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, τ 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 τ 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 τ 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.

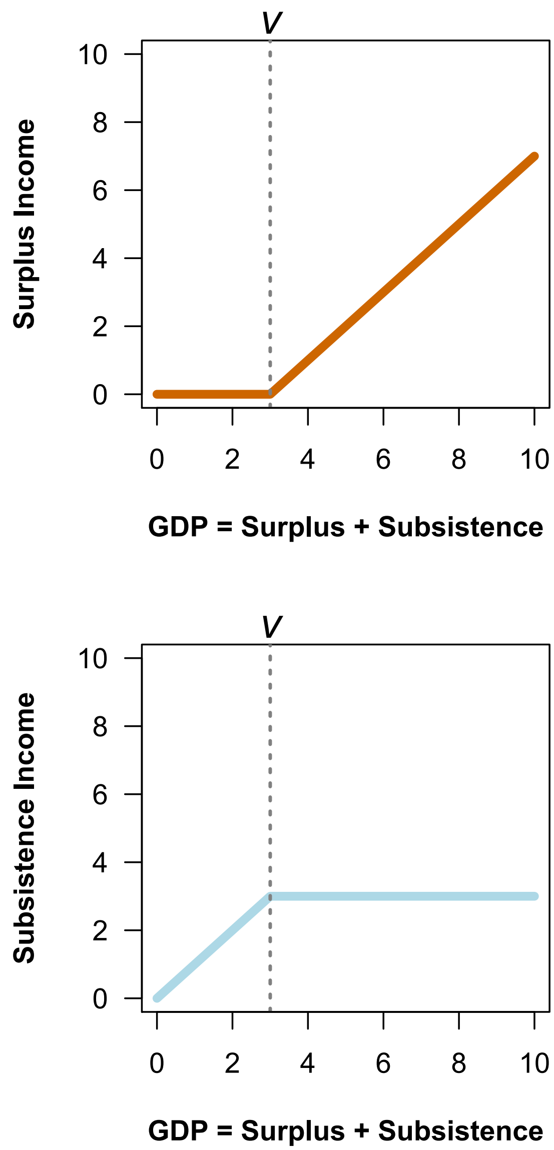


Figure 1: The dollar values displayed on the x-axes and y-axes in the panels above are in billions of $US. Suppose a country with a population of 2,739,726 people. Such a country needs to generate 3 billion $US dollars (365 days * $3 per-day * 2,739,726 people) per year to healthfully sustain each member 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.

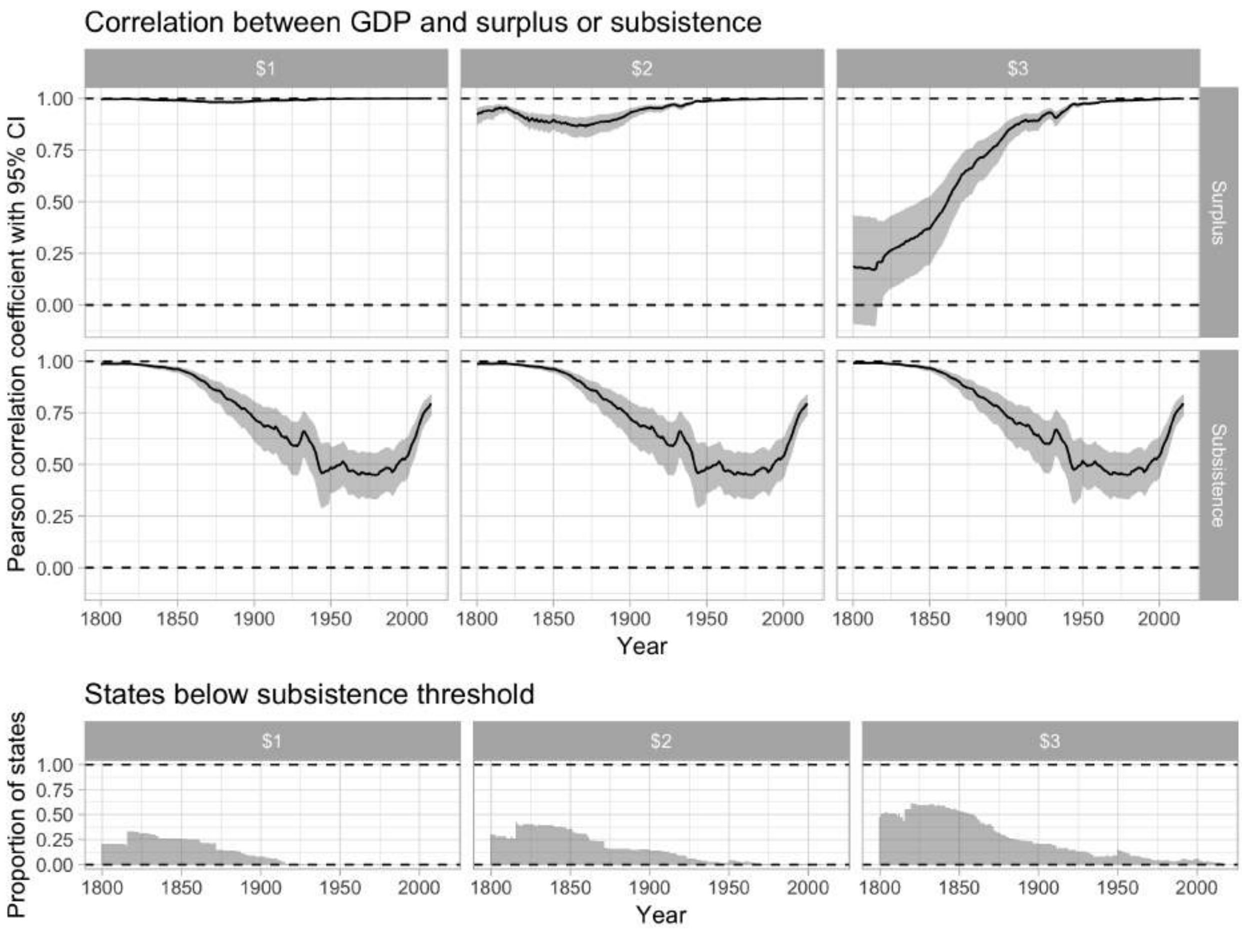


Figure 2: The top row of panels shows the yearly correlation between GDP and surplus income (SDP). The middle row of the panels shows the yearly correlation between GDP and subsistence income. The bottom row of panels shows the yearly proportion of countries that generate enough income to pass above 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.

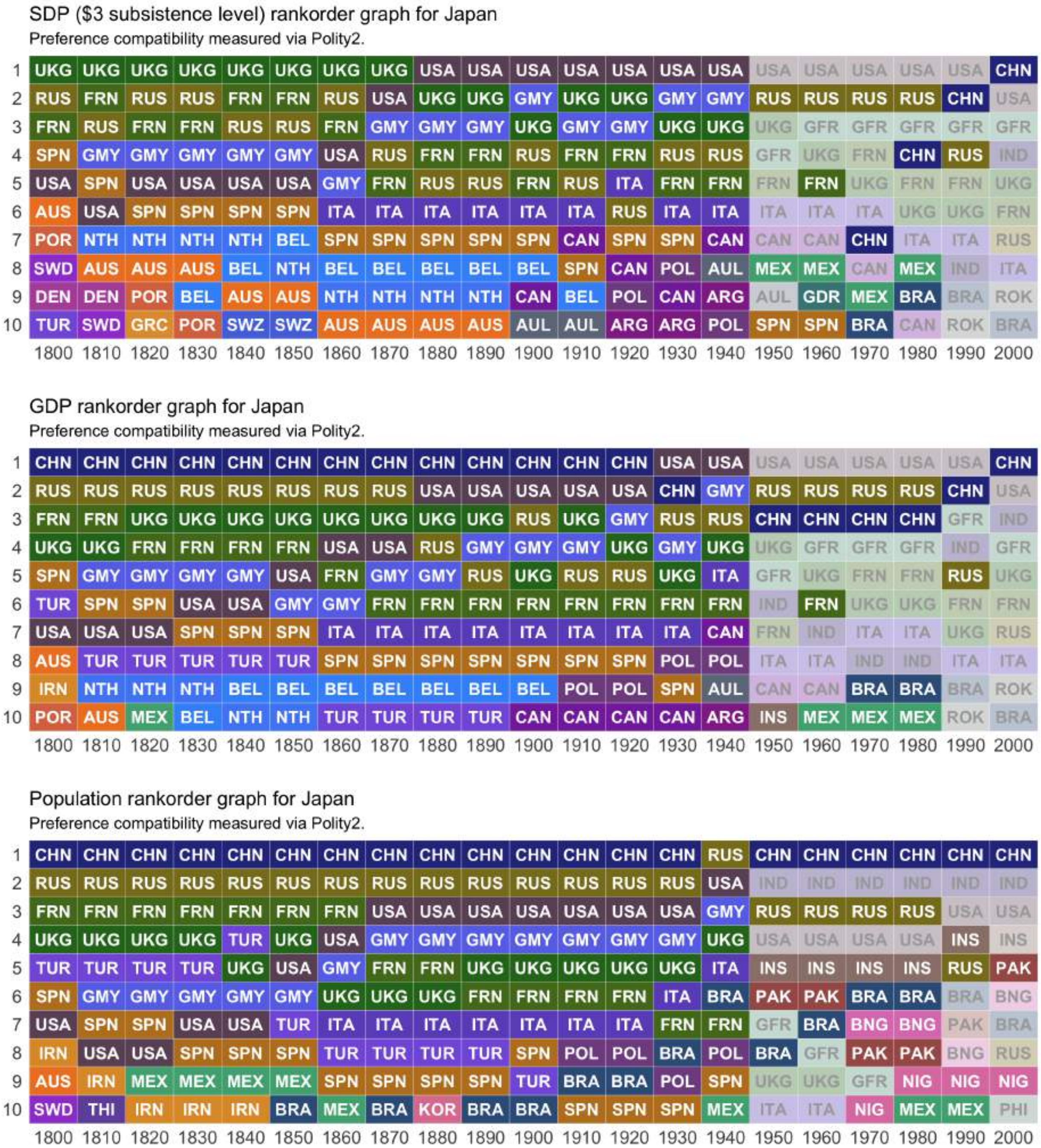


Figure 3: The Figure displays the top 10 potentially threatening states for Japan by decade for $3 per diem subsistence level relative SDP on the upper, standard relative GDP in the middle, and relative population on the lower panel. Dyads that *are not* jointly democratic are potentially threatening 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.

