital city of neighbor $j$ in year $t$. $d_{ijt}$ is defined for each country-year pair in each year using the longitude and latitude coordinates for each state's capital city

$$d_{ijt} = acos(sin(lat_{it}) * sin(lat_{jt}) + cos(lat_{it}) * cos(lat_{jt}) * cos(lon_{it} - lon_{jt})) * radius$$

Where $d_{ijt}$ is the distance between state $i$'s capital city and state $j$'s capital city. $lat_i$, $lat_j$, $lon_i$, $lon_j$, are the latitude and longitude locations for state $i$ and state $j$. These values vary little over time but we calculate $d_{ijt}$ for each year $t$. We transform the distance values into a proportion $w_{ijt}$, so that it falls on the unit interval $[0, 1]$. This captures the intuition that states that are geographically proximate (short distance between $i$ and $j$) should have more influential relationships than states that are geographically distant from one another. The loss of strength gradient increases the costs associated with projecting power. In many existing empirical applications, the transformation of distance to the unit interval is accomplished using either inverse distance or the inverse natural logarithm of this quantity. For the inverse natural logarithm, this is defined as:

$$w_{ijt} = \frac{1}{\ln(d_{ijt})}.$$

In words, $w_{ijt}$ is the the inverse of the natural log of distance $d_{ijt}$ in $km$ between state $i$ and state $j$ in year $t$. The measure captures the intuition that neighbors, which are geographically proximate (close neighbors), are more influential on the behavior of country $i$ than neighbors that are far away. Figure 9 provides visual examples of the distribution of this component measure.

**Potential threat** is defined as the total of each of these weighted relative power ratios for country $i$ in year $t$, based on state $i$'s relationship with all other $j$ states in the international system in each year. It is formally defined as

$$\text{Potential threat}_{it} = \sum_{j \in J} \left[ r_{ijt} \times w_{ijt} \times p_{ijt} \right].$$

Table 1 provides a summarization of each component part of Potential threat$_{it}$ for the economic resource-based version. Table 2 provides an analogous specification for the population-based potential threat measure. Figure 10 provides a step by step illustration of the construction of this measure.

| Concept | Measurement |
|---|---|
| Relative power ratio | $r_{ijt} = \frac{SDP_{jt}}{SDP_{jt}+SDP_{it}}$ |
| Loss of strength gradient | $w_{ijt} = \frac{1}{\ln(d_{ij})}$ |
| Preference compatibility (binary) | $p_{ijt} = \begin{cases} 0 & \text{if } i \text{ and } j \text{ jointly democratic} \\ 1 & \text{otherwise} \end{cases}$ |
| Preference compatibility (continuous) | $p_{ijt} = \Phi(-UDS_{it}) \times \Phi(-UDS_{jt})$ |
| Total potential threat (economic) | Potential threat$_{it} = \sum_{j \in J} [r_{ijt} \times w_{ijt} \times p_{ijt}]$ |

Table 1: Concepts and operational definitions of each of the component parts of the country-year potential threat measure based on economic resources.

| Concept | Measurement |
|---|---|
| Relative power ratio | $r_{ijt} = \frac{\text{Population}_{jt}}{\text{Population}_{jt}+\text{Population}_{it}}$ |
| Loss of strength gradient | $w_{ijt} = \frac{1}{\ln(d_{ij})}$ |
| Preference compatibility (binary) | $p_{ijt} = \begin{cases} 0 & \text{if } i \text{ and } j \text{ jointly democratic} \\ 1 & \text{otherwise} \end{cases}$ |
| Preference compatibility (continuous) | $p_{ijt} = \Phi(-UDS_{it}) \times \Phi(-UDS_{jt})$ |
| Total potential threat (population) | Potential threat$_{it} = \sum_{j \in J} [r_{ijt} \times w_{ijt} \times p_{ijt}]$ |

Table 2: Concepts and operational definitions of each of the component parts of the country-year potential threat measure based on economic resources.

We briefly describe the measurement process that generates the total relative power variable for a hypothetical three-state system. Suppose that in the year 1900 there are only three countries in the world: the United Kingdom, Germany, and the United States. The table below shows the computation of the level of potential threat that the United Kingdom faces if its strategic environment consists of only Germany and the United States. In 1900, the United Kingdom is coded as a democracy based on the categorical value of its democracy score based on Polity2. Its SDP was approximately 265 billion in constant 2011 international PPP dollars. In this year, the United Kingdom does not have compatible preferences with then-autocratic Germany, but is jointly democratic with the United States.[5] Based on the binary specification of regime type,

[5] Please note that this is only true for the binary joint democracy measures using the Polity and Boix et al. data. When using

only the relative power-resources of Germany, weighted by distance, contribute to the total level of potential threat, which is the sum of the relative power ratios that the United Kingdom faces in this three-state international environment.[6] Table 3 below illustrates the computation of the United Kingdom's potential threat in this three-state example. the United Kingdom's potential threat score in this hypothetical three-state international system is 0.07.

| | **Relative power-resources** | **Preference Compatibility** | **Loss of Strength Gradient** | **Weighted relative power-resources** |
|---|---|---|---|---|
| Germany | $\frac{234}{234+265} = 0.47$ | 1 | $\frac{1}{\ln(916)} = 0.15$ | $0.47 \times 1 \times 0.15 = 0.07$ |
| United States | $\frac{457}{457+265} = 0.63$ | 0 | $\frac{1}{\ln(5954)} = 0.12$ | $0.63 \times 0 \times 0.11 = 0$ |

Table 3: Hypothetical example of a three-state system. The example demonstrates how each component part of the potential threat variable is combined into the final value for this country-year variable.

the continuous joint democracy measure based on the UDS scale, the distance-weighted power of the United States, relative to the United Kingdom, would contribute to the total potential threat faced by the United Kingdom. However, the United States' contribution would be very small because the preference-weight will be close to zero

[6]The maximum distance between capital cities in our data is approximately 19949km. The distance London–Berlin is approximately 916km; London–Washington D.C. approximately 5954km.

## E.1 Geographic proximity and the loss of strength gradient

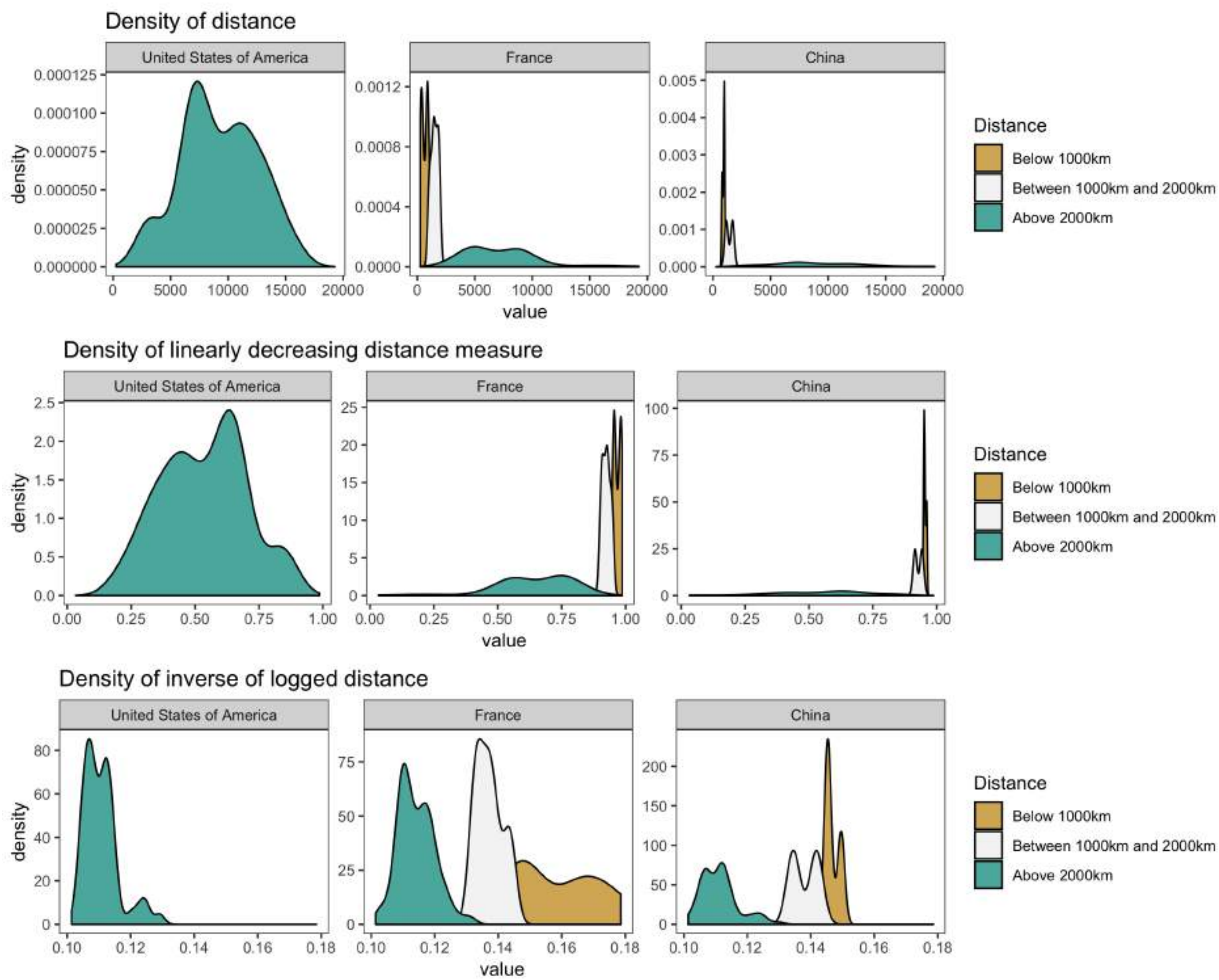



Figure 9: The graph compares the binning of countries by a) distance, b) a linear transformation of distance computed as $\frac{\max(\text{Distance})-\text{Distance}}{\max(\text{Distance})}$, and c) the inverse of the logged distance from the perspective of the United States, France, and China. Distance is measured as the great circle distance between capital cities in kilometers. The maximum distance in our data is 19911.7km. In the main specification of our potential threat measure, we use the transformed distance between capital cities as a continuous measure of geographic proximity, not the binned version. All else equal, states with higher values of the transformed distance (i.e. states that are closer) will be more threatening than states with lower values of the transformed distance (states that are farther away). For the purpose of illustration, we bin distances into three categories: states with capital cities that are less than 1000km away (Berlin–Paris would be in this category with approximately 880km distance), between 1000km but less than 2000km away (this would capture the distance Berlin–Moscow with approximately 1600km), and capital cities that are more than 2000km away from each other. Purely in terms of geography, the US faces enjoys an incredibly unthreatening neighborhood—with only Ottawa being in the closest category (the distance Washington D.C.–Ottawa is approximately 750km; Havana is 1800km away). In the main specification of our potential threat measure of each state's strategic environment, we weight relative power-resources by the inverse of the logged distance between capital cities.

## E.2 Construction of the potential threat measure

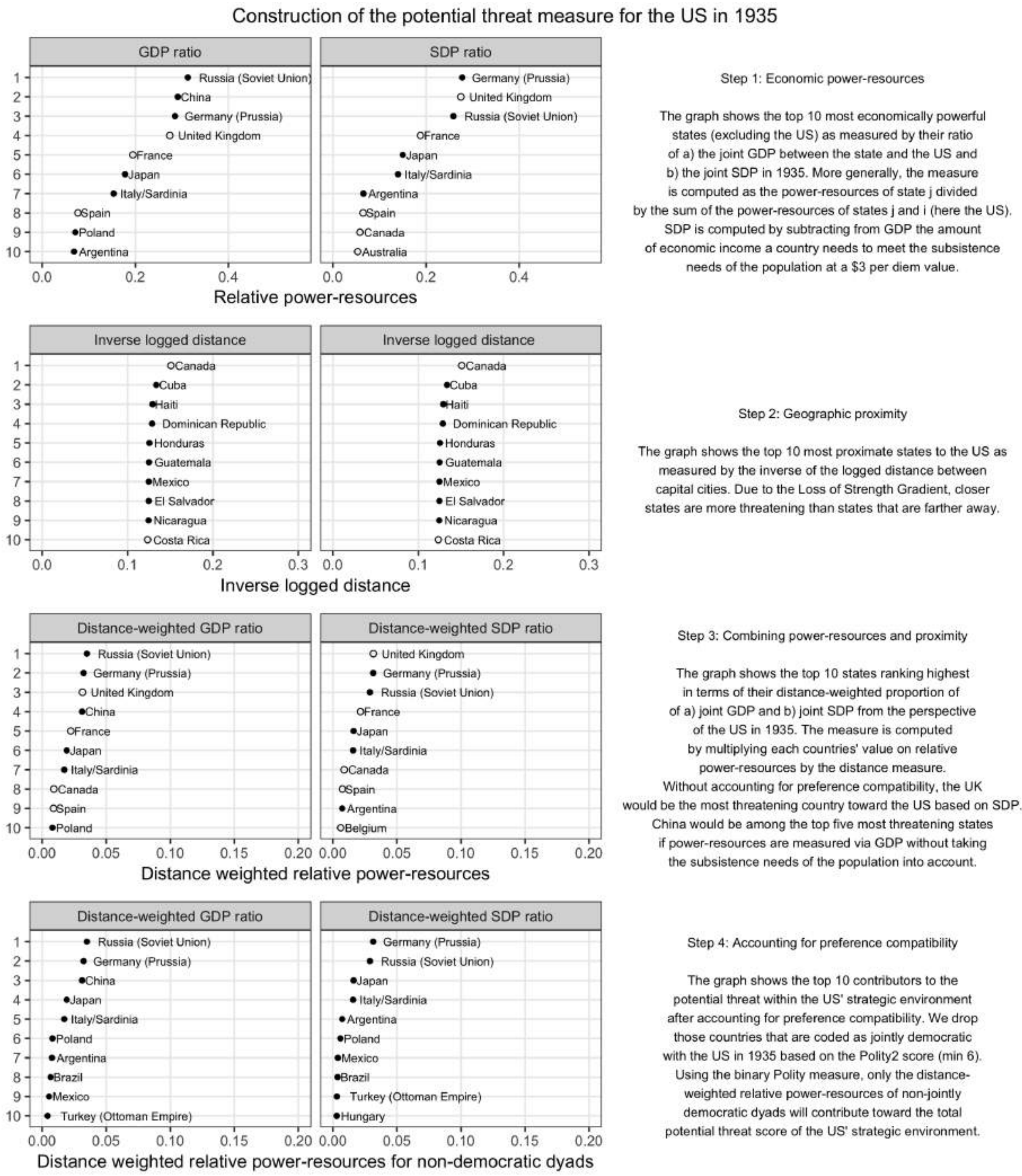



Figure 10: The graph illustrates the construction of the potential threat measure for the US in 1935. The SDP is based on a $3 per diem subsistence value.

## E.3 Correlation between alternative potential threat measures

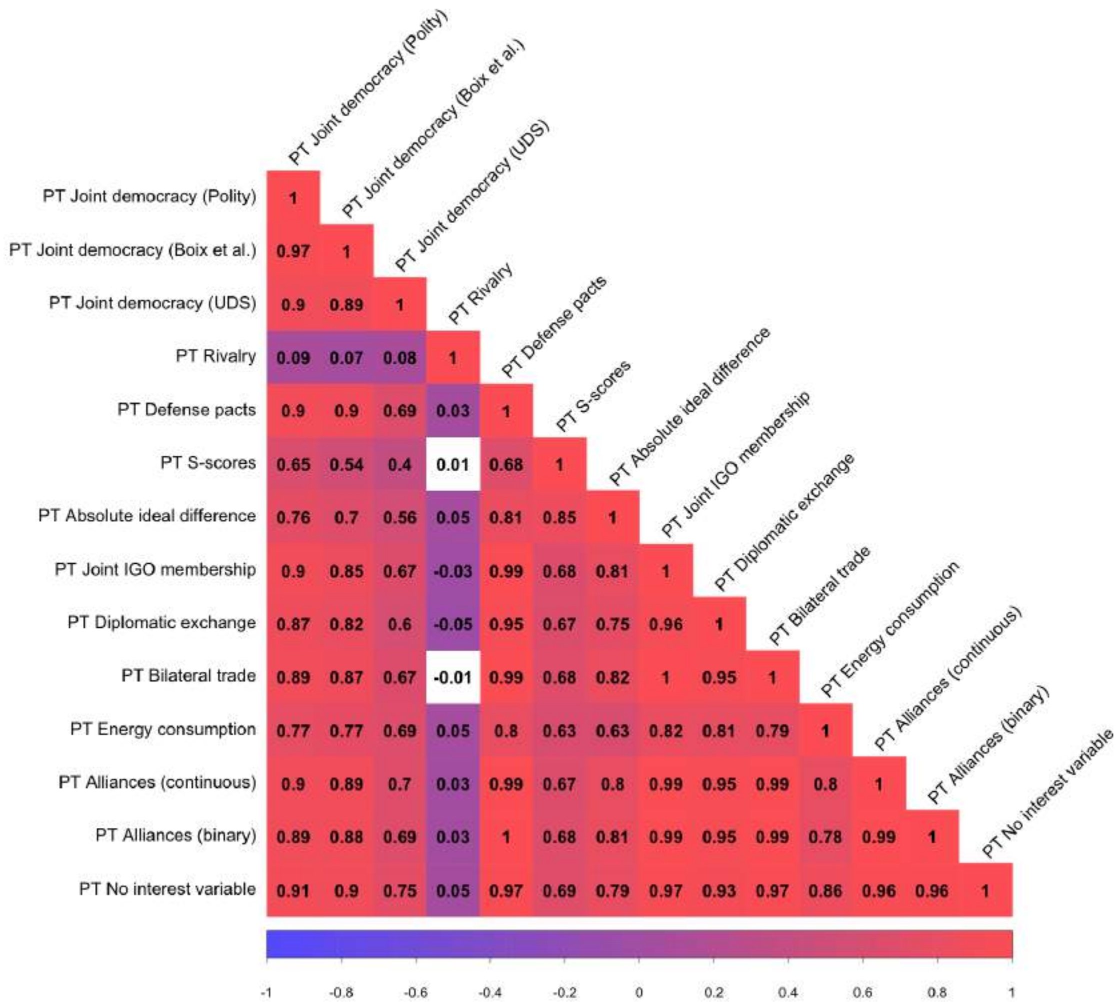



Figure 11: Correlation plot for alternative potential threat measures, using SDP ($3 subsistence level) to measure economic resources. Colored cells denote values that are significant at the minimum 5% level of significance. The plot demonstrates that all potential threat variables but one are statistically significantly positively correlated with one another. We compute potential threat measures for 13 alternative indicators of dyadic preference compatibility: joint democracy using data from Polity (Marshall et al., 2016), Boix et al. (2013), or the Unified Democracy Scores (Pemstein et al., 2010) data; rivalry from the RIV5.10 Rivalry Data Set (Klein et al., 2006); alliances using the defense alliance data from the Correlates of War (COW) Formal Interstate Alliance Dataset, 1816-2012, v4.1 (Gibler, 2009), as well as a binary and continuous alliance measure (Cohen's $\kappa$) from Alliance Treaty Obligations and Provisions (ATOP) data v4.01 (Leeds et al., 2002); United Nations General Assembly Voting Data affinity s-scores and absolute ideal difference measures from Bailey et al. (2017); joint IGO membership from the COW International Governmental Organizations Data Set v2.3 (Pevehouse et al., 2004); the COW Diplomatic Exchange Data Set v2006.1 (Bayer, 2006); bilateral trade from the COW Trade Data Set, v3.0 (Barbieri and Keshk, 2012; Barbieri et al., 2009); and a measure of dyadic per capita energy consumption (Markowitz et al., ming; Greig and Enterline, 2017).

## E.4 Global trends of potential threat

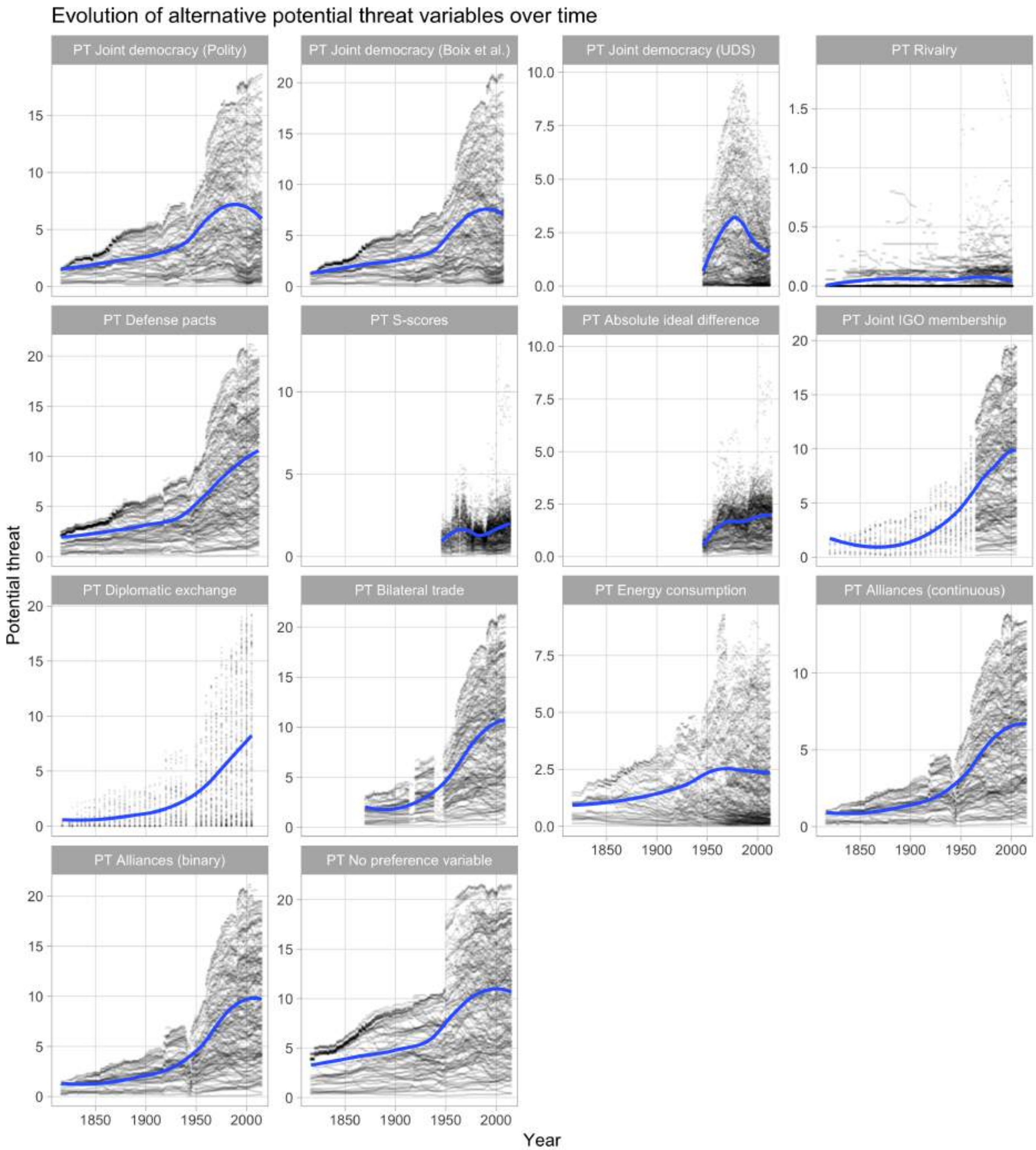



Figure 12: The graphs show the evolution of potential threat over time for alternative indicators of preference compatibility (see the caption of Figure 11 for data sources). Plotted is each country-year observation with the line denoting the loess smoothed trend over time across all countries.

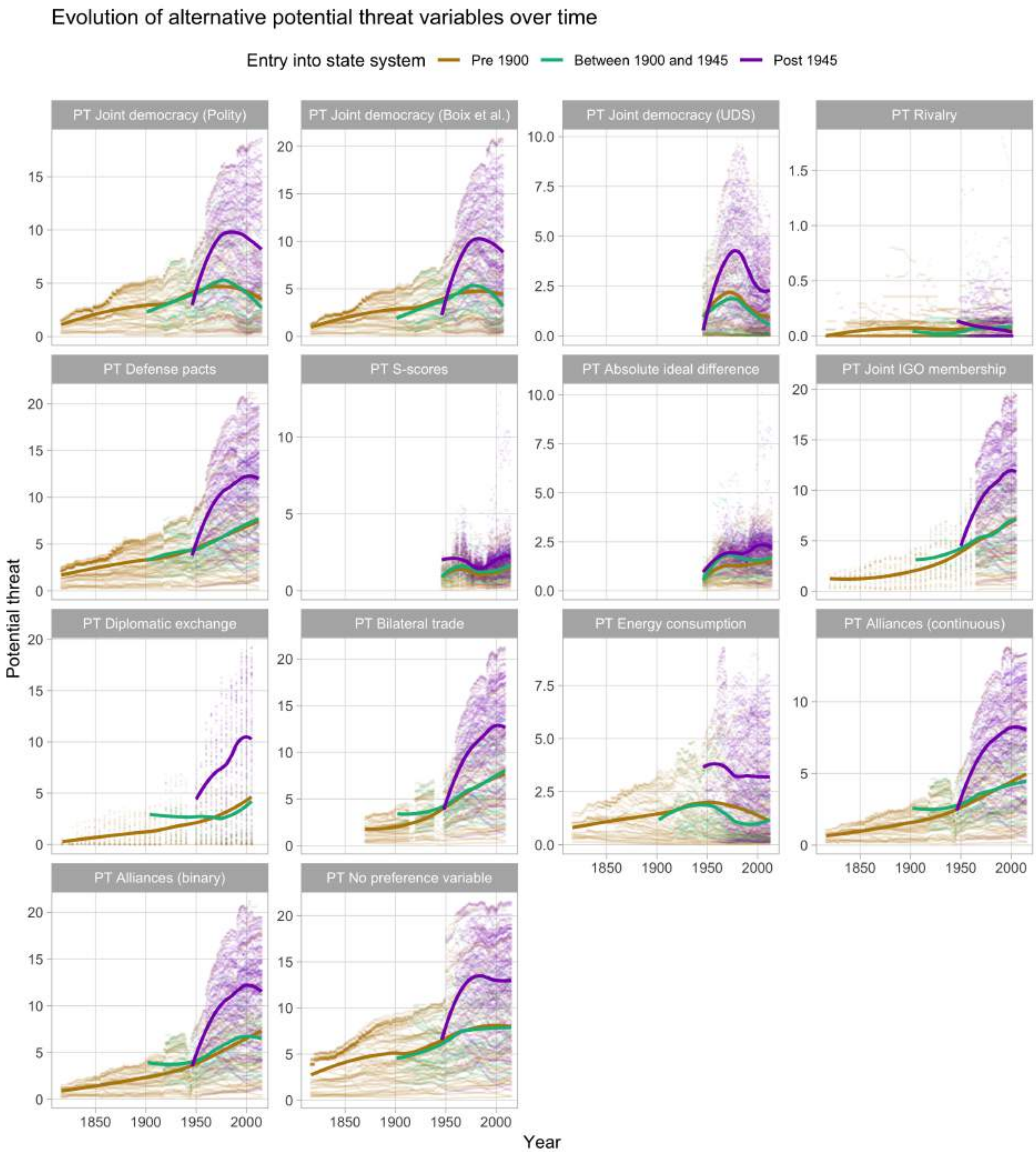



Figure 13: The graphs show the evolution of potential threat over time for alternative indicators of preference compatibility (see the caption of Figure 11 for data sources). Plotted is each country-year observation with the lines denoting the loess smoothed trend over time for three groups of states: a) states that enter the international system before 1900, b) states that enter between 1900 and 1945, and c) states that enter after 1945 based on Gleditsch and Ward (1999a). Globally, we find potential threat to be increasing over time for most indicators of preference compatibility. However, the trends are different for the three groups of countries; with those entering the system after 1945 facing the most threatening geopolitical environment.