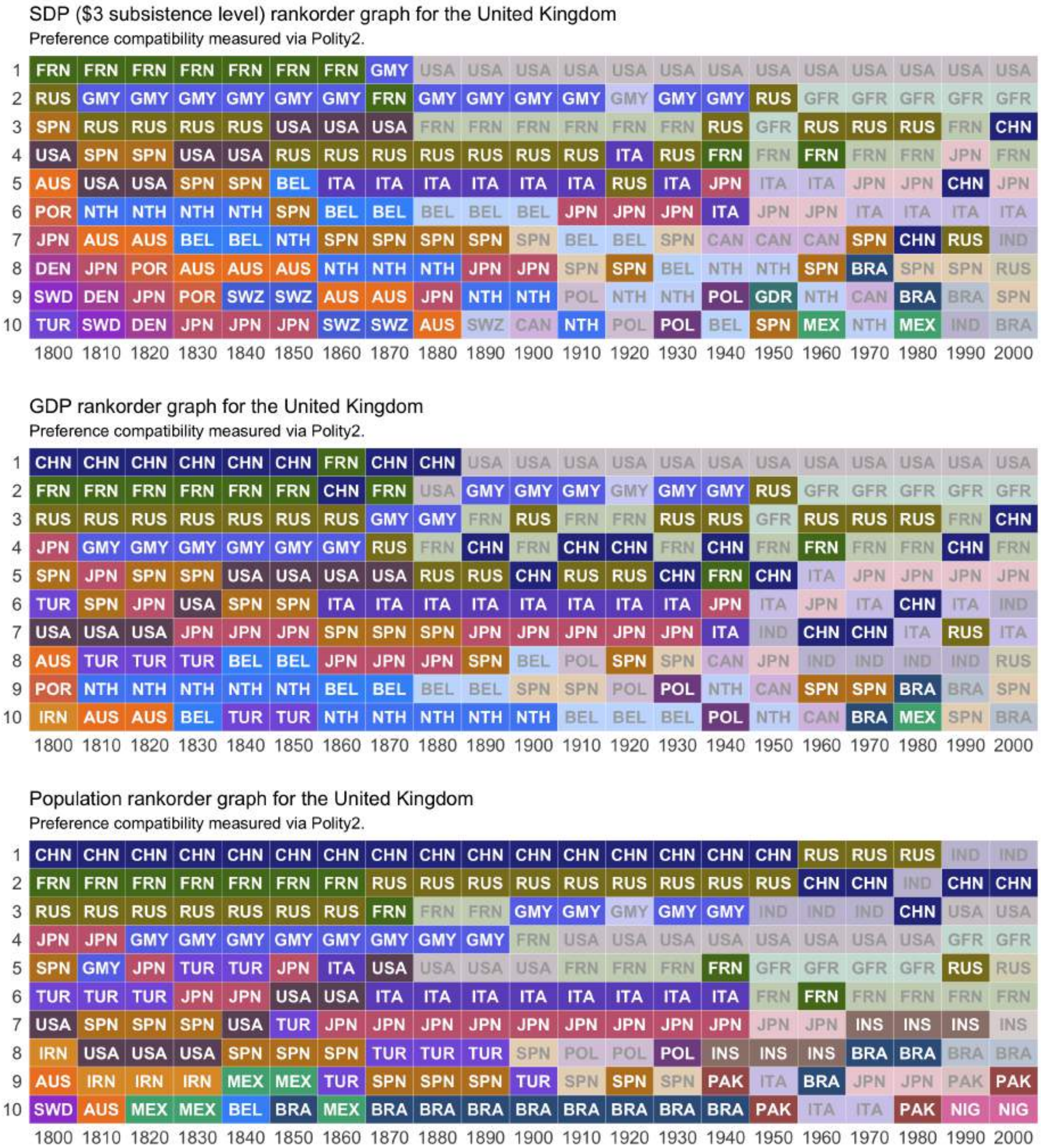


Figure 4: The Figure displays the top 10 potentially threatening states for the United Kingdom by decade for $3 per diem subsistence level relative SDP on the upper, standard relative GDP in the middle, and relative population on the lower panel. Dyads that *are not* jointly democratic are potentially 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).

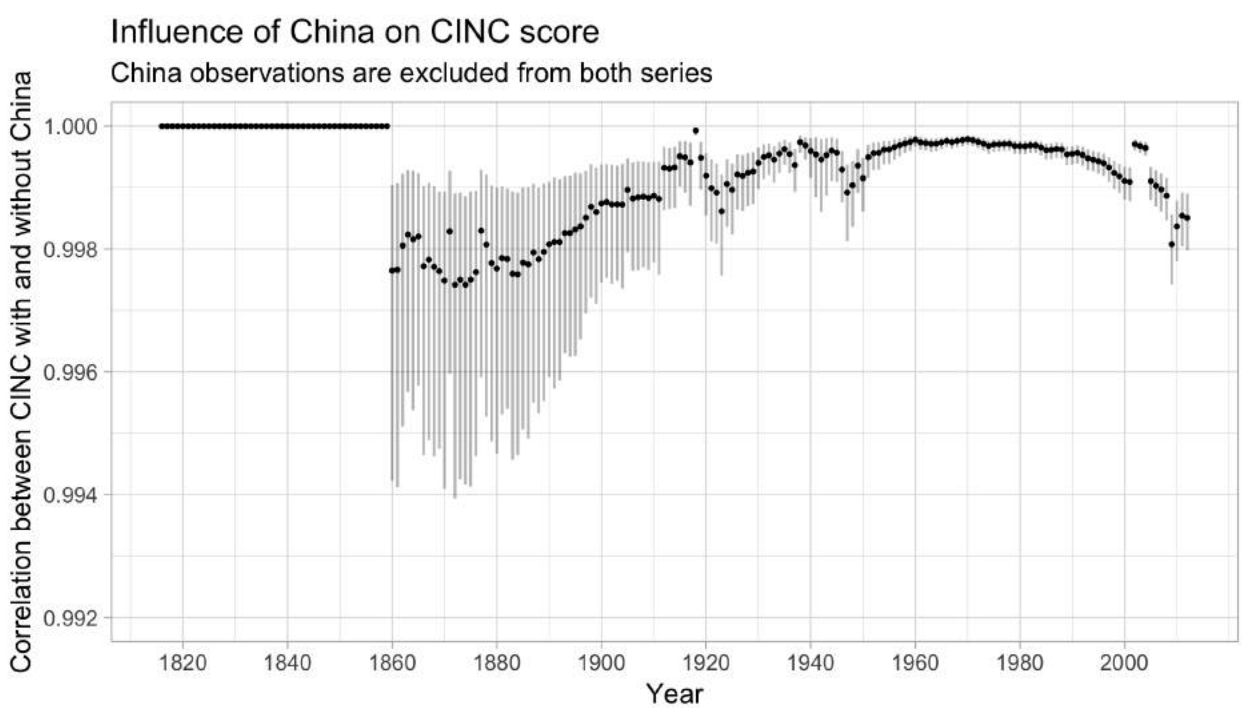
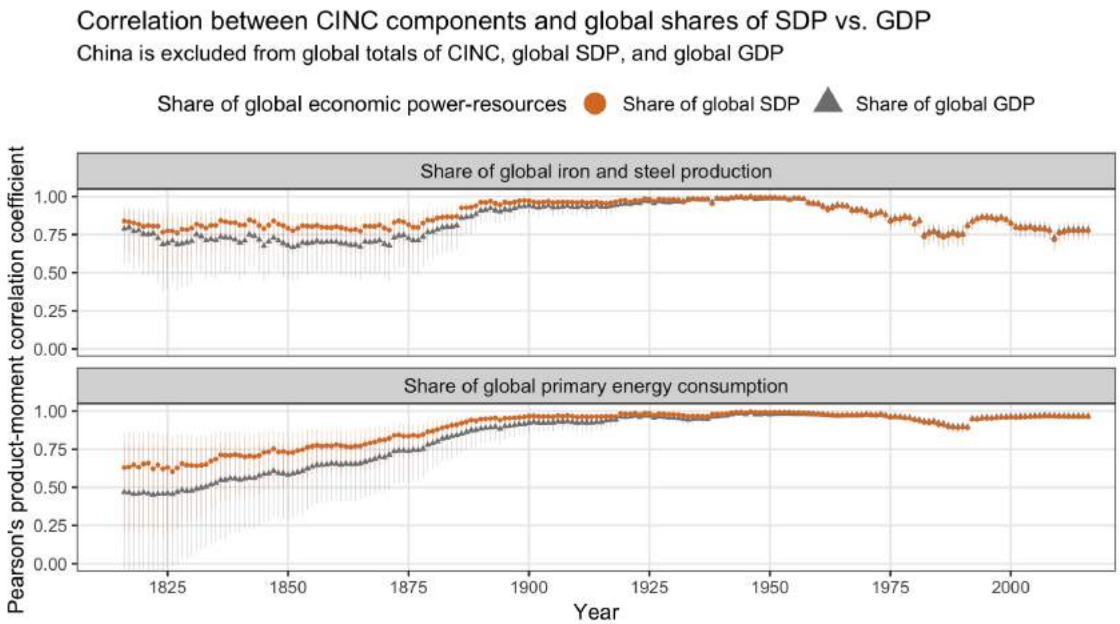


Figure 5: The plot shows the annual correlation with 95% confidence intervals between the original CINC score and a re-computation of CINC that drops China from the global sums of iron and steel production, primary energy consumption, total population, urban population, military expenditure, and military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

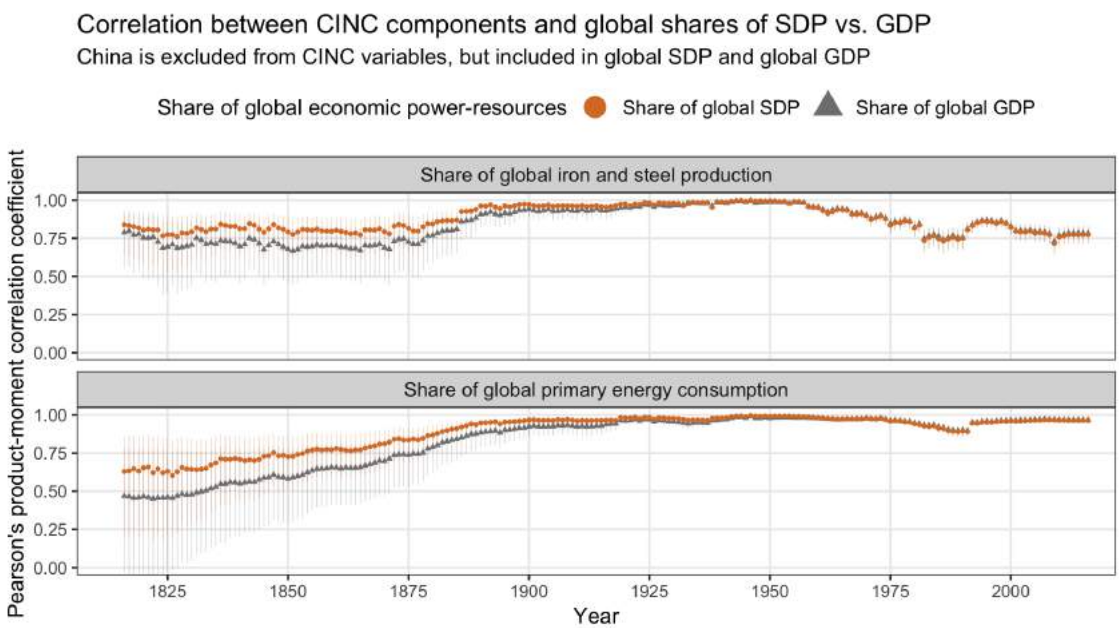
Our new measurement approach below makes population estimates available for a larger set of countries in the pre-industrial period and correct part of the bias resulting from the exclusion of units in CINC. For example, while CINC codes China as having 47% of the global population in 1860, our data code 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