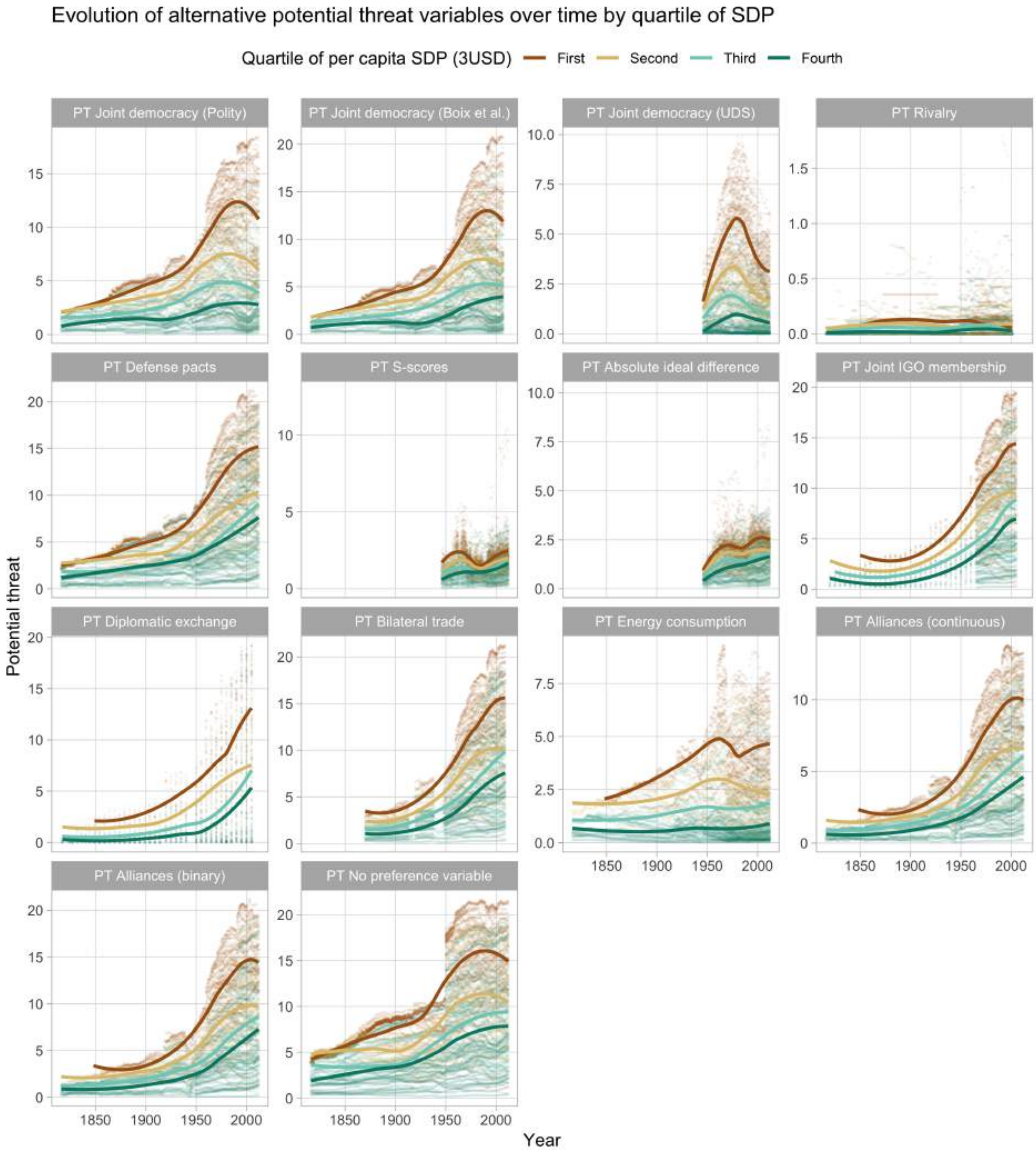



Figure 14: The graphs show the evolution of potential threat over time for each indicator of preference compatibility (see the caption of Figure 11 for data sources). Plotted is each country-year observation with the lines denoting the line of best fit over time for the first, second, third, and fourth quartile of states based on the distribution of the per capita SDP (using a $3 per diem subsistence level).

## E.5 Spatiotemporal variation of potential threat

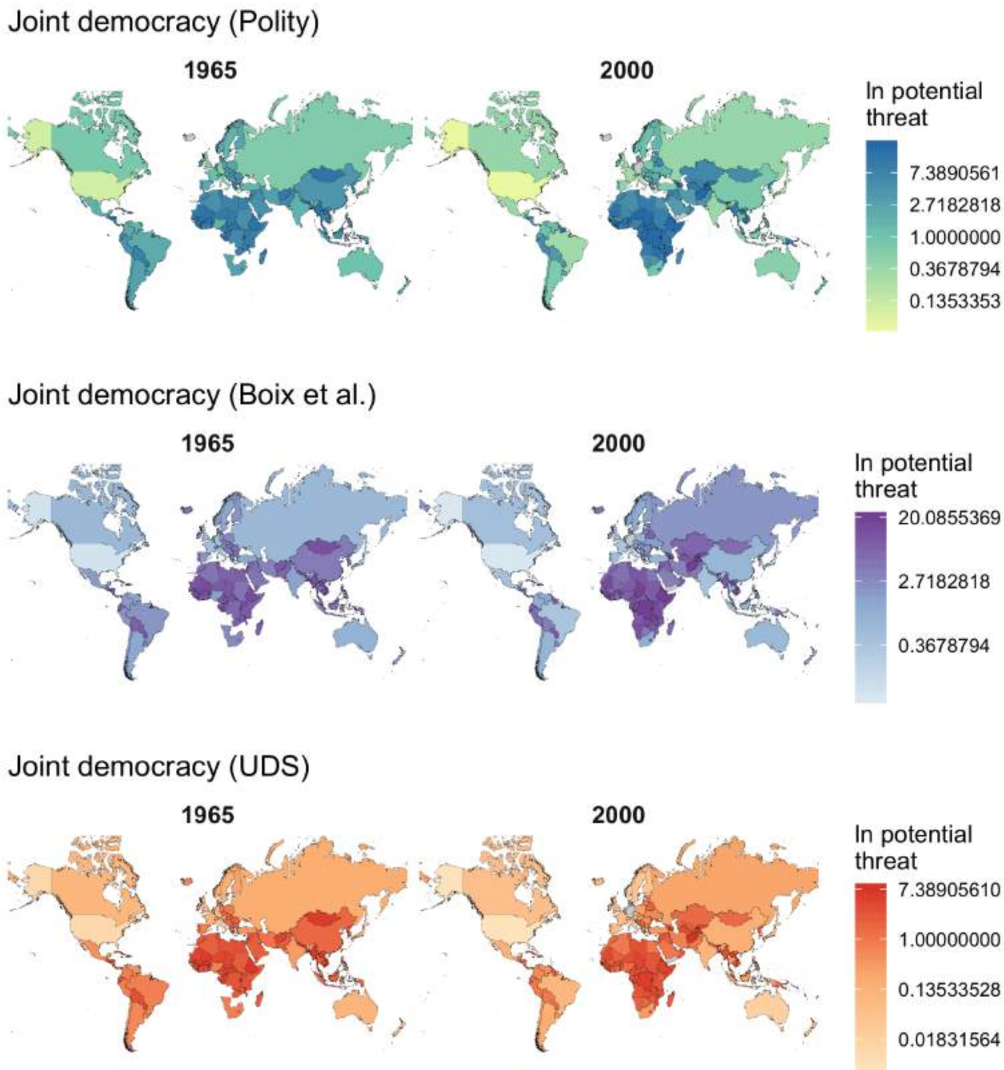



Figure 15: Maps plotting the spatiotemporal distribution of the natural log of the potential threat variable for the years 1965 and 2000. Potential threat is measured through joint democracy based on the Polity, Boix et al., and UDS scores, respectively. Power-resources are measured using the SDP indicator with a $3 per diem subsistence level. Grey shaded areas denote missing values. The maps are based on the borders for 1 January 1965 and 2000, respectively, using data from the `cshapes` library in `R` (Weidmann et al., 2010). The operationalization of each indicator is based on substantive choices of each coding team. Therefore, the coverage does not always perfectly map, either spatially or temporally. For example, some geographical spaces such as Greenland or former colonies in Africa are missing.

## E.6 Top 20 states facing the most threatening strategic environment

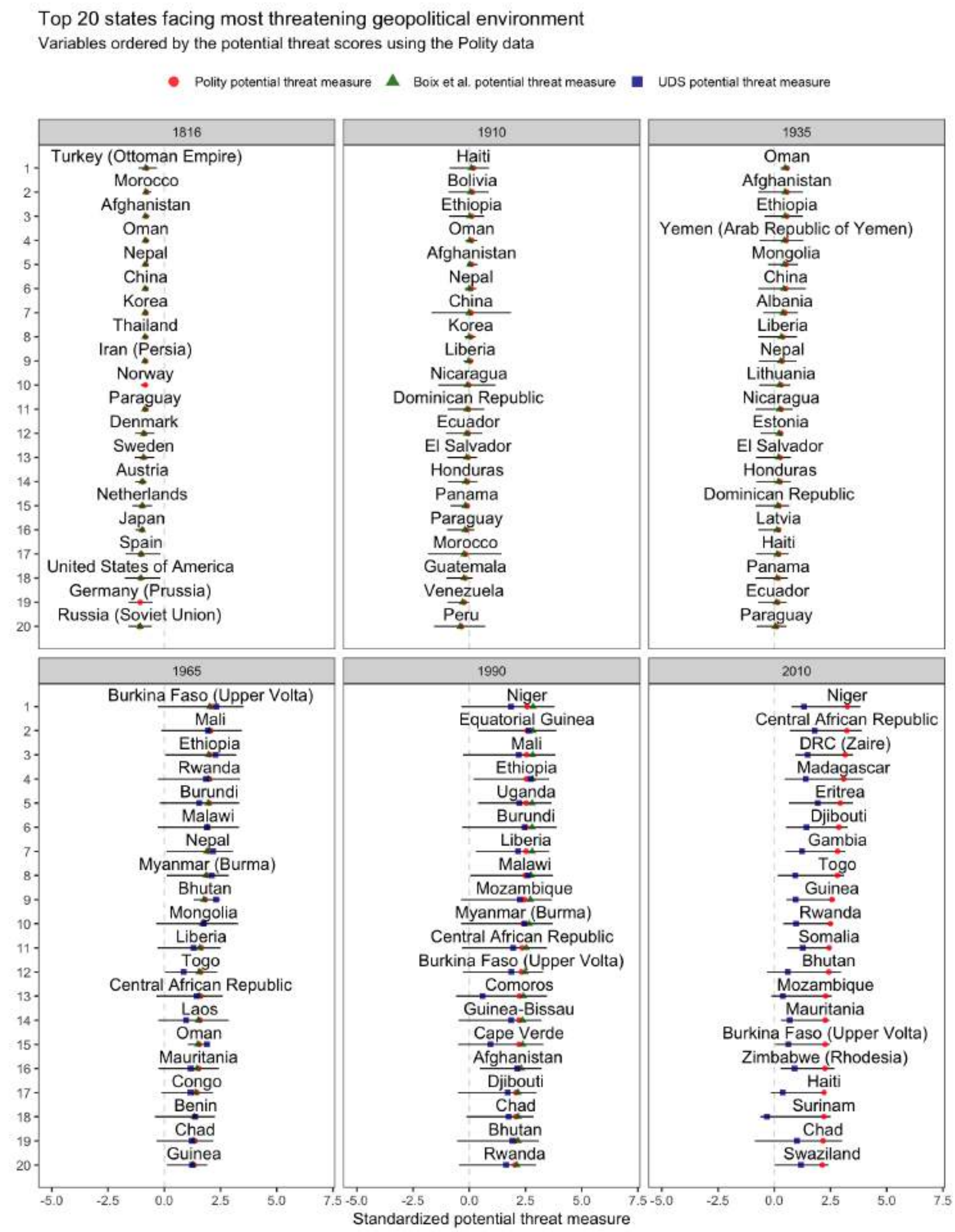



Figure 16: The plot shows the top 20 states facing the most potentially threatening strategic environment in 1816, 1910, 1935, 1965, 1990, and 2010. The red dots show our estimate of the total level of potential threat each country faces when using the Polity2 score to measure preference compatibility; green triangles the Boix et al. estimates, and blue squares the UDS potential threat scores. Countries are ranked based on the potential threat variable that measures preference compatibility via the Polity2 score. Error bars indicate the 95% confidence intervals for the average of *all* alternative potential threat measures (Polity, Boix et al., UDS, rivalry, defense alliances, S-scores, absolute ideal difference, joint IGO membership, diplomatic exchange, bilateral trade, and per capita primary energy consumption). All potential threat variables are standardized; hence, the x-axis measures are expressed in standard deviations.