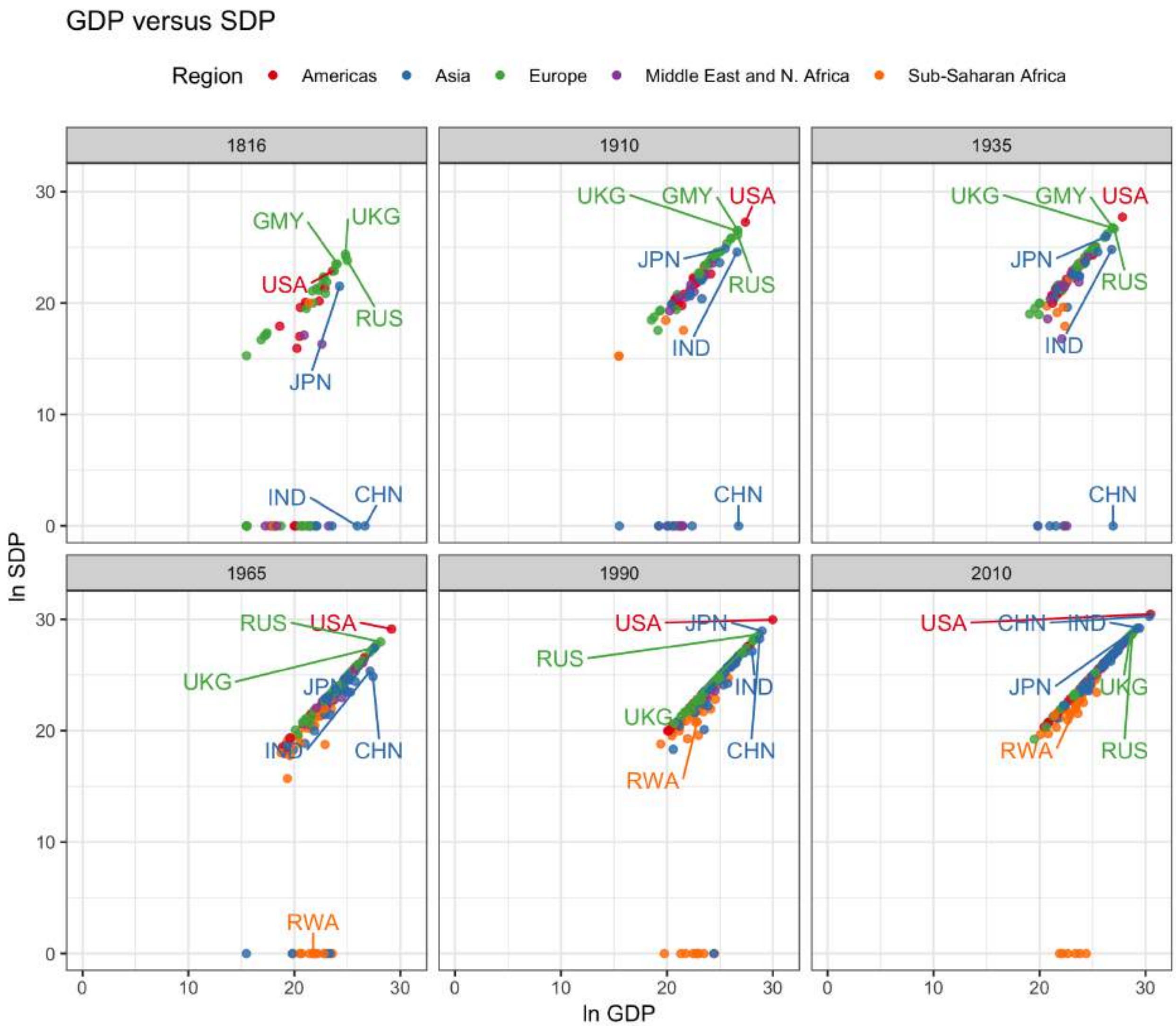


Figure 7: The graph plots the natural logarithm of GDP against the natural logarithm of SDP for select years. The plot shows that as time progresses and countries develop, SDP and GDP correlate highly. An exception are least developed countries, mostly in Sub-Saharan counties, who do not have a positive 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.

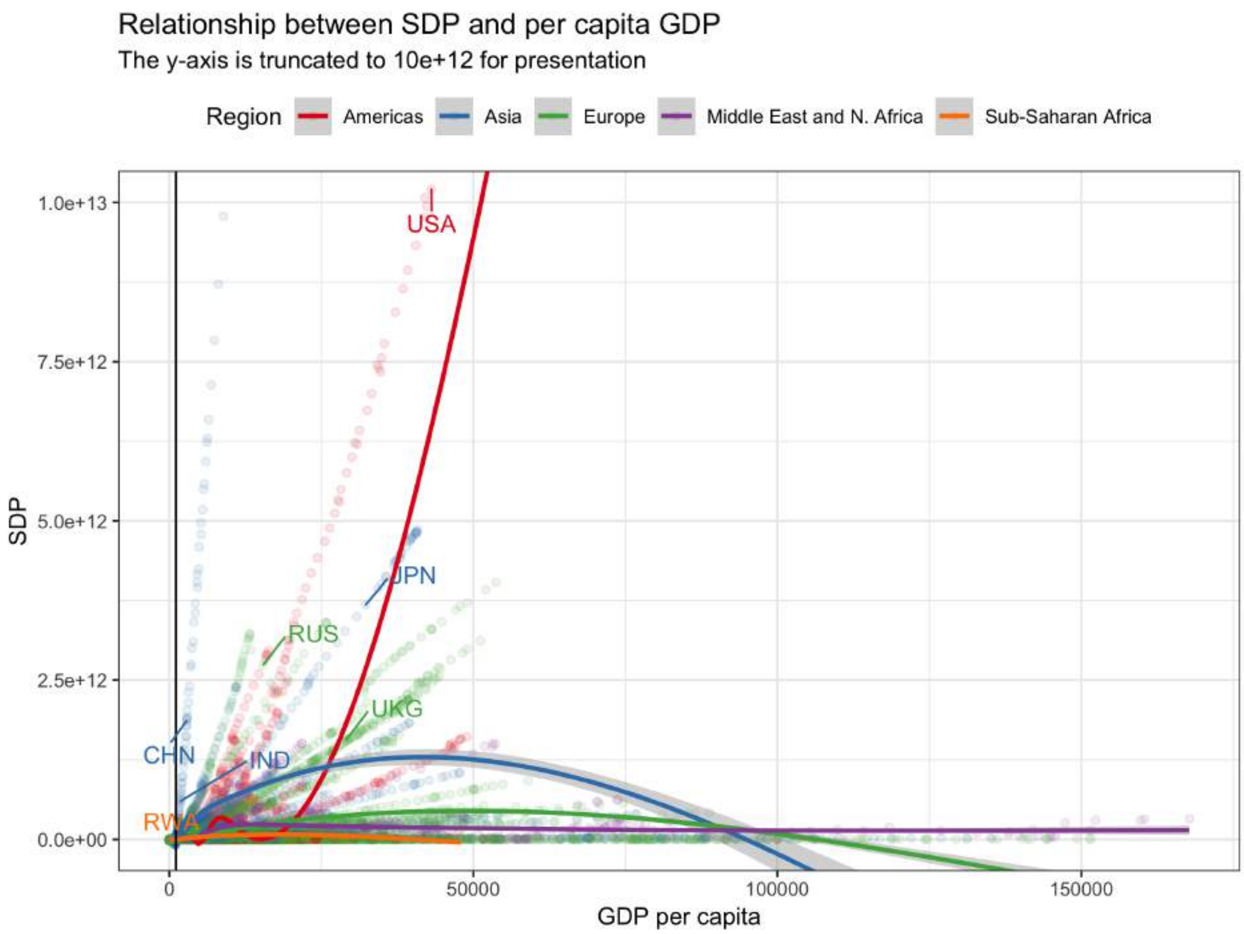


Figure 8: The graph plots GDP per capita against SDP across all country-years in the sample. The direction of the relationship between the two variables varies by world region. It is strong and positive for the most developed countries in the Americas (which includes the US and Canada). For most other 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 τ 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

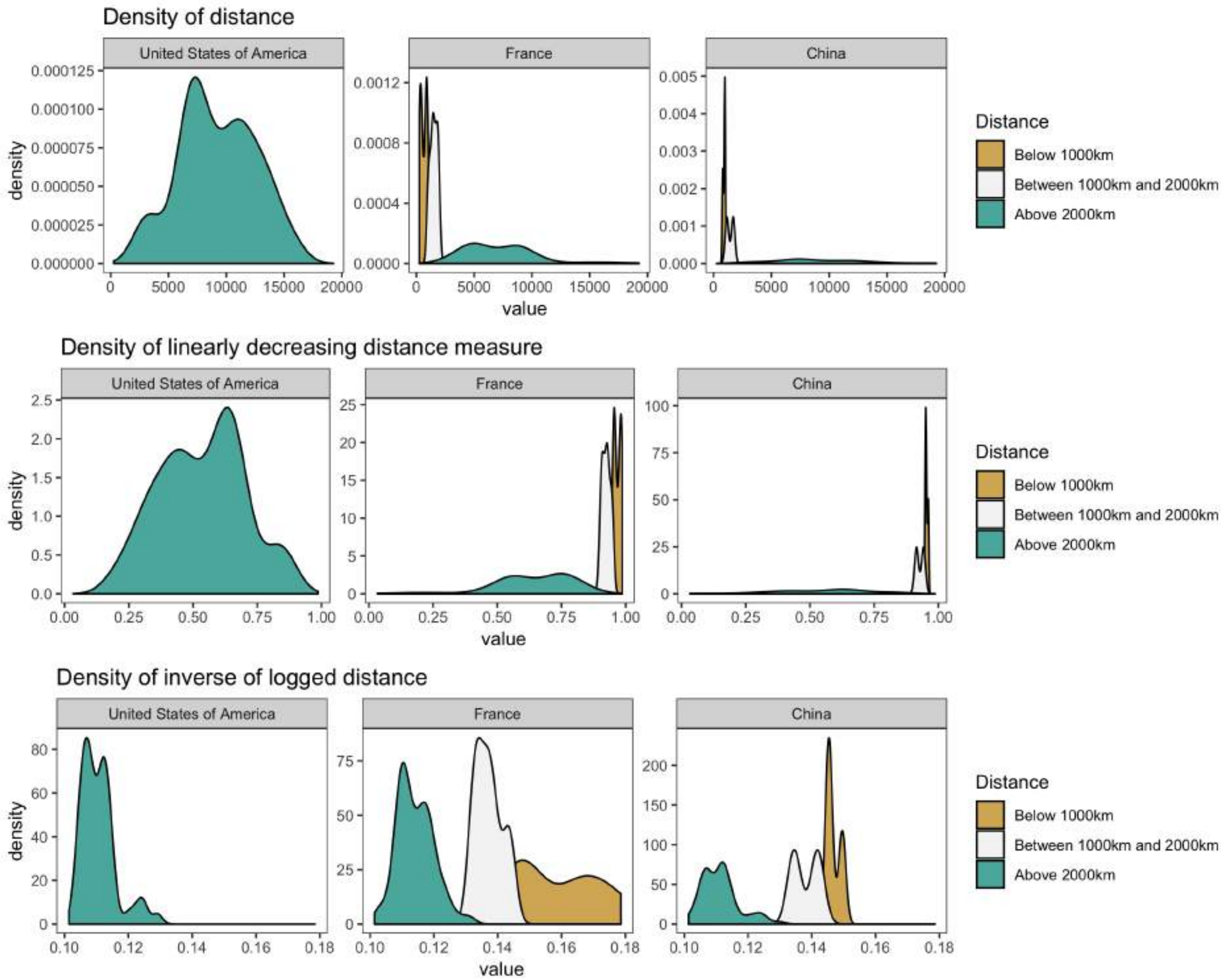


Figure 9: The graph compares the binning of countries by a) distance, b) a linear transformation of distance computed as $\frac{\max(\text{Distance})-\text{Distance}}{\max(\text{Distance})}$, and c) the inverse of the logged distance from the perspective of the United States, France, and China. Distance 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