## E.7 Economic-based potential threat versus population-based potential threat

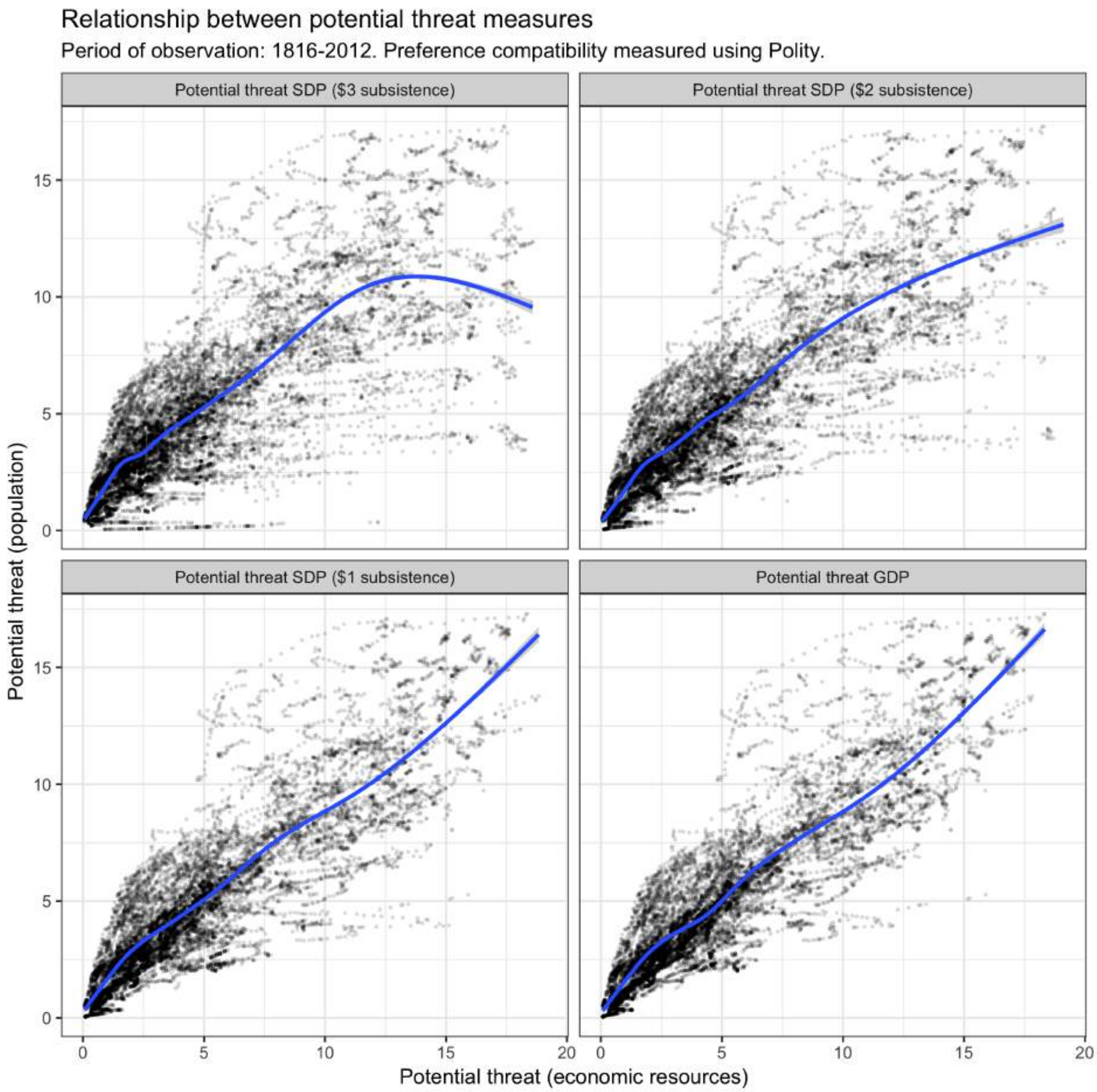



Figure 17: The graphs show the relationship between country-year values of potential threat based on population on the y-axis and potential threat based on economic resources for various subsistence thresholds on the x-axis. As the subsistence threshold increases, the strength of the association between the two alternative potential threat measures decreases.

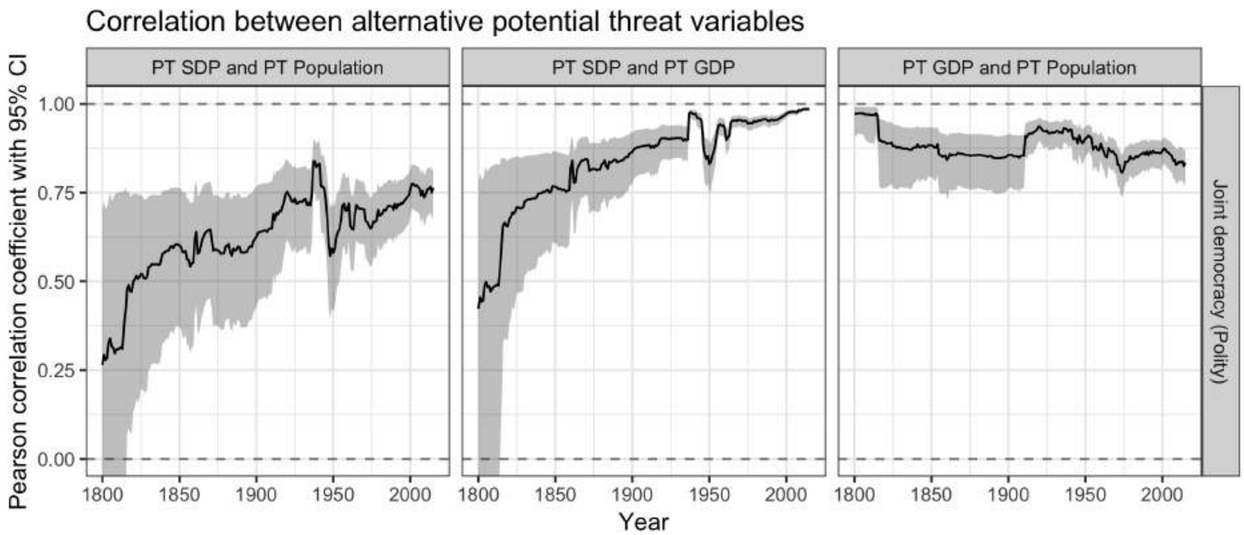



Figure 18: The graph shows the annual correlation between three alternative potential threat (PT) measures over time. We vary how power resources are measured across the three indicators: using GDP, using SDP ($3 subsistence threshold), and using population. For all indicators, preference compatibility is measured via joint democracy (Polity) and power resources are weighted by the inverse of the logged distance between capital cities. The graph shows that the association between the PT measures using SDP and population grows stronger over time. However, even in modern times, the correlation has a maximum of approximately 0.75. This highlights the importance of distinguishing between SDP and population-based potential threat measures, in particular in earlier years of the series. Prior to industrialization, few countries had a GDP that exceeded subsistence income. However, this does not mean that they did not have power resources; they relied more heavily on their population, as opposed to surplus resources, for military capacity. The association between the potential threat measures incorporating SDP and GDP becomes increasingly strong over time and reaches an correlation coefficient of approximately 1 after the year 2000. Conversely, the correlation between GDP and population-based potential threat variables experiences a slight decrease over time, but remains high even in recent years. The graph illustrates that the GDP-based measure is heavily driven by the size of a country's population, and mask the influence of surplus resources that can be invested in arming or power projection.

# F Dependent variables: Military investments

## F.1 Evolution of the military investments over time

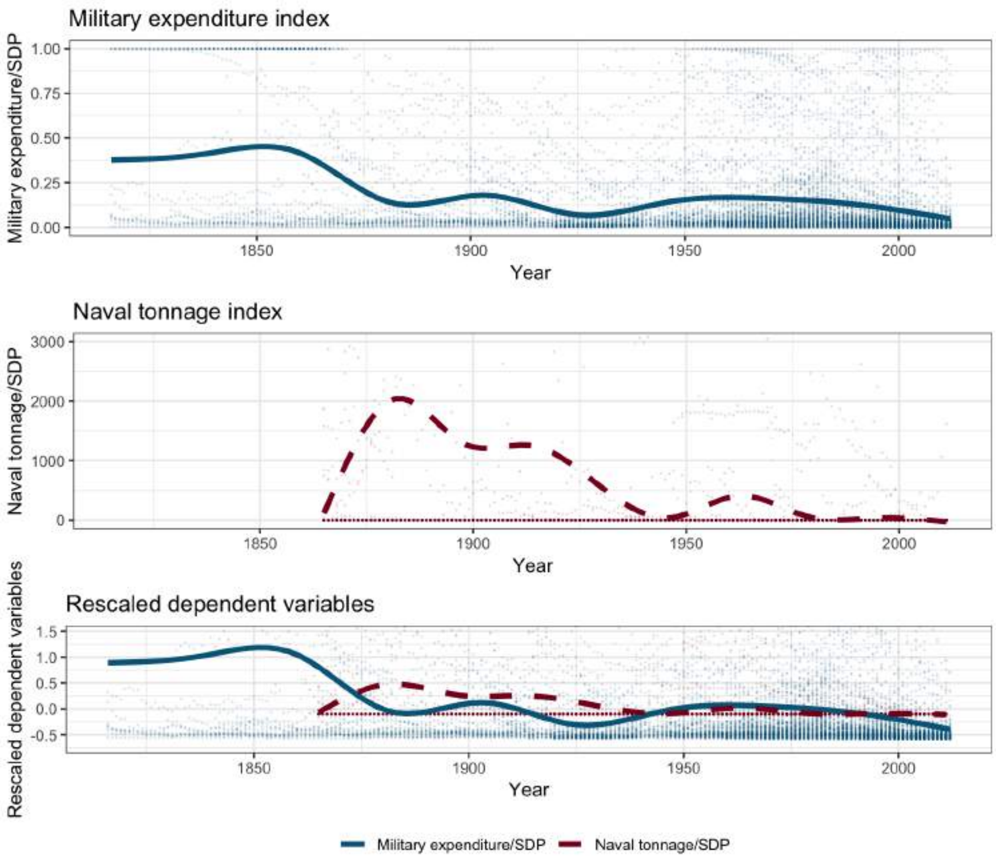



Figure 19: The two upper plots illustrate the temporal evolution of the military expenditure index (military expenditure/SDP in 2011 constant US PPP dollars) and naval tonnage index (naval tonnage/SDP in constant 2011 international PPP dollars), respectively. Dots denote the point values for each country in each year. Lines illustrate the smoothed average over all countries. SDP is computed based on a $3 per diem subsistence level. The lower panel plots standardized values for the military expenditure and naval tonnage indices. Surplus values above zero denote positive deviations from the mean in standard deviations; scores below zero negative deviations. Adopting several dependent variables of military effort allows us to compare how these measures vary over time and to assess the degree to which they respond to the variation in the potential threat of the strategic environment. While these measures of military effort show an overall negative time trend, we observe differences in the evolution of these indicators over time. The smoothed average of military expenditure as percentage of SDP experienced a pronounced decrease since the 1860s. Before 1860, we observe a number of countries spending the entirety of their surplus resources on the military. While global levels of military spending as a percentage of surplus resources in the 20th century are at much lower levels than throughout the 19th and early 20th centuries, the levels of military burden show spikes during the WWI, WWII, and the Cold War. The transition from steam boats to internal combustion engines and gas turbines in the late 19th and early 20th centuries was accompanied by a substantial increase in naval forces—a trend that is not matched by the other dependent variables.

## F.2 Correlations

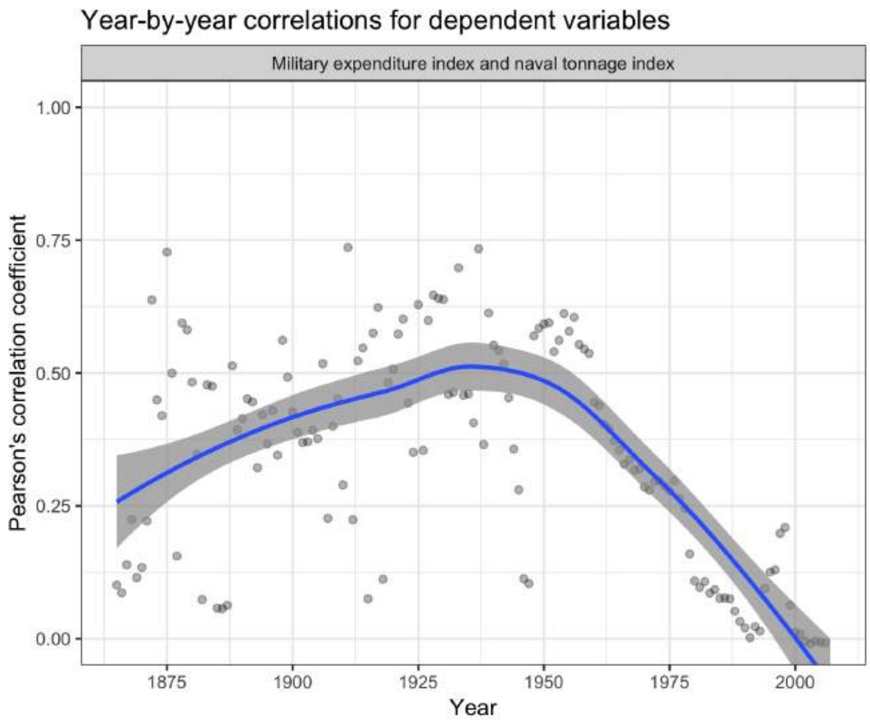



Figure 20: The plot illustrates the correlations between the two alternative dependent variables over time. Points denote the correlation coefficients for each year between 1965 and 2007. Lines represent the Loess smooth over those points.

## F.3 Comparing individual countries over time

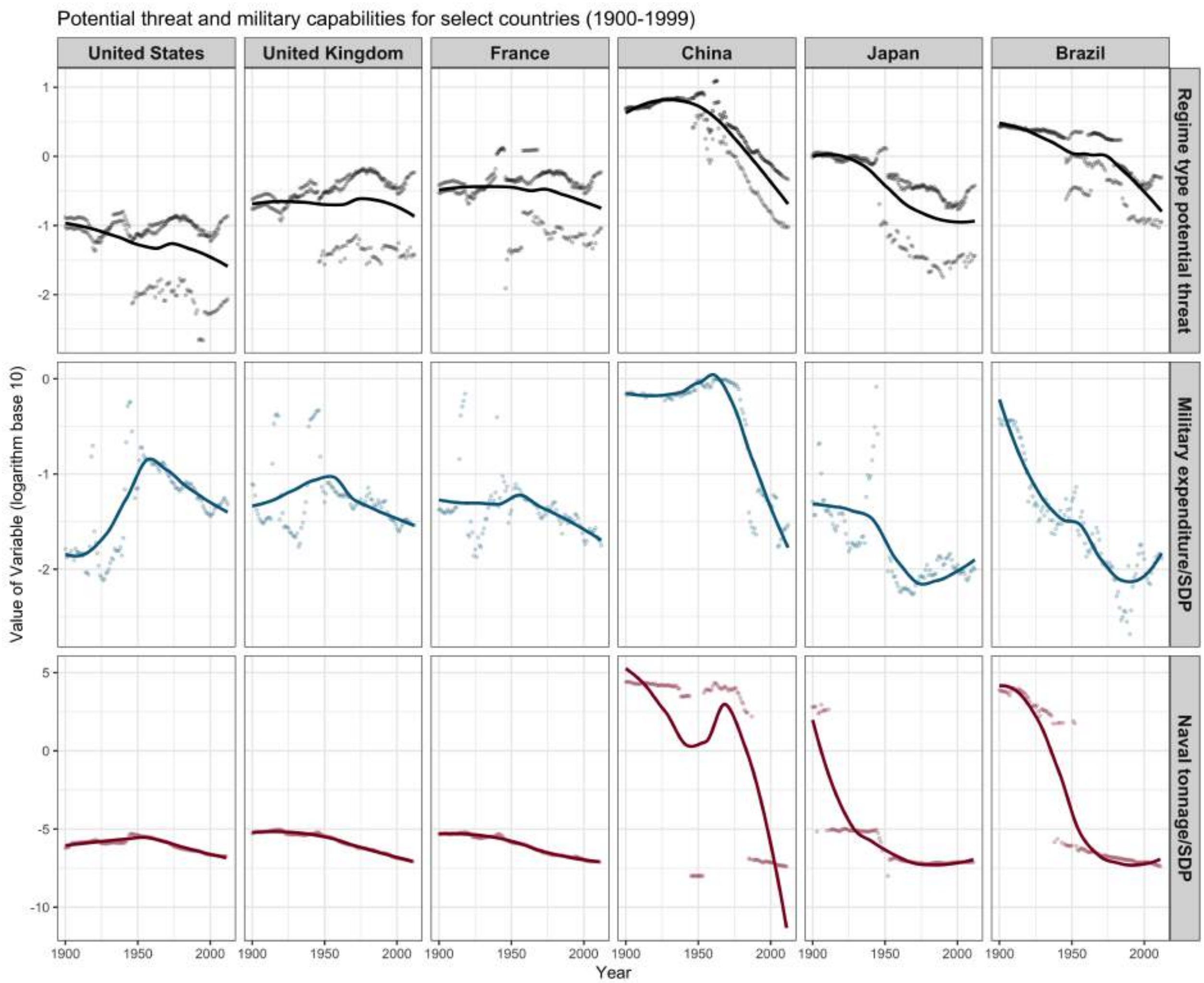



Figure 21: The plot demonstrates the ability of our measurement strategy to obtain scores for the level of potential threat that individual countries face at any given point in time (granted data availability). Plotted in the first row are economic resource-based potential threat scores using alternative regime type indicators to measure preference compatibility for the United States, the United Kingdom, France, Japan, China, and Brazil in the 20th century—the line representing a smoothed trend across all variables. In the rows below, we graph the time trends for the two dependent variables military expenditure as a proportion of SDP and naval tonnage as a proportion of SDP. All variables are shown on a logarithmic scale with base 10. It is striking how closely military burden and power projection follow the sharp decrease in level of potential threat that, for example, Japan experienced after the end of WWII.

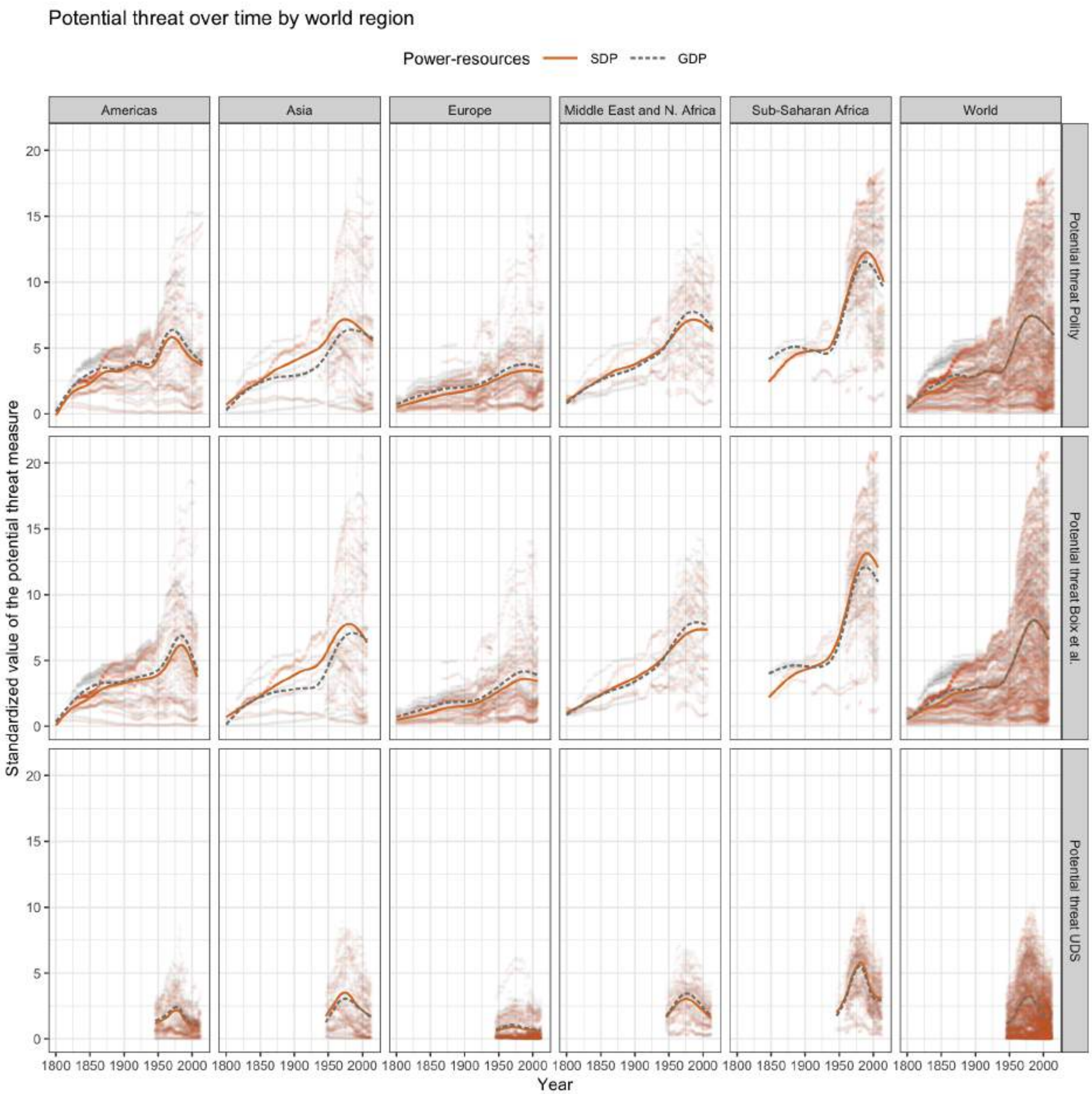



Figure 22: The graph shows the evolution of the potential threat across world regions for alternative measures of preference compatibility. The graph demonstrates that our estimates of potential threat differ depending on the measurement of power-resources. For example, using GDP we would lead us to overestimate the level of potential threat faced by European nations, as well as the Middle East and the Americas in the recent past. Conversely, measuring power-resources via GDP would also lead us to underestimate the level of potential threat faced by states in Asia and Sub-Saharan Africa. The measurement of power-resources via SDP corrects for the fact that poor countries today and most countries in the past had little resources beyond the subsistence needs of their population. Upon accounting for subsistence needs, in the early 19th century, the average global level of potential threat is somewhat lower than GDP would suggest. However, our estimates of the average global level of potential threat for GDP and SDP converge toward late 19th century.

# G Regression models

## G.1 Dropping the population-based potential threat variable

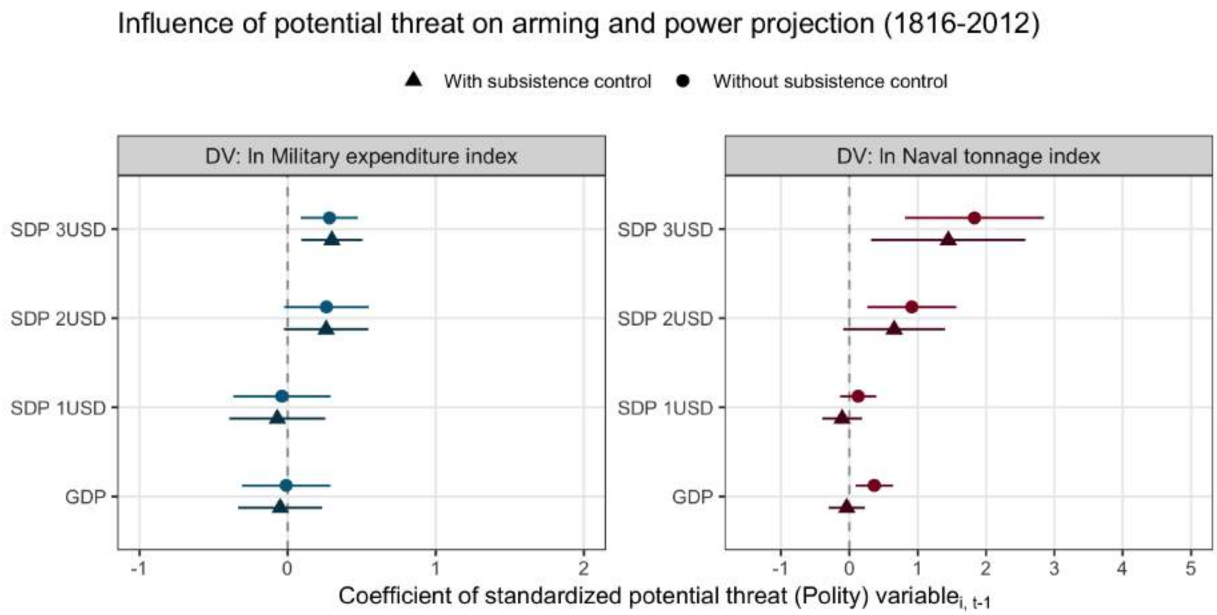



Figure 23: The graph plots the coefficients and 95% confidence intervals of standardized potential threat variables for regression models of two dependent variables — the military expenditure index and naval tonnage index — on potential threat and control variables. Preference compatibility is measured via joint democracy using Polity scores. The loss of strength gradient is measured as the inverse of logged distance. All models include controls for the natural log of income (SDP or GDP) and a country's Polity2 score. We distinguish between models that control for subsistence (or population for the GDP models) and those that do not. Standard errors are clustered by country; right-hand side variables are lagged by one year. Within each panel, rows distinguish between alternative measurements of economic power, that is SDP using a \$3, \$2, and \$1 per diem subsistence level as well as standard GDP.

## G.2 Bivariate regressions

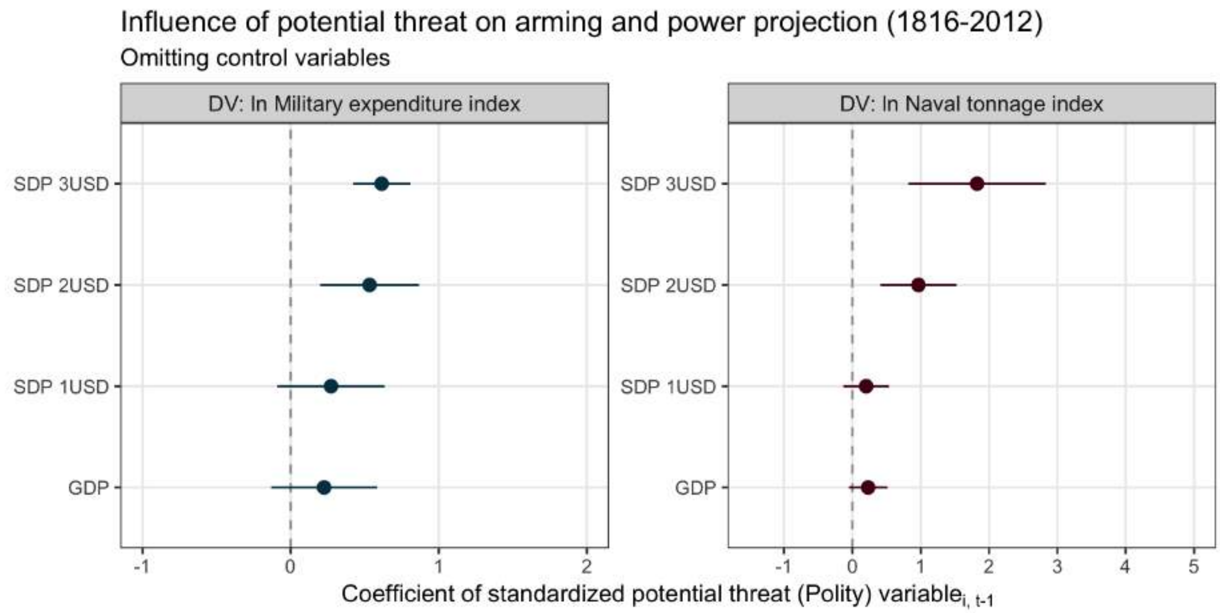



Figure 24: The graph plots the coefficients and 95% confidence intervals of standardized potential threat variables for regression models of two dependent variables — the military expenditure index and naval tonnage index. Preference compatibility is measured via joint democracy using Polity scores. The loss of strength gradient is measured as the inverse of logged distance. Standard errors are clustered by country; right-hand side variables are lagged by one year. Within each panel, rows distinguish between alternative measurements of economic power, that is SDP using a \$3, \$2, and \$1 per diem subsistence level as well as standard GDP.