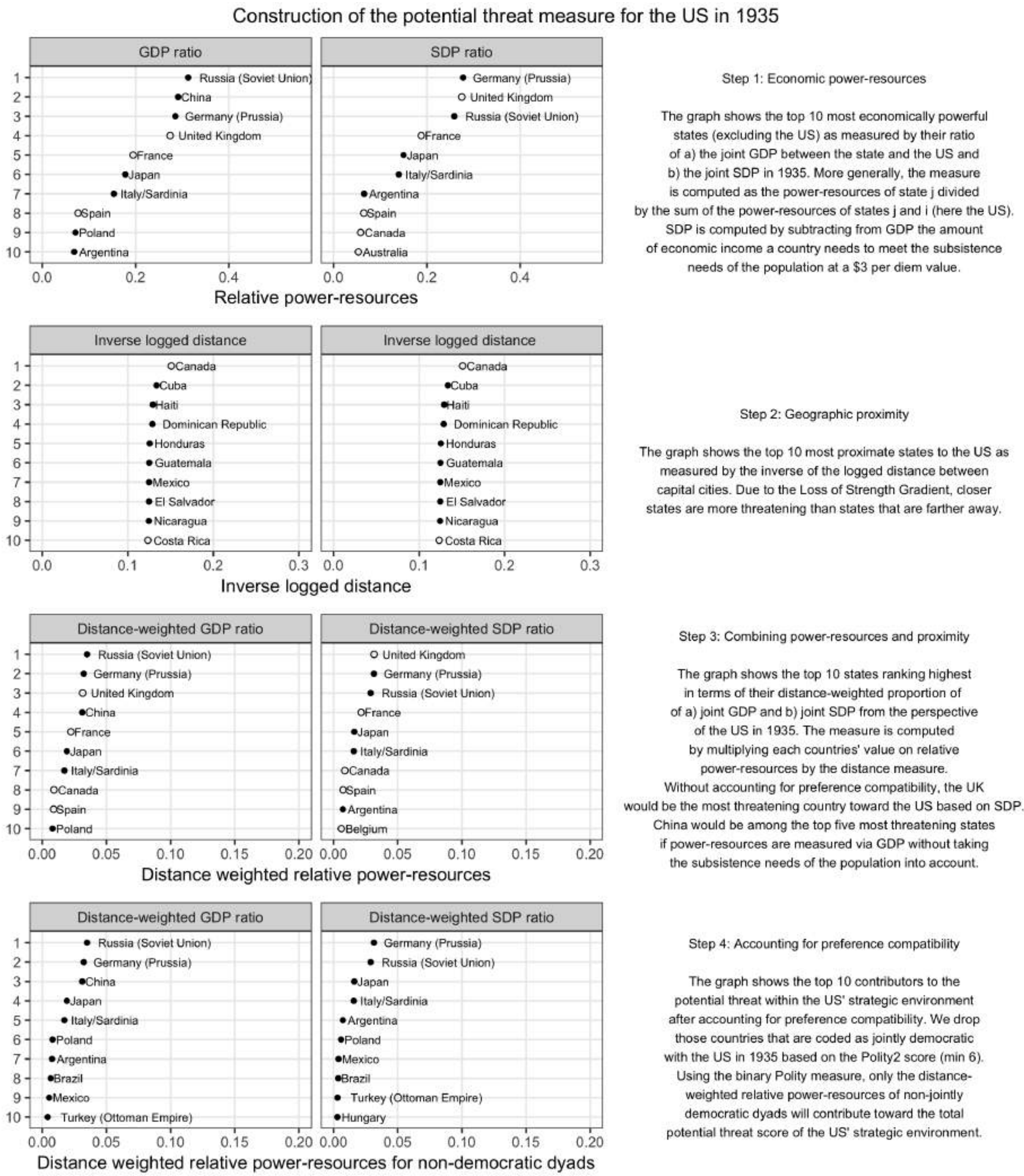


Figure 10: The graph illustrates the construction of the potential threat measure for the US in 1935. The SDP is based on a $3 per diem subsistence value.

E.3 Correlation between alternative potential threat measures

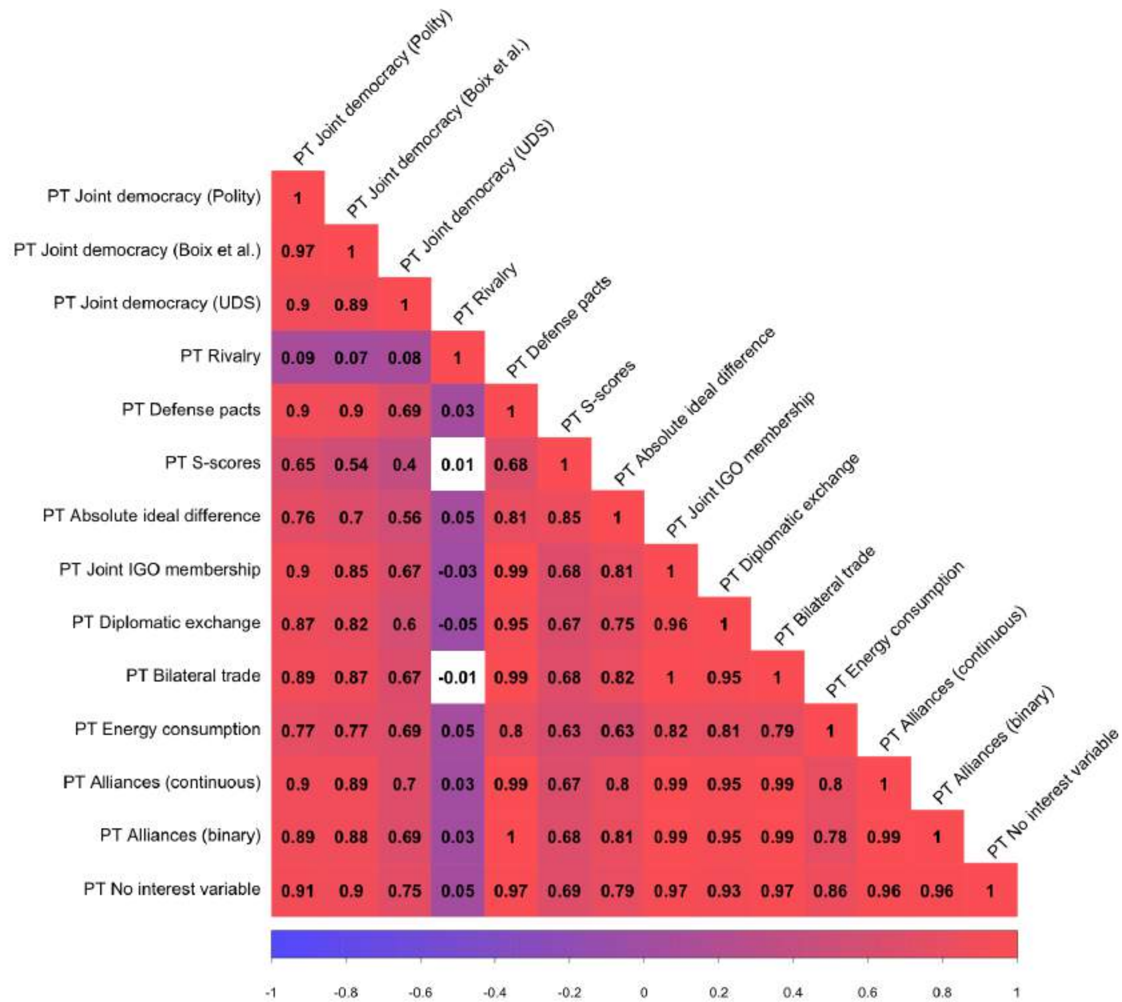


Figure 11: Correlation plot for alternative potential threat measures, using SDP ($3 subsistence level) to measure economic resources. Colored cells denote values that are significant at the minimum 5% level of significance. The plot demonstrates that all potential threat variables but one are statistically 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 κ) 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

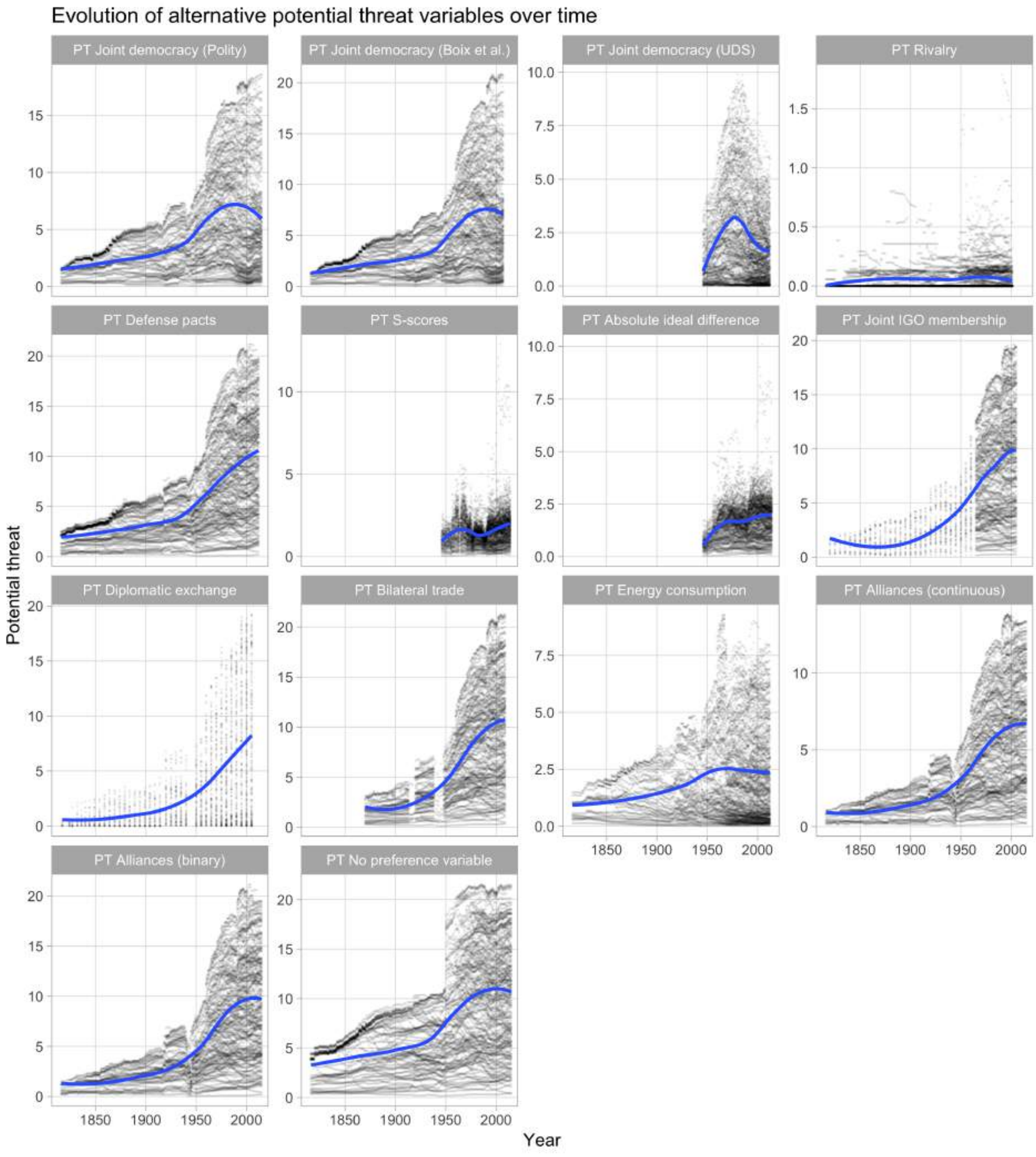


Figure 12: The graphs show the evolution of potential threat over time for alternative indicators of preference compatibility (see the caption of Figure 11 for data sources). Plotted is each country-year observation with the line denoting the loess smoothed trend over time across all countries.