## G.3 Post-WWII sample

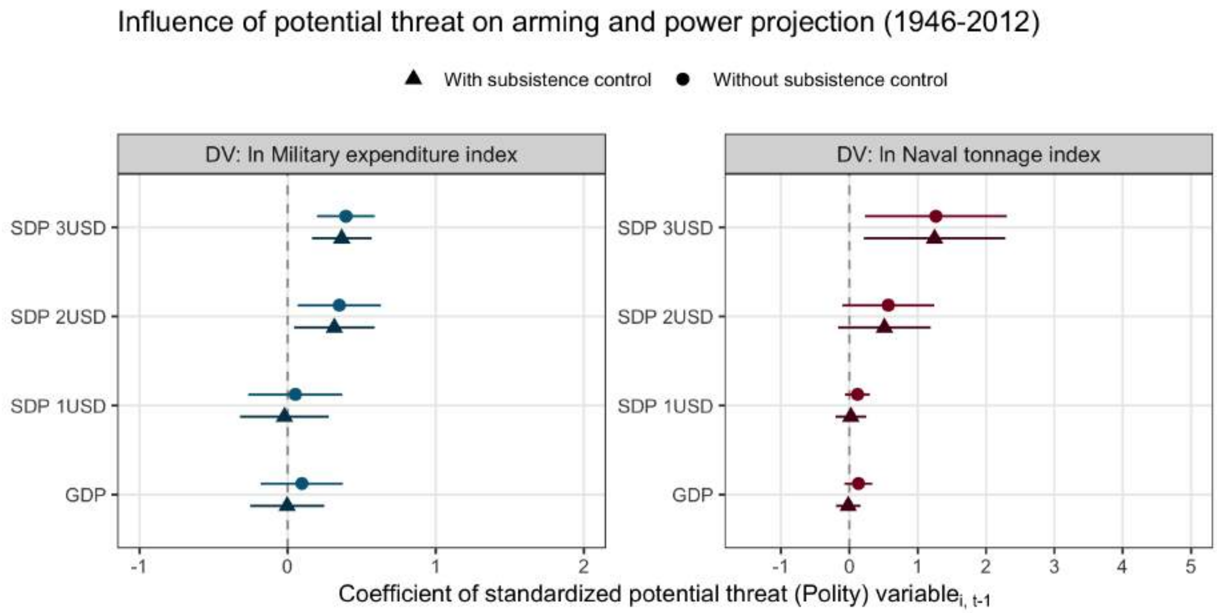



Figure 25: The graph plots the coefficients and 95% confidence intervals of standardized potential threat variables for regression models of two dependent variables — the military expenditure index and naval tonnage index — on potential threat and control variables for a post-WWII sample. Preference compatibility is measured via joint democracy using the Polity2 scores. The loss of strength gradient is measured as the inverse of logged distance. All models include controls for the natural log of income (SDP or GDP), a country's Polity2 score, and a measure of potential threat based on population. We distinguish between models that control for subsistence (or population for the GDP models) and those that do not. Standard errors are clustered by country; right-hand side variables are lagged by one year. Within each panel, rows distinguish between alternative measurements of economic power, that is SDP using a \$3, \$2, and \$1 per diem subsistence level as well as standard GDP.

## G.4 Alternative interest compatibility measures

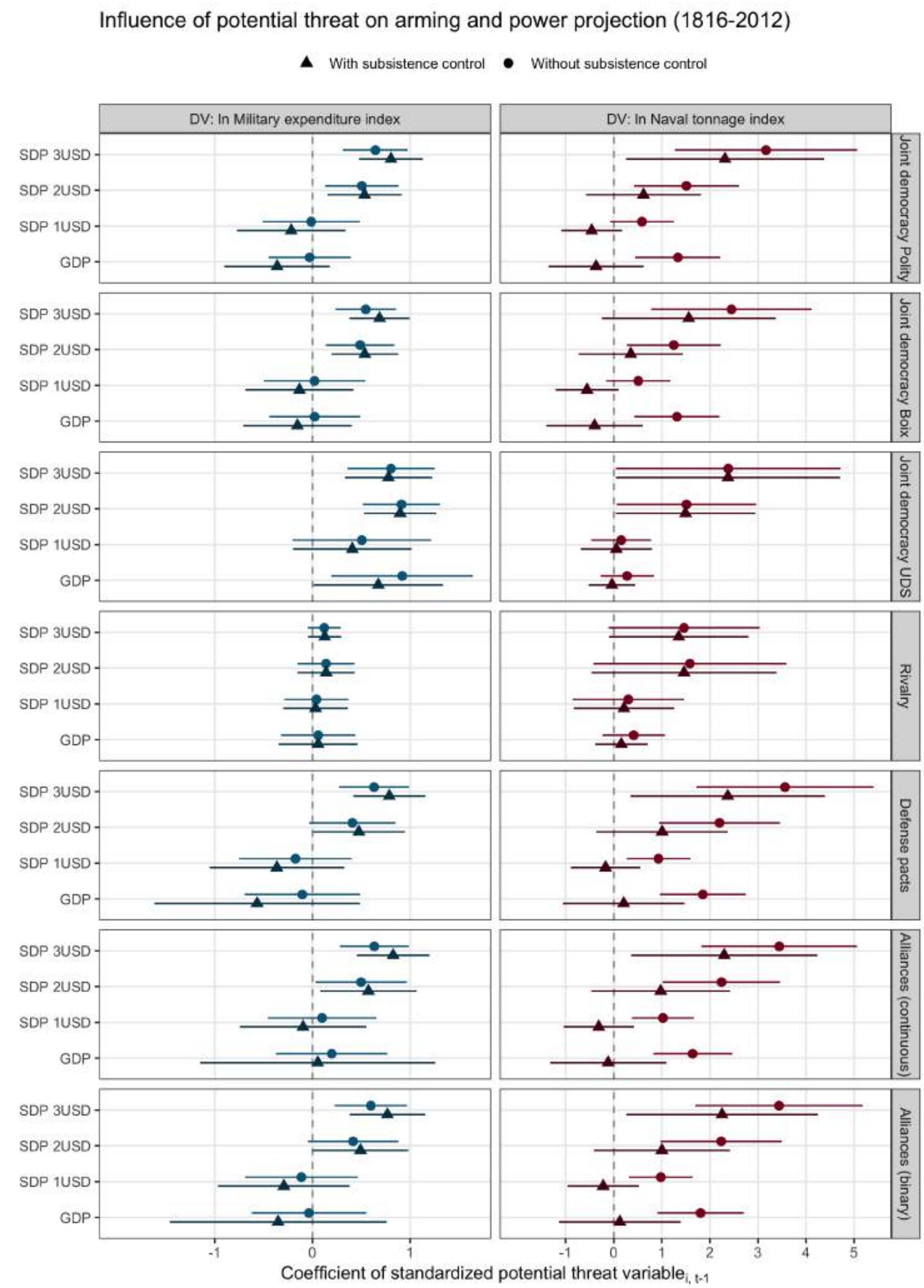



Figure 26: The graph plots the coefficients and 95% confidence intervals of standardized potential threat variables for regression models of two dependent variables — the military expenditure index and naval tonnage index — on potential threat and control variables. Preference compatibility is measured via several alternative indicators of preference compatibility. The loss of strength gradient is measured as the inverse of logged distance. All models include controls for the natural log of income (SDP or GDP), a country's Polity2 score, and a measure of potential threat based on population. We distinguish between models that control for subsistence (or population for the GDP models) and those that do not. Standard errors are clustered by country; right-hand side variables are lagged by one year. Within each panel, rows distinguish between alternative measurements of economic power, that is SDP using a \$3, \$2, and \$1 per diem subsistence level, as well as standard GDP.

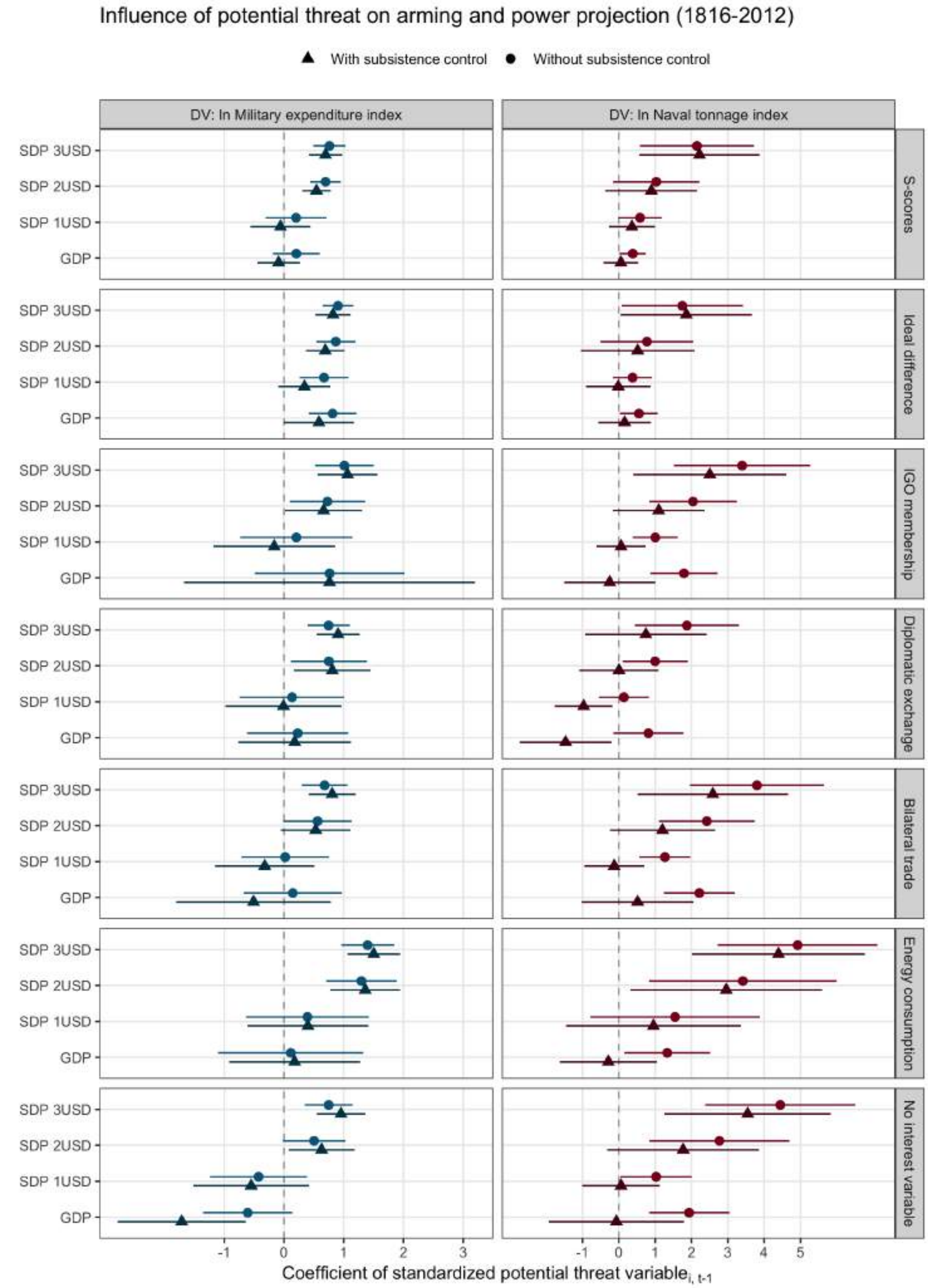



Figure 27: The graph plots the coefficients and 95% confidence intervals of standardized potential threat variables for regression models of two dependent variables — the military expenditure index and naval tonnage index — on potential threat and control variables. Preference compatibility is measured via several alternative indicators of preference compatibility. The loss of strength gradient is measured as the inverse of logged distance. All models include controls for the natural log of income (SDP or GDP), a country's Polity2 score, and a measure of potential threat based on population. We distinguish between models that control for subsistence (or population for the GDP models) and those that do not. Standard errors are clustered by country; right-hand side variables are lagged by one year. Within each panel, rows distinguish between alternative measurements of economic power, that is SDP using a \$3, \$2, and \$1 per diem subsistence level, as well as standard GDP.

## G.5 Summary statistics

Table 4: Summary statistics for key variables.

| Statistic | Min | Median | Mean | Max | St. Dev. | N |
|---|---|---|---|---|---|---|
| ln Military expenditure/GDP | −20.72 | −3.98 | −4.55 | 0.00 | 3.69 | 13,097 |
| ln Naval tonnage/GDP | −20.72 | −20.72 | −18.74 | −11.74 | 2.80 | 13,321 |
| ln Military Expenditure/SDP | −20.72 | −3.35 | −3.68 | 0.00 | 3.75 | 13,097 |
| ln Naval tonnage/SDP | −20.72 | −20.72 | −17.36 | 12.02 | 6.70 | 13,321 |
| Potential threat (SDP) joint democracy (Polity) | 0.06 | 3.61 | 4.92 | 18.58 | 4.16 | 16,001 |
| Potential threat (SDP) joint democracy (Boix et al.) | 0.04 | 3.63 | 5.05 | 20.82 | 4.44 | 15,284 |
| Potential threat (SDP) joint democracy (UDS) | 0.002 | 1.74 | 2.29 | 9.90 | 2.08 | 9,694 |
| Potential threat (SDP) rivalry | 0.00 | 0.00 | 0.06 | 1.79 | 0.12 | 12,427 |
| Potential threat (SDP) defense alliances | 0.07 | 4.56 | 6.17 | 21.16 | 4.78 | 17,355 |
| Potential threat (SDP) s-scores | 0.04 | 1.47 | 1.59 | 12.88 | 0.98 | 9,424 |
| Potential threat (SDP) absolute ideal difference | 0.03 | 1.70 | 1.72 | 10.10 | 0.94 | 9,537 |
| Potential threat (SDP) joint IGO membership | 0.11 | 7.25 | 7.71 | 19.64 | 4.96 | 7,933 |
| Potential threat (SDP) diplomatic exchange | 0.00 | 2.34 | 4.29 | 19.24 | 4.81 | 2,660 |
| Potential threat (SDP) bilateral trade | 0.12 | 5.80 | 7.24 | 21.27 | 5.20 | 11,519 |
| Potential threat (SDP) per capita energy consumption | 0.01 | 1.54 | 2.11 | 9.31 | 1.91 | 14,402 |
| Potential threat (SDP) ATOP Alliances (continuous) | 0.04 | 3.14 | 4.32 | 13.82 | 3.41 | 15,081 |
| Potential threat (SDP) ATOP Alliances (binary) | 0.05 | 4.70 | 6.39 | 21.16 | 5.03 | 15,081 |
| Potential threat (SDP) no interest variable | 0.15 | 6.22 | 7.62 | 21.66 | 5.19 | 26,067 |
| Potential threat (GDP) joint democracy (Polity) | 0.07 | 3.93 | 5.00 | 18.33 | 3.94 | 16,001 |
| Potential threat (GDP) joint democracy (Boix et al.) | 0.08 | 3.98 | 5.14 | 20.10 | 4.18 | 15,284 |
| Potential threat (GDP) joint democracy (UDS) | 0.003 | 1.85 | 2.30 | 10.24 | 1.98 | 9,694 |
| Potential threat (GDP) rivalry | 0.00 | 0.00 | 0.06 | 1.70 | 0.11 | 12,427 |
| Potential threat (GDP) defense alliances | 0.10 | 4.90 | 6.29 | 20.86 | 4.61 | 17,355 |
| Potential threat (GDP) s-scores | 0.05 | 1.46 | 1.59 | 13.69 | 0.99 | 9,424 |
| Potential threat (GDP) absolute ideal difference | 0.04 | 1.70 | 1.72 | 10.76 | 0.94 | 9,537 |
| Potential threat (GDP) joint IGO membership | 0.16 | 7.53 | 7.73 | 19.67 | 4.82 | 7,933 |
| Potential threat (GDP) diplomatic exchange | 0.00 | 2.43 | 4.31 | 20.22 | 4.77 | 2,660 |
| Potential threat (GDP) bilateral trade | 0.17 | 5.93 | 7.27 | 20.81 | 5.06 | 11,519 |
| Potential threat (GDP) per capita energy consumption | 0.01 | 1.63 | 2.13 | 9.76 | 1.83 | 14,402 |
| Potential threat (GDP) ATOP Alliances (continuous) | 0.05 | 3.26 | 4.34 | 13.94 | 3.30 | 15,081 |
| Potential threat (GDP) ATOP Alliances (binary) | 0.06 | 4.89 | 6.42 | 20.86 | 4.87 | 15,081 |
| Potential threat (GDP) no interest variable | 0.15 | 7.28 | 7.98 | 21.21 | 5.01 | 26,067 |
| Potential threat (population) joint democracy (Polity) | 0.05 | 4.14 | 5.00 | 17.29 | 3.52 | 16,001 |
| Potential threat (population) joint democracy (Boix et al.) | 0.05 | 4.12 | 5.14 | 19.64 | 3.81 | 15,284 |
| Potential threat (population) joint democracy (UDS) | 0.01 | 1.85 | 2.30 | 10.08 | 1.91 | 9,694 |
| Potential threat (population) rivalry | 0.00 | 0.00 | 0.06 | 1.11 | 0.10 | 12,427 |
| Potential threat (population) defense alliances | 0.08 | 5.02 | 6.29 | 22.18 | 4.49 | 17,355 |
| Potential threat (population) s-scores | 0.05 | 1.39 | 1.59 | 13.88 | 1.08 | 9,424 |
| Potential threat (population) absolute ideal difference | 0.05 | 1.59 | 1.72 | 10.92 | 1.01 | 9,537 |
| Potential threat (population) joint IGO membership | 0.11 | 7.29 | 7.73 | 21.17 | 4.68 | 7,933 |
| Potential threat (population) diplomatic exchange | 0.00 | 2.58 | 4.31 | 21.43 | 4.62 | 2,660 |
| Potential threat (population) bilateral trade | 0.12 | 6.31 | 7.30 | 21.93 | 4.94 | 11,519 |
| Potential threat (population) per capita energy consumption | 0.05 | 1.69 | 2.13 | 9.39 | 1.68 | 14,402 |
| Potential threat (population) ATOP Alliances (continuous) | 0.05 | 3.59 | 4.34 | 13.63 | 3.16 | 15,081 |
| Potential threat (population) ATOP Alliances (binary) | 0.07 | 5.21 | 6.42 | 22.18 | 4.71 | 15,081 |
| Potential threat (population) no interest variable | 0.13 | 6.80 | 7.98 | 22.39 | 4.93 | 26,067 |
| ln GDP | 15.36 | 22.77 | 22.55 | 30.74 | 2.78 | 27,321 |
| ln SDP | 0.00 | 21.70 | 17.78 | 30.68 | 9.52 | 27,321 |
| ln Subsistence | 13.80 | 21.86 | 21.44 | 28.00 | 2.50 | 27,321 |
| ln Population | 6.80 | 14.99 | 14.76 | 21.00 | 2.21 | 27,321 |
| Polity2 score | −10.00 | −3.00 | −0.55 | 10.00 | 7.07 | 16,974 |