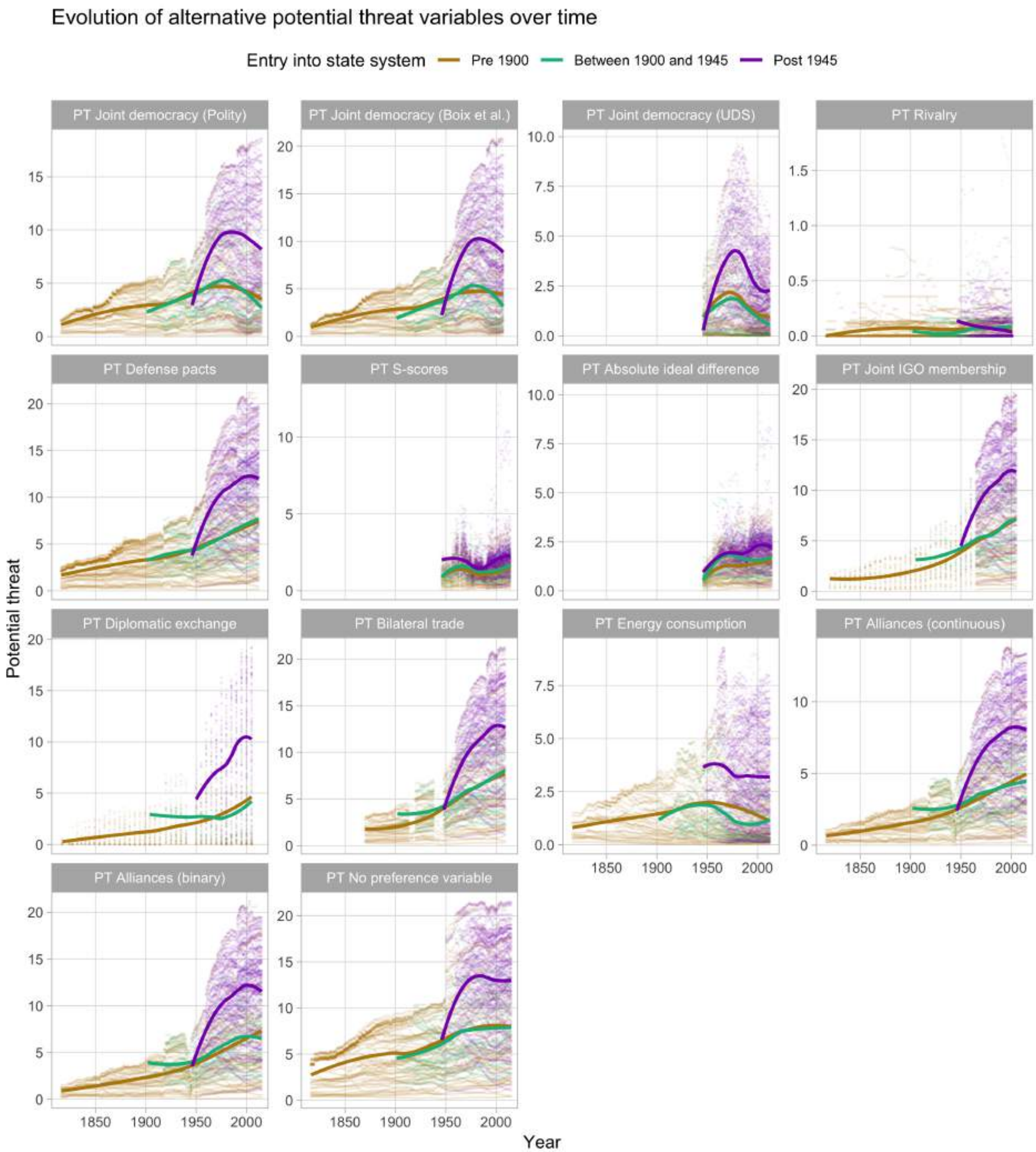


Figure 13: The graphs show the evolution of potential threat over time for alternative indicators of preference compatibility (see the caption of Figure 11 for data sources). Plotted is each country-year observation with the lines denoting the loess smoothed trend over time for three groups of states: 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.

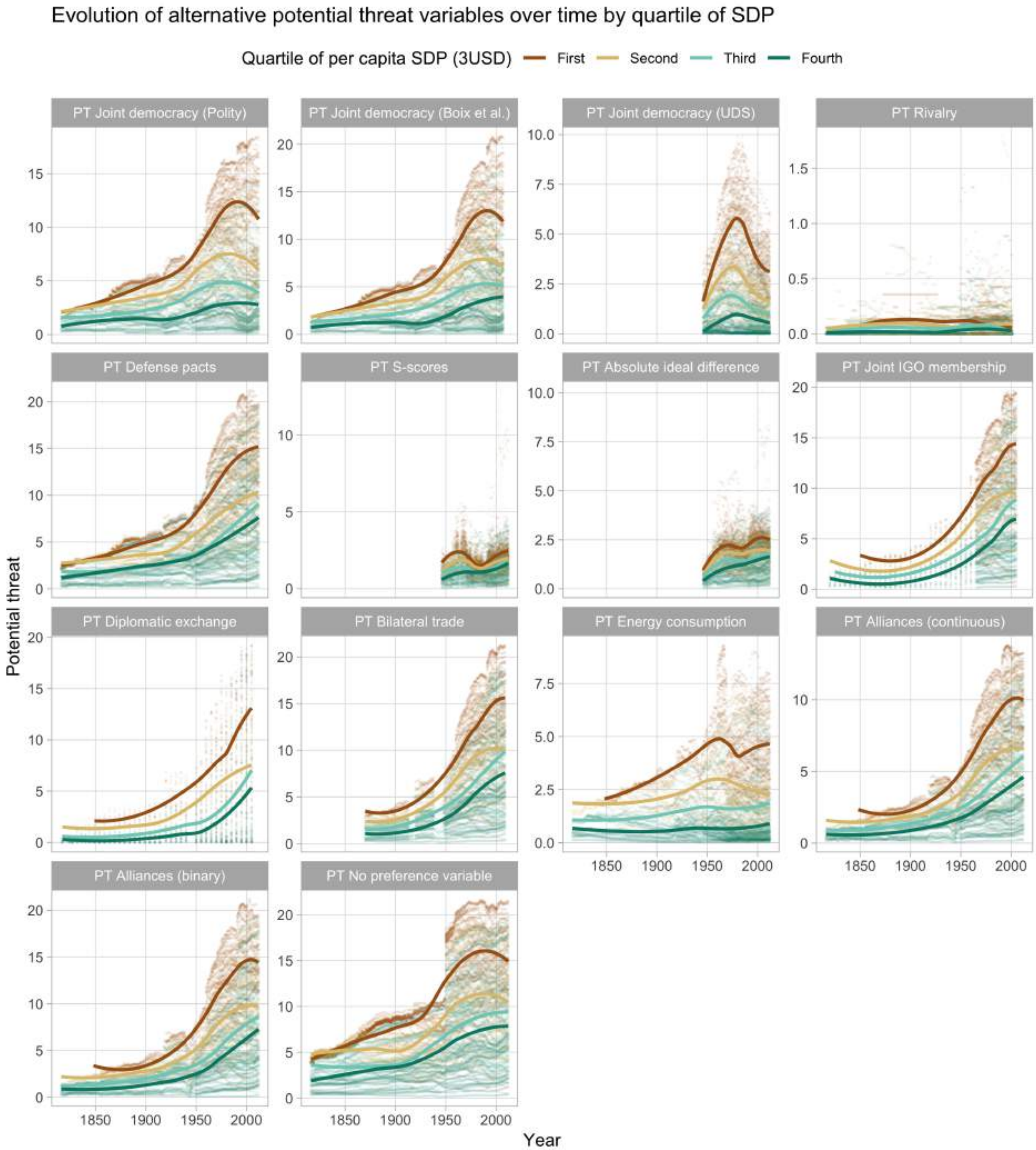


Figure 14: The graphs show the evolution of potential threat over time for each indicator of preference compatibility (see the caption of Figure 11 for data sources). Plotted is each country-year observation with the lines denoting the line of best fit over time for the first, second, third, and fourth 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

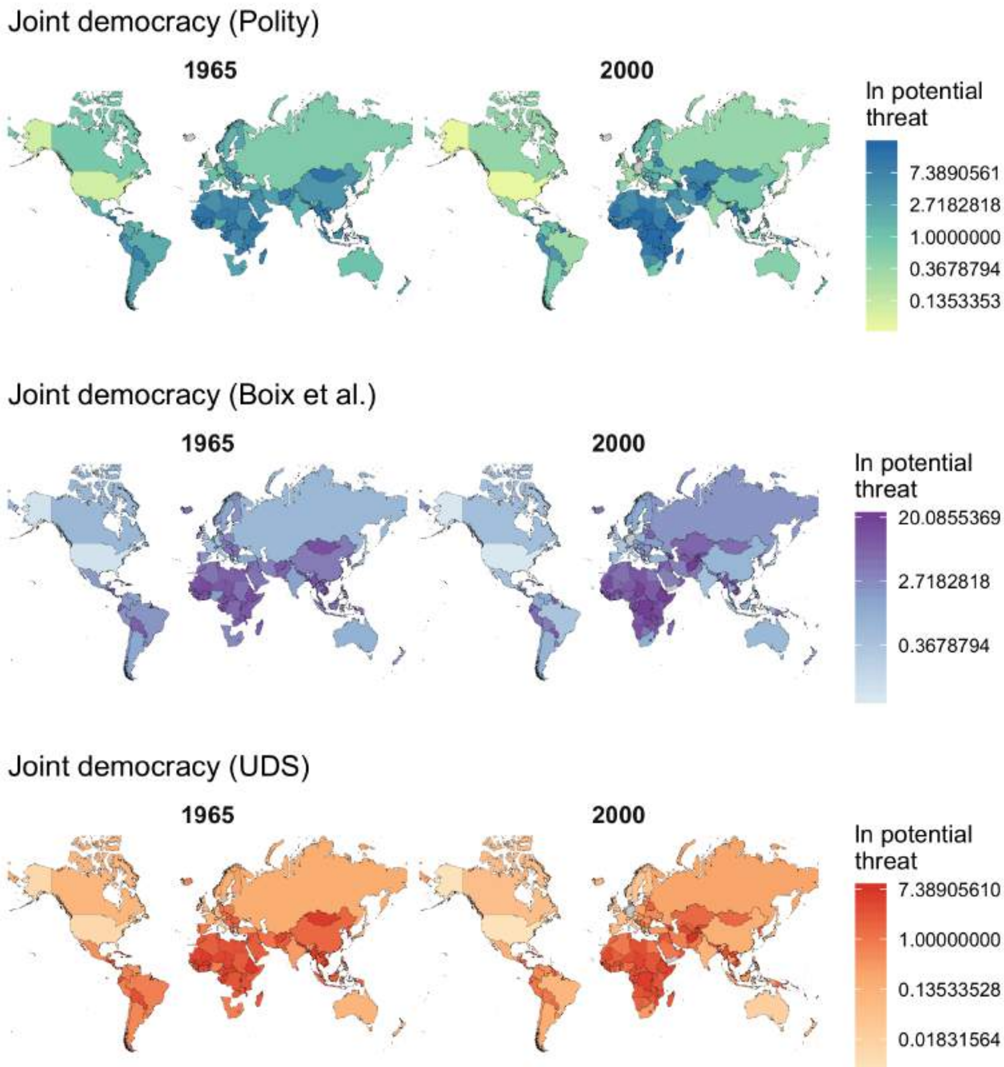


Figure 15: Maps plotting the spatiotemporal distribution of the natural log of the potential threat variable for the years 1965 and 2000. Potential threat is measured through joint democracy based on the Polity, Boix et al., and UDS scores, respectively. Power-resources are measured using the SDP indicator 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

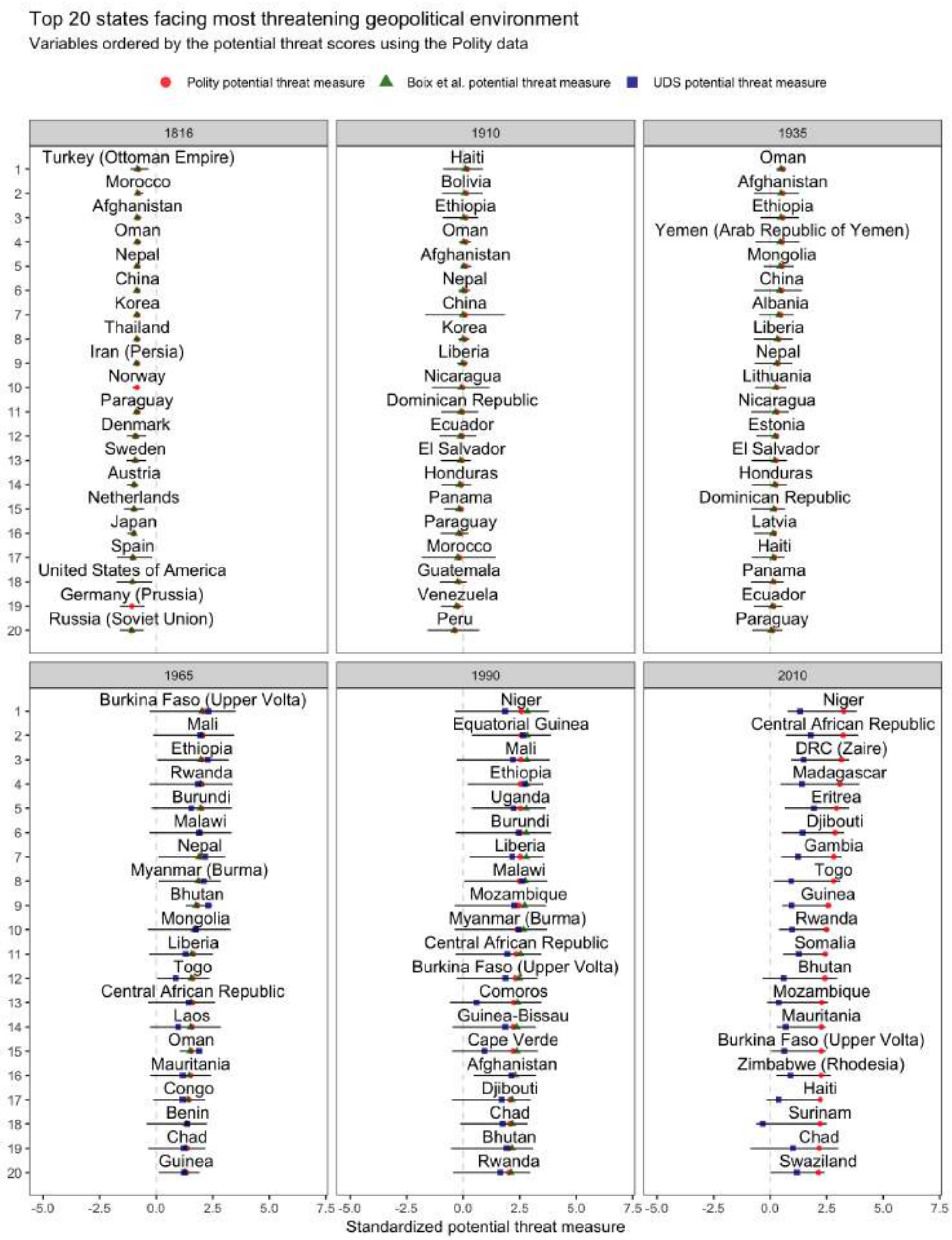


Figure 16: The plot shows the top 20 states facing the most potentially threatening strategic environment in 1816, 1910, 1935, 1965, 1990, and 2010. The red dots show our estimate of the total level of potential threat each country faces when using the Polity2 score to measure preference compatibility; 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

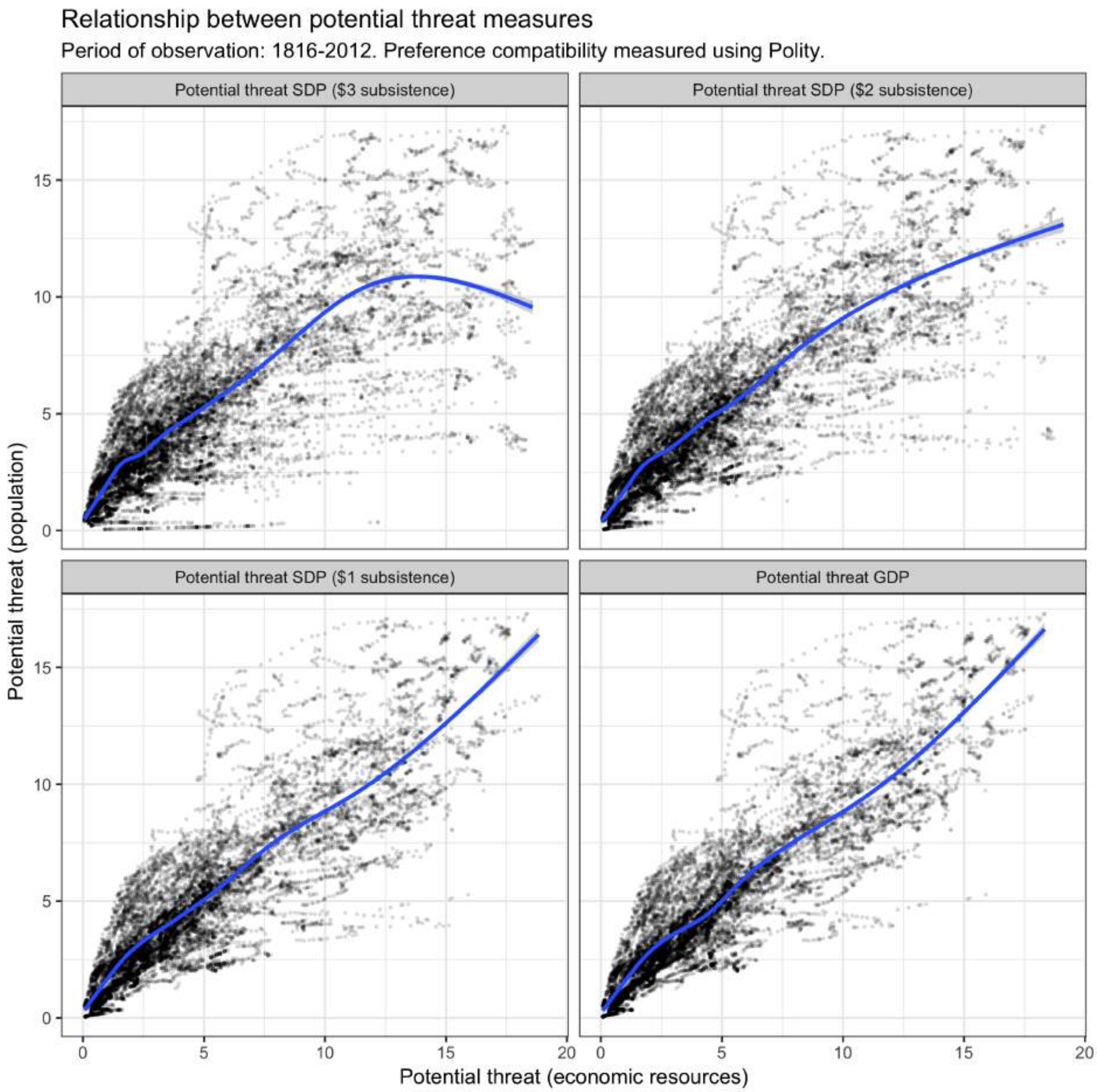


Figure 17: The graphs show the relationship between country-year values of potential threat based on population on the y-axis and potential threat based on economic resources for various subsistence thresholds on the x-axis. As the subsistence threshold increases, the strength of the association between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

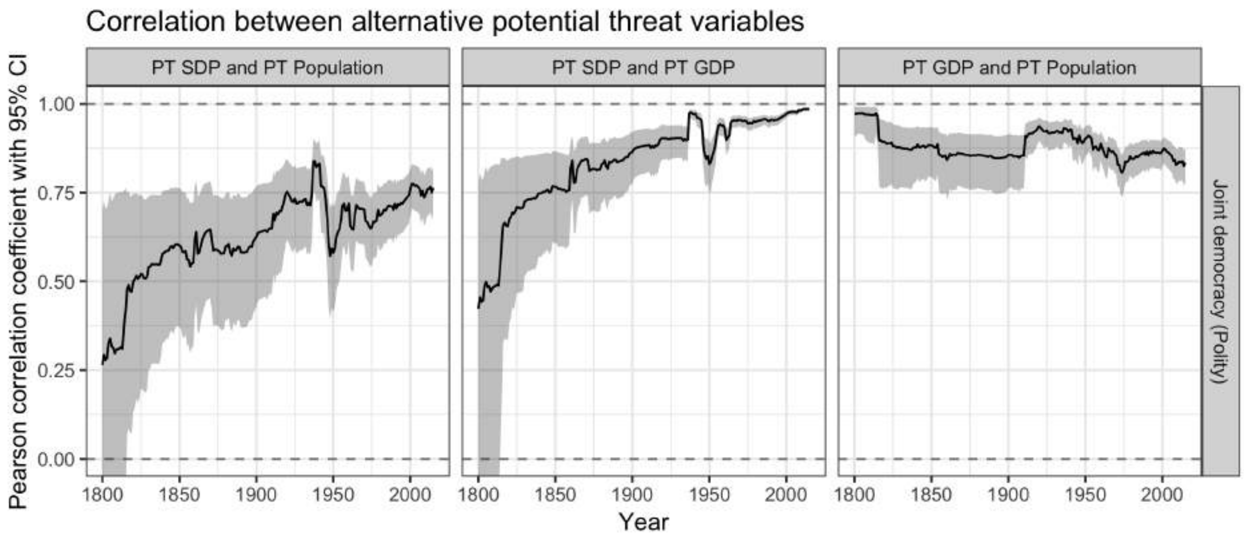


Figure 18: The graph shows the annual correlation between three alternative potential threat (PT) measures over time. We vary how power resources are measured across the three indicators: using GDP, using SDP ($3 subsistence threshold), and using population. For all indicators, preference compatibility 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

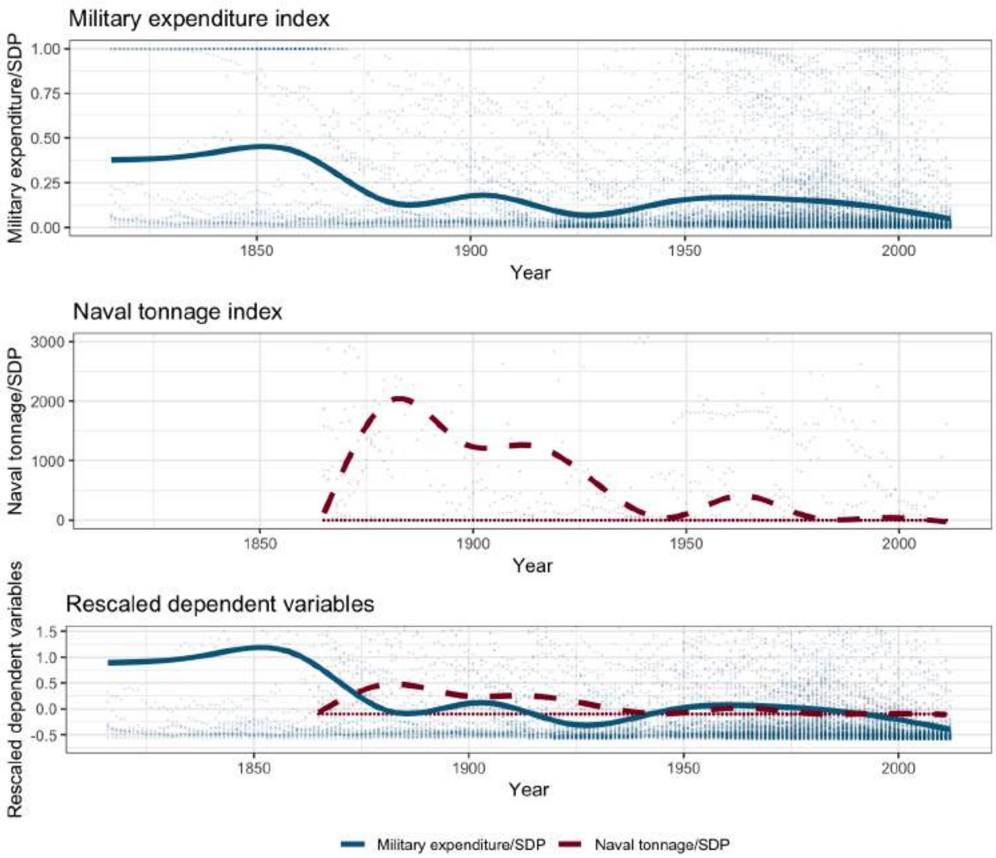


Figure 19: The two upper plots illustrate the temporal evolution of the military expenditure index (military expenditure/SDP in 2011 constant US PPP dollars) and naval tonnage index (naval tonnage/SDP in constant 2011 international PPP dollars), respectively. Dots denote the point values for each country 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