*Notes*:
GDP measure in constant 2011 international PPP dollars.
SDP is based on a \$3 per diem subsistence level.
Loss of strength gradient measured using the following formula $\frac{1}{\log(distance)}$.
Very small values are rounded to 0 in the output above.

## H GDP, Population, and GDPpc Component Datasets

Total power-resources are measured using GDP data in constant 2011 international dollars from the World Development Indicators (World Bank, 2016), and supplemented with a number of historic GDP data estimates that are combined using a measurement model to estimate a GDP series that covers the entire period of observation 1816–2012. The latent variable model that is used to compute the GDP and population data for the analysis is estimated based on data for Gross Domestic Product (GDP)[7], GDP per capita[8], and population.[9] Details on the sources, measurement choices, and coverage of the component variables are provided in Table 5. For each component dataset, we extract relevant indicators, attach unique country identifiers, and reshape the data into a common country-year format. We consulted the codebooks of each dataset to drop observations that are interpolated or extrapolated by the authors of the dataset, or already covered by other datasets (e.g., the data generated by Gleditsch (2002) includes some interpolated values and values taken from the Maddison Project). Details on the underlying source materials for each component measure and coding decisions are provided below and are documented in the `R` code we use to merge the constituent datasets together.

When merging the different variables together we relied on the available country-year units as prepared by the authors of the original datasets. We use the Gleditsch and Ward (1999b) revised list of independent states as the base set of units. For years prior to the start year of this data set (1816 A. D.) we again use the date the year the unit enters the dataset or 1500 A.D. As we discussed in each dataset description, different datasets sometimes use different spatial definitions for units. 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 Germany had completely 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.

[7] For observed data on GDP see World Bank (2016); Feenstra et al. (2015); Broadberry and Klein (2012); Maddison (2010); Gleditsch (2002); Bairoch (1976).

[8] For observed data on GDP per capita see World Bank (2016); Broadberry (2015); The Maddison-Project (2013); Broadberry and Klein (2012); Gleditsch (2002); Bairoch (1976).

[9] For observed data on population see World Bank (2016); Feenstra et al. (2015); Broadberry and Klein (2012); Maddison (2010); Gleditsch (2002); Singer et al. (1972).

Table 5: Component Measures for GDP, GDP per capita, and Population Latent Variable Model

| Variable Descriptions | Coverage in Original | Coverage in Model | Source Material and Citations |
|---|---|---|---|
| GDP data are measured in 1990 international dollars. | 1AD–2008 | 1500–2008 | Historical GDP data collected by Angus Maddison (Maddison, 2010). |
| GDP data are measured as total real GDP at 2005 prices (PPP). | 1950-2011 | 1950-2011 | Expanded GDP data version 6.0 beta, September 2014 (Gleditsch, 2002). |
| GDP data are measured in constant 2010 USD. | 1960–2015 | 1960–2015 | World Development Indicators (World Bank, 2016) |
| GDP data are measured in constant 2011 international dollars (PPP). | 1990–2015 | 1960–2015 | World Development Indicators (World Bank, 2016) |
| GDP data limited to European countries and the United States, after accounting for changing country boundaries. GDP is measured in millions of 1990 international dollars (national currencies are converted to international dollars using Angus Maddison's purchasing power parities) | 1870–2001 | 1870–2001 | Broadberry and Klein (2012). |
| GNP data limited to European countries, after accounting for changing country boundaries. GNP is measured at market prices and expressed in constant 1960 US dollars. | 1830–1973 | 1830–1973 | Bairoch (1976). |
| GDP (expenditure oriented) in millions of constant 2011 international dollars (PPP). | 1950–2014 | 1950–2014 | Penn World Tables version 9.0 (Feenstra et al., 2015). |
| GDP (output oriented) in millions of constant 2011 international dollars (PPP). | 1950–2014 | 1950–2014 | Penn World Tables version 9.0 (Feenstra et al., 2015). |
| GDP per capita data are measured in 1990 international dollars. | 1AD-2010 | 1500–2010 | Extension of Angus Maddison's historical GDP and population estimates (The Maddison-Project, 2013). |
| GDP per capita data are measured as total real GDP at 2005 prices (PPP). | 1950-2011 | 1950-2011 | Expanded GDP data version 6.0 beta, September 2014 (Gleditsch, 2002). |
| GDP per capita are measured in constant 2010 USD. | 1960–2015 | 1960–2015 | World Development Indicators (World Bank, 2016) |
| GDP per capita are measured in constant 2011 international dollars (PPP). | 1990–2015 | 1960–2015 | World Development Indicators (World Bank, 2016) |
| GDP per capita data limited to European countries and the United States, after accounting for changing country boundaries. GDP is measured in millions of 1990 international dollars. | 1870–2001 | 1870–2001 | Broadberry and Klein (2012). |
| GNP per capita data are limited to European countries, after accounting for changing country boundaries. GNP is measured at market prices and expressed in constant 1960 US dollars. | 1830–1973 | 1830–1973 | Bairoch (1976). |
| GDP per capita data limited England/Great Britain, Holland/Netherlands, Italy, Spain, Japan, China, and India. GDP is measured in millions of 1990 international dollars. | 725–1850 | 1500–1850 | Broadberry (2015). |
| Total population measured in thousands at mid-year. | 1AD–2030 | 1500–2010 | Historical population data collected by Angus Maddison (Maddison, 2010). |
| Total population measured in thousands. | 1950-2011 | 1950-2011 | Expanded GDP data version 6.0 beta, September 2014 (Gleditsch, 2002). |
| Population data limited to European countries and the United States. | 1870–2001 | 1870–2001 | Broadberry and Klein (2012). |
| Total population. | 1960–2015 | 1960–2015 | World Development Indicators (World Bank, 2016) |
| Total population measured in thousands. | 1816–2001 | 1816–2001 | The Correlates of War Project's National Material Capabilities data version 4.0 (Singer et al., 1972) |
| Population (in millions). | 1950–2014 | 1950–2014 | Penn World Tables version 9.0 (Feenstra et al., 2015). |

**The Maddison Project (Maddison, 2010; The Maddison-Project, 2013):** Maddison's original GDP, GDP per capita, and population variables are derived from a large number of country-level sources (Maddison, 2003, 2001, 1995). Because the underlying source materials employed by Maddison are expansive and country-specific, we refrain from describing them in detail. The more recent version of these data, The Maddison-Project (2013), is based on a collaboration of researchers dedicated to continuing Angus Maddison's data collection efforts by extending and, if warranted, revising his estimates. Due to the collaborative nature of the effort, different research teams use different methods and source material to obtain their estimates. With a few exceptions, data from 1990–2010 were revised using figures from the Total Economy Database of the Conference Board (Bolt and van Zanden, 2014). Other estimates are based on historical national statistics from country-specific sources (Bolt and van Zanden, 2014). We subset the data from the Maddison Project to include only country-year observations starting in 1500. The original Maddison (2010) data includes both GDP and population values. The updated version only included GDP per capita estimates. We include both data versions in our model since, as we describe below, it is capable of linking all of these observed indicators together in united model that leverages the information from each type of variable. Unlike some of the other datasets we describe below, these datasets do not contain origin codes that classify the source material used to inform the country-year values.

**Expanded GDP data version 6.0 beta (Gleditsch, 2002):** Gleditsch (2002)'s (beta) version 6.0 of the Expanded GDP data is based primarily on the Penn World Tables (PWT) 8.0, and supplemented with data from the PWT 5.6, the Maddison Project Database, and the World Bank Global development indicators. In addition, Gleditsch (2002) constructed his data using imputations for the lead and tail values, as well as interpolation for estimates within the series. We use only the values that stem from the PWT figures in the latent variable model (origin codes 0, -1, and 3) and exclude data from the Maddison Project, as well as interpolated or imputed figures (origin codes -2, 1, and 2). In the Validity section below, we consider the model fit for the latent variables estimates that do include these variables compared to the latent variable model estimates that exclude them and demonstrate the model fit is improved by estimating these missing values using our model-based approach instead of using interpolation or extrapolation.

**World Development Indicators (World Bank, 2016):** We include GDP, GDP per capita, and population from the World Bank (2016). Where possible, we use the metadata for each indicator provided by the World Bank's DataBank to determine the underlying source material of the GDP, GDP per capita, and population values. As with the Gleditsch (2002) data, we drop values that are interpolated or extrapolated and allow our model to generate new estimates for these units. We describe each of these variables in turn.

We include the World Bank (2016) GDP indicator measured in constant 2010 US dollars in our latent variable model. The figures are compiled from the World Bank and OECD national accounts data. The documentation in the metadata file indicates that the series is based on an underlying interpolation of component data upon aggregating it to a "gap-filled total." Unfortunately, we do not have information on the details of this aggregation process. We therefore use the full series of GDP as provided by the World Bank (2016)'s online data portal DataBank. In future versions of our model, we plan to identify these cases when possible and adjust our model accordingly.

The per capita GDP series is based on the World Bank (2016)'s GDP in constant 2010 US dollars and the total population figures (for the underlying source material see below). According to the metadata, the data is aggregated using weighted averages. We exclude observations from our model that the metadata indicates as being preliminary, extrapolated, or interpolated. Information on which country-years were excluded based on the metadata is provided in the replication material that accompanies this paper.

The population figures from World Bank (2016) are based on national population censuses. The census data that informs this measure stem from a variety of sources, including the United Nations World Population Prospects (for the majority of developing countries), Eurostat (for European countries), and national statistical agencies. The data are interpolated for all years between census years. Since we do not have information on the years that a census was conducted for each country, we retain the interpolated data for the use in the latent variable model. We do, however, exclude population figures that are explicitly indicated as being extrapolated, interpolated, or preliminary in the metadata. Information on which country-years-units were excluded is provided in the replication material that accompanies this paper. In future versions of our model, we plan to identify the other interpolated cases when possible and again adjust our model accordingly.

**Broadberry and Klein (2012):** The GDP, GDP per capita, and population variables in Broadberry and Klein (2012) are limited to European nations, including Russia and Turkey, as well as the United States. A detailed list of underlying source material is available in the paper's appendix (Broadberry and Klein, 2012, pp. 105). For GDP, these sources include the data from Maddison (2010), official national account statistics, and the work of country-expert historians. Data on population are drawn mainly from Mitchell (2003) and Maddison (2010), and supplemented with country-specific data from official national statistics and historians. We exclude those country-year observations that are taken from Maddison (2010) in our model.

**Bairoch (1976):** The underlying source material for the data by Bairoch is detailed in the paper's methodological appendix. For GNP, these sources include the work of historians and official national statistics for earlier country-years, as well as OECD figures for years starting in 1950 (Bairoch, 1976, 329 et seq.). For the 19th century and the year 1900, three-year annual averages are available for every decade starting from 1830 and expressed in 1960 U.S. dollars (Bairoch, 1976, 286). For the 20th century, data are available for select years between 1913 and 1973 and expressed in 1960 U.S. dollars as well (Bairoch, 1976, 297). For population, Bairoch relies on United Nations Demographic yearbooks, data from the League of Nations, and national statistical agencies to assemble his data (321). We incorporate all of Bairoch's estimates in our model, including the ones flagged as having a larger-than-average margin of error (the figures presented in parentheses). The data from Bairoch (1976) cover the total and per capita gross *national* product (GNP), not gross *domestic* product (GDP). Bairoch's definition is based on the United Nations' 1953 System of National accounts (United Nations, 1953). With the exception of the data from Bairoch (1976), the data on economic size are measured as the gross domestic product (GDP). Bairoch (1976) uses gross national product (GNP) instead. While the GNP excludes value added by foreign firms, this measure is highly correlated with GDP. The correlation between GNP and GDP is quite high, with correlation coefficients between 0.865 and 0.995 for country-year units within the period 1830–1973. The strength of the positive relationship varies over time but rarely falls below 0.9. We anticipate that in future years, the correlation between the two measures should drop as globalization increases and the internationalization of production and investment increases the relevance of the conceptual difference between GNP and GDP. Additional estimates of GNP and GDP from more recent years would help researchers determine how this empirical relationship evolves over time. The evaluation of this distinction is one possible avenue that our new latent variable model opens up for exploration, which we discuss below.

**Broadberry (2015):** The GDP per capita estimates in Broadberry (2015) are based on historical national accounting data that is constructed from documents such as "government accounts, customs accounts, poll tax returns, parish registers, city records, trading company records, hospital and educational establishment records, manorial accounts, probate inventories, farm accounts, tithe files and other records of religious institutions." (Broadberry, 2015, 5). Broadberry lists the data sources for each country in the main text.[10] As with the Maddison data, we exclude cases for years prior to 1500 from our model.

**COW National Military Capabilities data v4.0 (Singer et al., 1972):** The Correlates of War Project provides a variety of country-level estimates including population beginning in the year 1816. For country-years starting in 1919, the population estimates by Singer et al. (1972) are based primarily on the estimates of the United Nations Statistical Office. The population estimates for years prior to 1919 are based on national government censuses. For these earlier years in the series, the authors of the population dataset selected country-specific data that presents the greatest continuity with the data from the United Nations.[11] The authors of the data use a variety of methods to bridge gaps in the data, including interpolation, regression, and extrapolation. Quality codes for the estimates of the total population figure are specified — indicating whether a data point stems from an identified source, is missing, derived through interpolation, regression,

[10] Pages 6 and 7 contain the underlying source material for Britain, the Netherlands, Italy, and Spain; page 8 contains the data for China, Japan, and India.

[11] For details, please refer to the codebook for version 4.0 of the data: Correlates of War Project National Material Capabilities Data Documentation Version 4.0, `http://cow.dss.ucdavis.edu/data-sets/national-material-capabilities/nmc-codebook/at_download/file`, accessed 1 December 2016.

or extrapolation. We retain only those data points that stem from an identified source (quality code A).

# I Latent Variable Model Specification

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 $y_{itj}$. Because the observed variables that enter the model represent three different concepts—GDP, population, and GDP per capita—we estimate three latent variable parameters, where $k = 1, 2, 3$, indexes the three categories *gdp*, *pop*, *gdppc*. This allows us to define the set of $y_{itj}$ that we observe for each of the $k$ dimensions of the latent variable model, where $y_{itj} * 1\{y \in \pi_k\}$. This notation allows us to denote the set of observed variables used to estimate each of the three underlying latent variables such that $\pi_{gdp} = \{y_{it1}, y_{it2}, y_{it3}, y_{it4}, y_{it5}\}$, $\pi_{pop} = \{y_{it6}, y_{it7}, y_{it8}, y_{it9}, y_{it10}\}$, $\pi_{gdppc} = \{y_{it11}, y_{it12}, y_{it13}, y_{it14}, y_{it15}, y_{it16}\}$.[12]

With knowledge of how the observed variables relate to each category $k$, we can denote how the three dimensions of the latent variable relate to them as well. The model assumes that the latent variables take the form: $\theta_{itk} \sim \mathcal{N}(0, 1)$ 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: $\theta_{itk} \sim \mathcal{N}(\theta_{it-1,k}, \sigma_k)$.

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 $t = 1$, $\sigma_{gdp}$ and $\sigma_{pop}$ are drawn from a uniform distribution $U(0, 1)$. For the latent GDP per capita variable, the latent estimates and associated uncertainty are deterministically determined by the GDP and Population latent variables themselves such that $\theta_{it,gdppc} \leftarrow \frac{\theta_{it,gdp}}{\theta_{it,pop}}$. This modeling innovation allows information form the three types of observed variables to inform more than just one of the latent variables.

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 the three types of variables $\tau_k$: $\{\tau_{gdp}, \tau_{pop}, \tau_{gdppc}\}$. These $\tau_k$ parameters are estimates of model level uncertainty, which link each of the latent variables to the sets of observed GDP, population, or GDP per capita variables. Shape parameters translate the observed variables from their original unit-of-measurement into the latent variable unit-of-measurement. Because we specify a Gaussian link function, these shape parameters are the intercept and slope from the linear model. For the intercept parameters $\alpha_j$, we center the standard normal prior around the the mean value of the observed data with a relatively large variance (low precision): $\alpha_j \sim \mathcal{N}(\bar{y}_j, 4)$. We

[12] A useful feature of this notation is that the sets of observed variables do not need to be mutually exclusive. Though we do not allow the observed variables to inform the estimation of multiple latent variables in the application presented here, this is a possibility in other applications. See Gelman and Hill (2007); Imai et al. (2017) for more details. We thank James Lo for this notational suggestion.

choose the mean value of the observed variables because the mean of latent traits themselves are centered around 0.[13] The intercept parameter therefore transforms the latent trait into the unit-of-measurement of the original observed variable. For identification of the model we set $\beta_j = 1$ because we assume a one-unit change in the latent trait is equivalent to a one-unit change in the original observed variable.[14] All of the prior distributions are summarized in Table 6. Recall that we organize the three types of observed variables in three sets such that $y_{itj} * 1\{y \in \pi_k\}$. Therefore, the likelihood function that links the observed data to the estimated parameters is:

$$\mathcal{L}(\beta, \alpha, \tau, \theta | y_{itj} * 1\{y \in \pi_k\}) = \prod_{i=1}^{N} \prod_{t=1}^{T} \prod_{j=1}^{J} \prod_{k=1}^{K} \mathcal{N}(\alpha_j + \theta_{itk}\beta_j, \tau_k)$$

Table 6: Prior Distribution for Latent Variables and Model Level Parameter Estimates

| Parameter | Prior | | |
|---|---|---|---|
| Country $i$ latent GDP estimate in first year $t$ | $\theta_{it=1,gdp}$ | $\sim$ | $\mathcal{N}(0, 1)$ |
| Country $i$ latent GDP estimate in all other years | $\theta_{it,gdp}$ | $\sim$ | $\mathcal{N}(\theta_{t-1,gdp}, \sigma_{gdp})$ |
| Latent GDP uncertainty | $\sigma_{gdp}$ | $\sim$ | $U(0, 1)$ |
| Country $i$ latent population estimate in first year $t$ | $\theta_{it=1,pop}$ | $\sim$ | $\mathcal{N}(0, 1)$ |
| Country $i$ latent population estimate in all other years | $\theta_{it,pop}$ | $\sim$ | $\mathcal{N}(\theta_{t-1,pop}, \sigma_{pop})$ |
| Latent population uncertainty | $\sigma_{pop}$ | $\sim$ | $U(0, 1)$ |
| Country $i$ latent GDP per capita estimate | $\theta_{it,gdppc}$ | $\leftarrow$ | $\frac{\theta_{it,gdp}}{\theta_{it,pop}}$ |
| Model $j$ intercept "difficulty parameter" | $\alpha_j$ | $\sim$ | $\mathcal{N}(\bar{y}_{itj}, 4)$ |
| Model $j$ slope "discrimination parameter" | $\beta_j$ | $\leftarrow$ | 1 |
| Model uncertainty for all GDP items | $\tau_{gdp}$ | $\sim$ | $\mathcal{G}(0.001, 0.001)$ |
| Model uncertainty for all population items | $\tau_{pop}$ | $\sim$ | $\mathcal{G}(0.001, 0.001)$ |
| Model uncertainty for all GDP per capita items | $\tau_{gdppc}$ | $\sim$ | $\mathcal{G}(0.001, 0.001)$ |

The model is estimated with five MCMC chains, run for 100,000 iterations each. The first 50,000 iterations were thrown away as burn-in and the rest were used to generate the posterior prediction intervals for the original observed variables.[15]

[13]We set this parameter to the empirical mean of the Maddison GDP and population variables as an identification constraint.

[14]This assumption can be relaxed to examine the relative strength of the relationship between one measure compared to another. We leave this analysis to future research. Relaxing this assumption would allow for analysts to explore the relative relationship between measures of GDP and GNP as functions of the underlying latent trait. We view this as a useful extension to the model we present here.

[15]The Gibbs sampler was implemented in Martyn Plummer's JAGS software (Plummer, 2010). The JAGS code used is displayed in the Appendix. Conventional diagnostics all suggest convergence.

# References


Bailey, M. A., A. Strezhnev, and E. Voeten (2017, 2018/02/15). Estimating dynamic state preferences from united nations voting data. *Journal of Conflict Resolution 61*(2), 430–456.

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

Barbieri, K. and O. Keshk (2012). Correlates of war project trade data set codebook. Online: http://correlatesofwar.org.

Barbieri, K., O. M. Keshk, and B. M. Pollins (2009). Trading data: Evaluating our assumptions and coding rules. *Conflict Management and Peace Science 26*(5), 471–491.

Bayer, R. (2006). Diplomatic exchange data set, v2006.1. Online: http://correlatesofwar.org, retreived 10 October 2016.

Boix, C., M. K. Miller, and S. Rosato (2013). A complete data set of political regimes, 1800–2007. *Comparative Political Studies 46*(12), 1523–1554.

Bolt, J. and J. L. van Zanden (2014). The Maddison Project: Collaborative research on historical national accounts. *The Economic History Review 67*(3), 627–651.

Boulding, K. E. (1962). *Conflict and Defense: A General Theory*. New York, NY: Harper and Row.

Broadberry, S. (2015, September). Accounting for the great divergence. Online. https://www.nuffield.ox.ac.uk/users/Broadberry/AccountingGreatDivergence6.pdf, accessed 23 November 2016.

Broadberry, S. and A. 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.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015, October). The next generation of the penn world table. *American Economic Review 105*(10), 3150–82.

Gelman, A. and J. Hill (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, MA: Cambridge University Press.

Gibler, D. M. (2009). *International Military Alliances, 1648–2008*. Washington, D.C.: CQ Press.

Gleditsch, K. S. (2002). Expanded trade and GDP data. *Journal of Conflict Resolution 46*(5), 712–724.

Gleditsch, K. S. and M. D. Ward (1999a). Interstate System Membership: A Revised List of the Independent States since 1816. *International Interactions 25*, 393–413.

Gleditsch, K. S. and M. D. Ward (1999b). A revised list of independent states since the congress of vienna. *International Interactions 25*(4), 393–413.

Gleditsch, K. S. and M. D. Ward (2001). Measuring space: A minimum-distance database and applications to international studies. *Journal of Peace Research 38*(6), 739–758.

Greig, J. M. and A. J. Enterline (2017). National material capabilities (nmc) data documentation version 5.0. Technical report, Department of Political Science, University of North Texas.

Imai, K., J. Lo, and J. Olmsted (2017). Fast estimation of ideal points with massive data. *American Political Science Review forthcoming*.

Klein, J. P., G. Goertz, and P. F. Diehl (2006). The new rivalry dataset: Procedures and patterns. *Journal of Peace Research 43*(3), 331–348.

Leeds, B., J. Ritter, S. Mitchell, and A. Long (2002). Alliance treaty obligations and provisions, 1815-1944. *International Interactions 28*(3), 237–260.

Maddison, A. (1995). *Monitoring the World Economy, 1820-1992*. Development Centre Studies. Paris: OECD.

Maddison, A. (2001). *The World Economy. A Millenial Perspective*. Development Centre Studies. Paris: OECD.

Maddison, A. (2003). *The World Economy. Historial Statistics*. Development Centre Studies. Paris: OECD.

Maddison, A. (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, J., C. Fariss, and R. B. McMahon (Forthcoming). Producing goods and projecting power: How what you make influences what you take. *Journal of Conflict Resolution*.

Markowitz, J. N. and C. J. Fariss (2013). Going the distance: The price of projecting power. *International Interactions 39*(2), 119–143.

Marshall, M. G., T. R. Gurr, and K. Jaggers (2016). Polity IV Project. Political Regime Charateristics and Transitions, 1800-2015, Dataset Users' Manual. Technical report, Center for Systemic Peace, Vienna, VA.

Mitchell, B. R. (2003). *International Historical Statistics: Europe, 1750 2000* (5 ed.). New York: Palgrave Macmillan.

Pemstein, D., S. A. Meserve, and J. Melton (2010). Democratic compromise: A latent variable analysis of ten measures of regime type. *Political Analysis 18*(4), 426–449.

Pevehouse, J. C., T. Nordstrom, and K. Warnke (2004). The cow-2 international organizations dataset version 2.0. *Conflict Management and Peace Science 21*(2), 101–119.

Plummer, M. (2010). Jags (just another gibbs sampler) 1.0.3 universal. Online. http://www-fis.iarc.fr/~martyn/software/jags/.

Singer, D. J., S. Bremer, and J. Stuckey (1972). Capability distribution, uncertainty, and major power war, 1820-1965. In B. Russett (Ed.), *Peace, War, and Numbers*, pp. 19–48. Beverly Hills: Sage.

The Maddison-Project (2013). Online. http://www.ggdc.net/maddison/maddison-project/home.htm, retrieved August 2016.

Trochim, W. M. K. and J. P. Donelly (2008). *The Research Methods Knowledge Base* (3rd ed.). Mason, OH: Atomic Dog.

United Nations (1953). A system of national accounts and supporting tables. Studies in Methods 2, Department of Economic Affairs Statistical Office, New York.

Weidmann, N. B., D. Kuse, and K. S. Gleditsch (2010). The geography of the international system: The cshapes dataset. *International Interactions 36*(1), 86–106.

World Bank (2016). World development indicators.