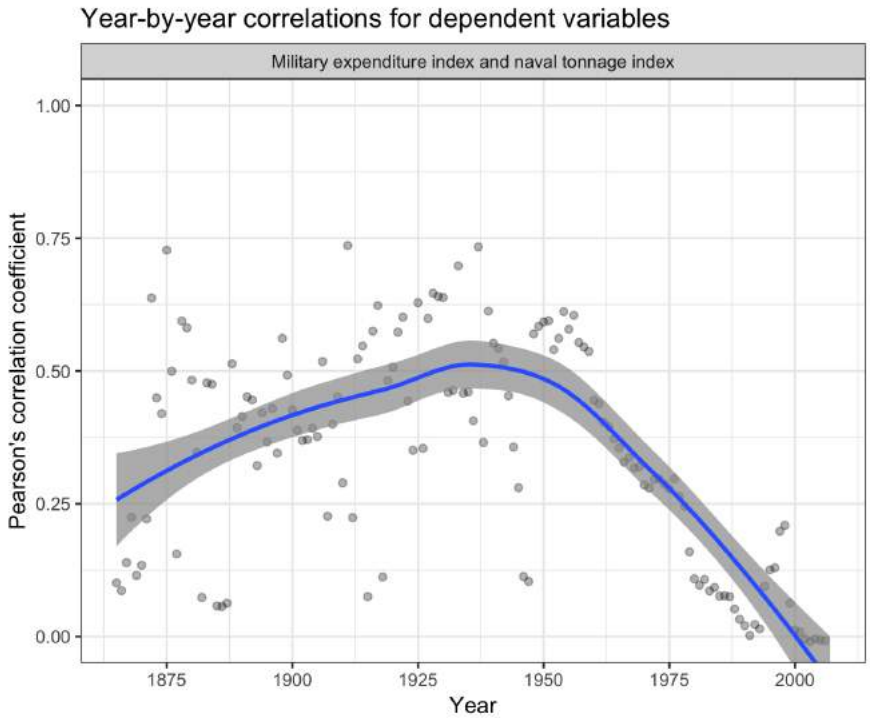


Figure 20: The plot illustrates the correlations between the two alternative dependent variables over time. Points denote the correlation coefficients for each year between 1965 and 2007. Lines represent the Loess smooth over those points.

F.3 Comparing individual countries over time

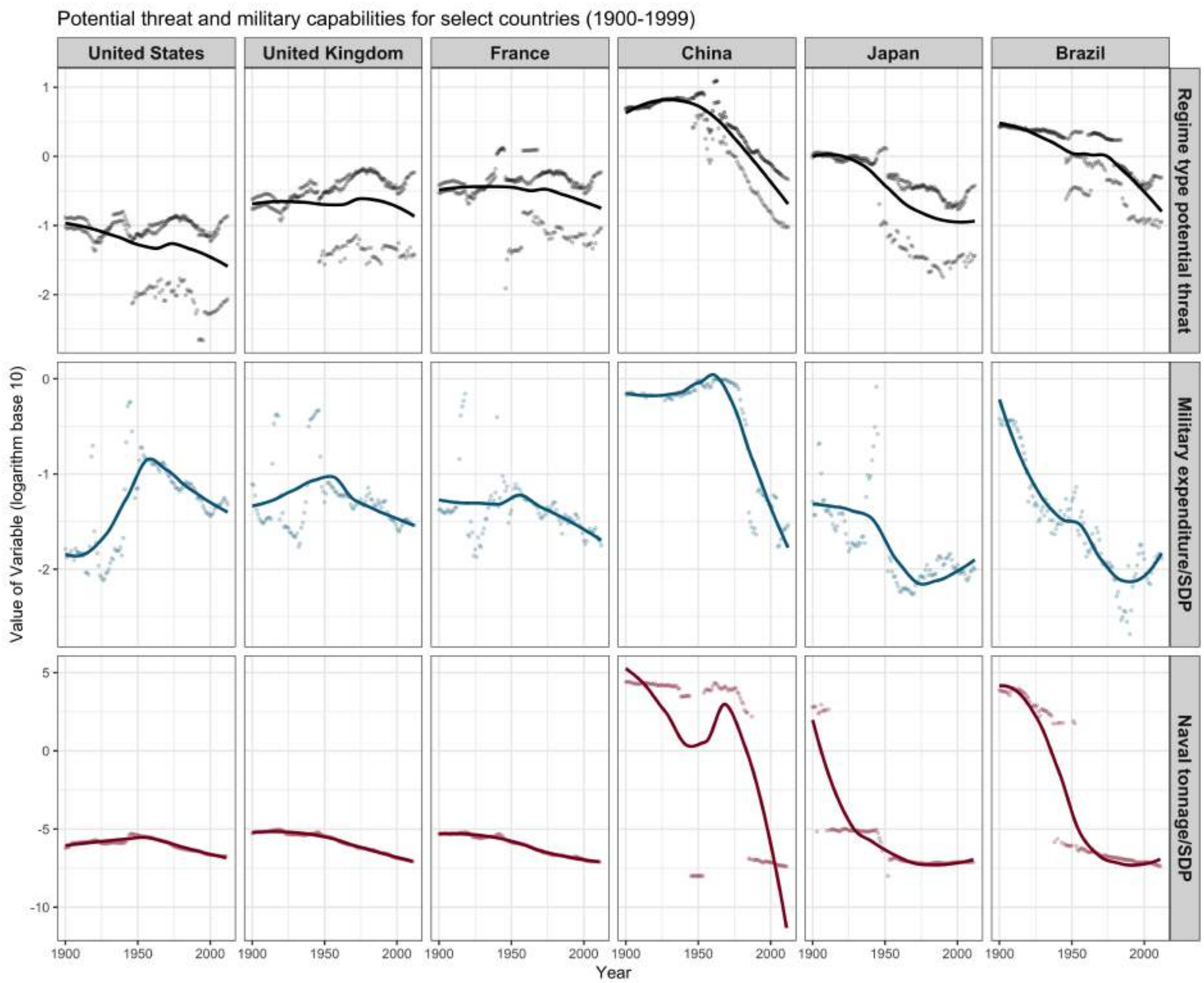


Figure 21: The plot demonstrates the ability of our measurement strategy to obtain scores for the level of potential threat that individual countries face at any given point in time (granted data availability). Plotted in the first row are economic resource-based potential threat scores using alternative 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.

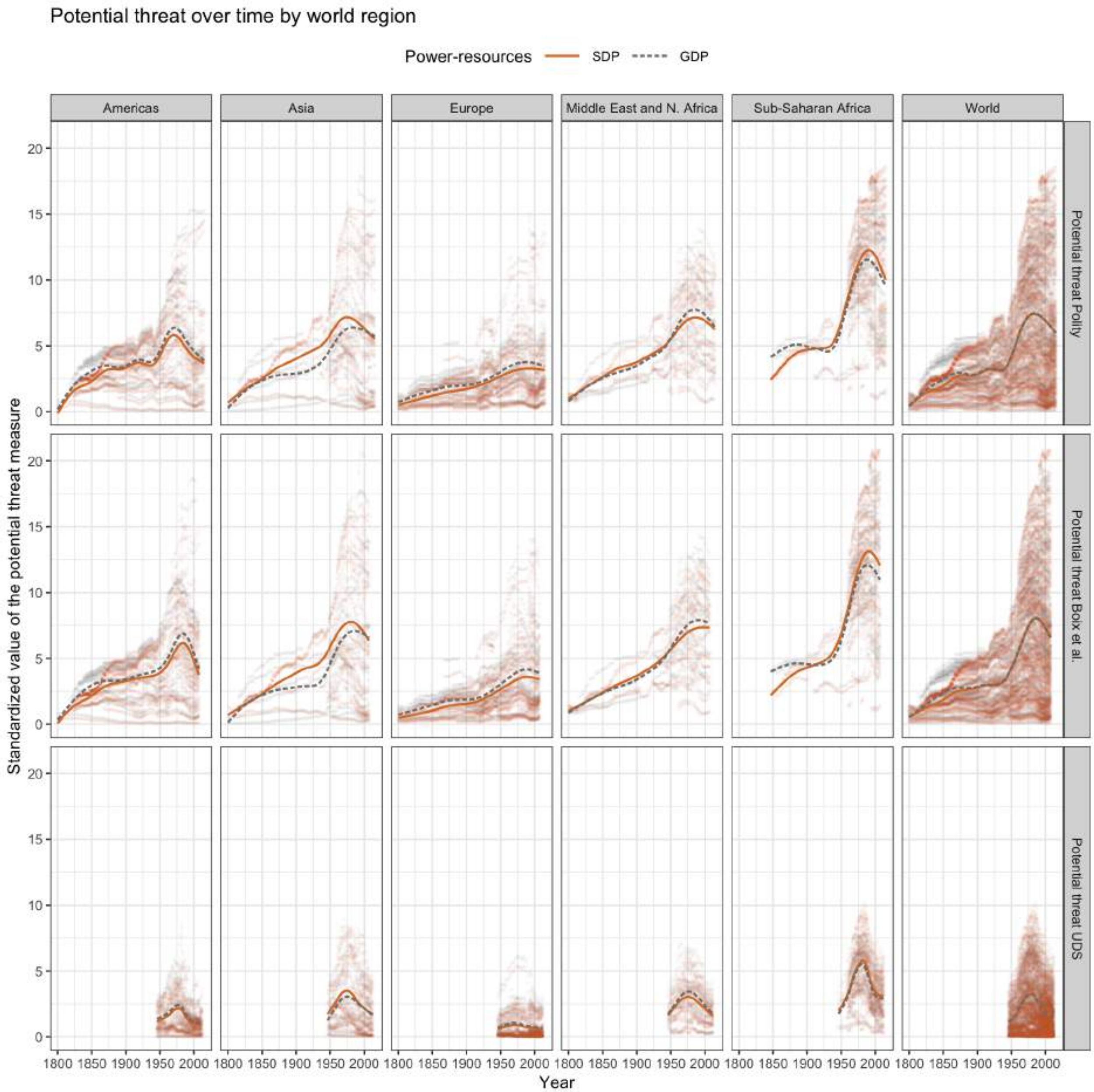


Figure 22: The graph shows the evolution of the potential threat across world regions for alternative measures of preference compatibility. The graph demonstrates that our estimates of potential threat differ depending on the measurement of power-resources. For example, using GDP we would lead us to 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

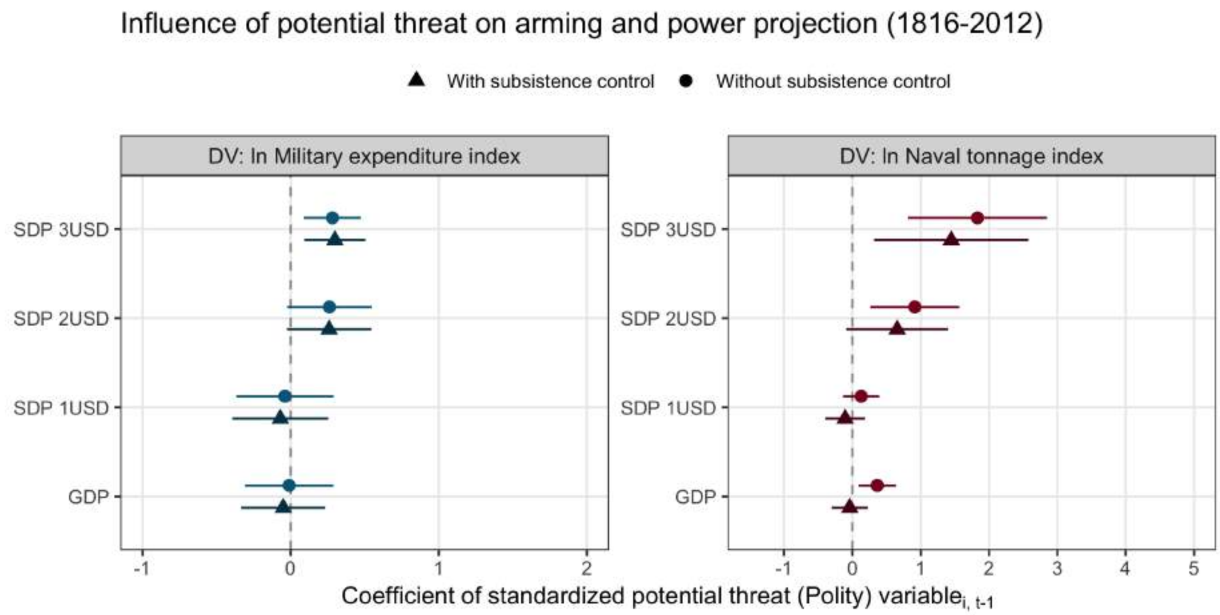


Figure 23: The graph plots the coefficients and 95% confidence intervals of standardized potential threat variables for regression models of two dependent variables — the military expenditure index and naval tonnage index — on potential threat and control variables. Preference compatibility is measured 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

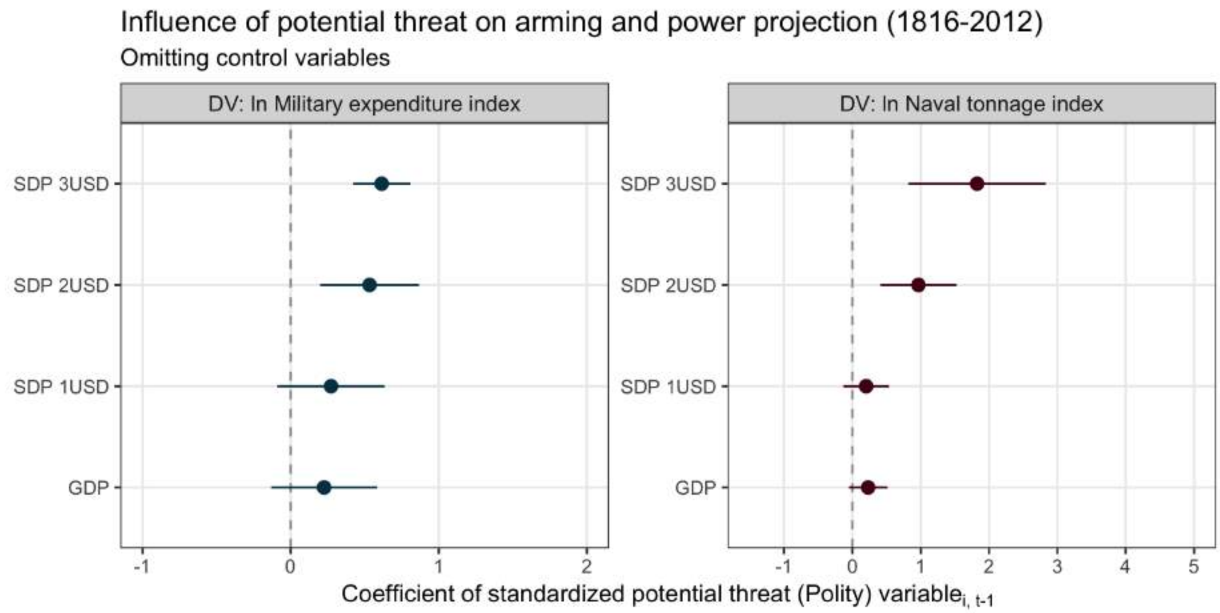


Figure 24: The graph plots the coefficients and 95% confidence intervals of standardized potential threat variables for regression models of two dependent variables — the military expenditure index and naval tonnage index. Preference compatibility is measured via joint democracy using Polity scores. 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

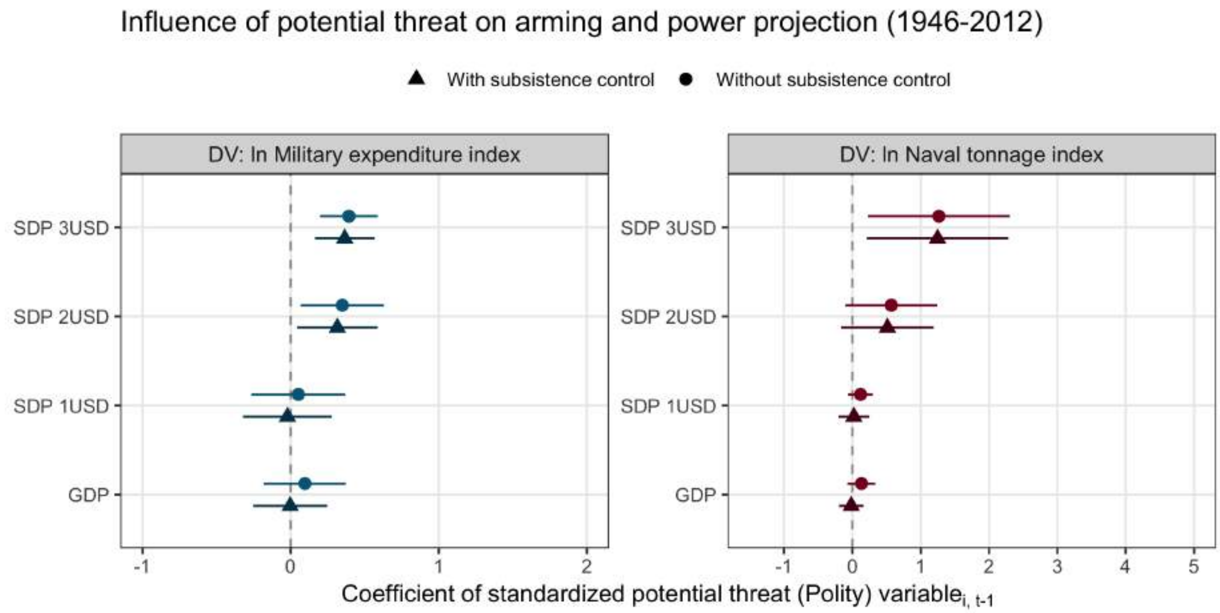


Figure 25: The graph plots the coefficients and 95% confidence intervals of standardized potential threat variables for regression models of two dependent variables — the military expenditure index and naval tonnage index — on potential threat and control variables for a post-WWII sample. Preference 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

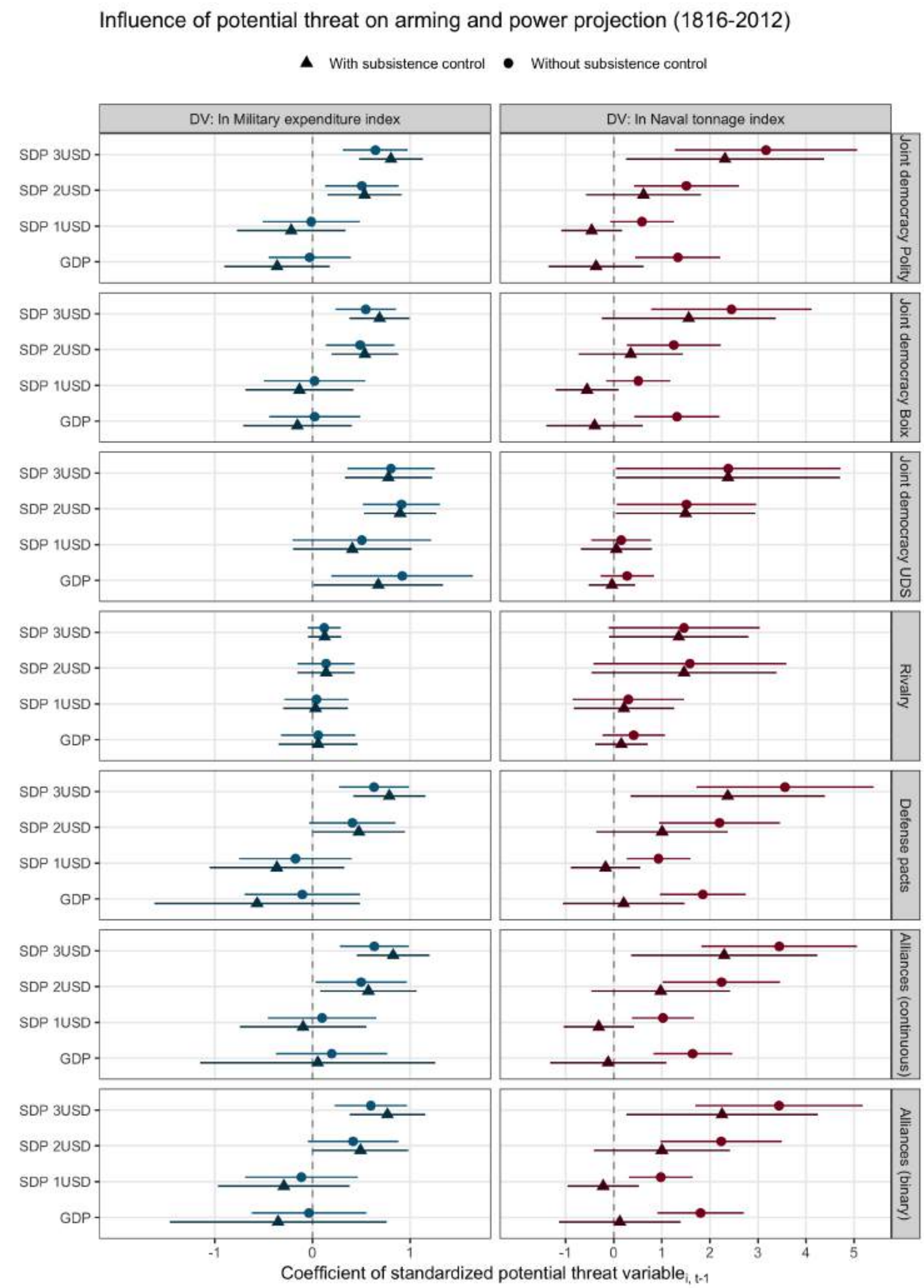


Figure 26: The graph plots the coefficients and 95% confidence intervals of standardized potential threat variables for regression models of two dependent variables — the military expenditure index and naval tonnage index — on potential threat and control variables. Preference compatibility is measured 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.

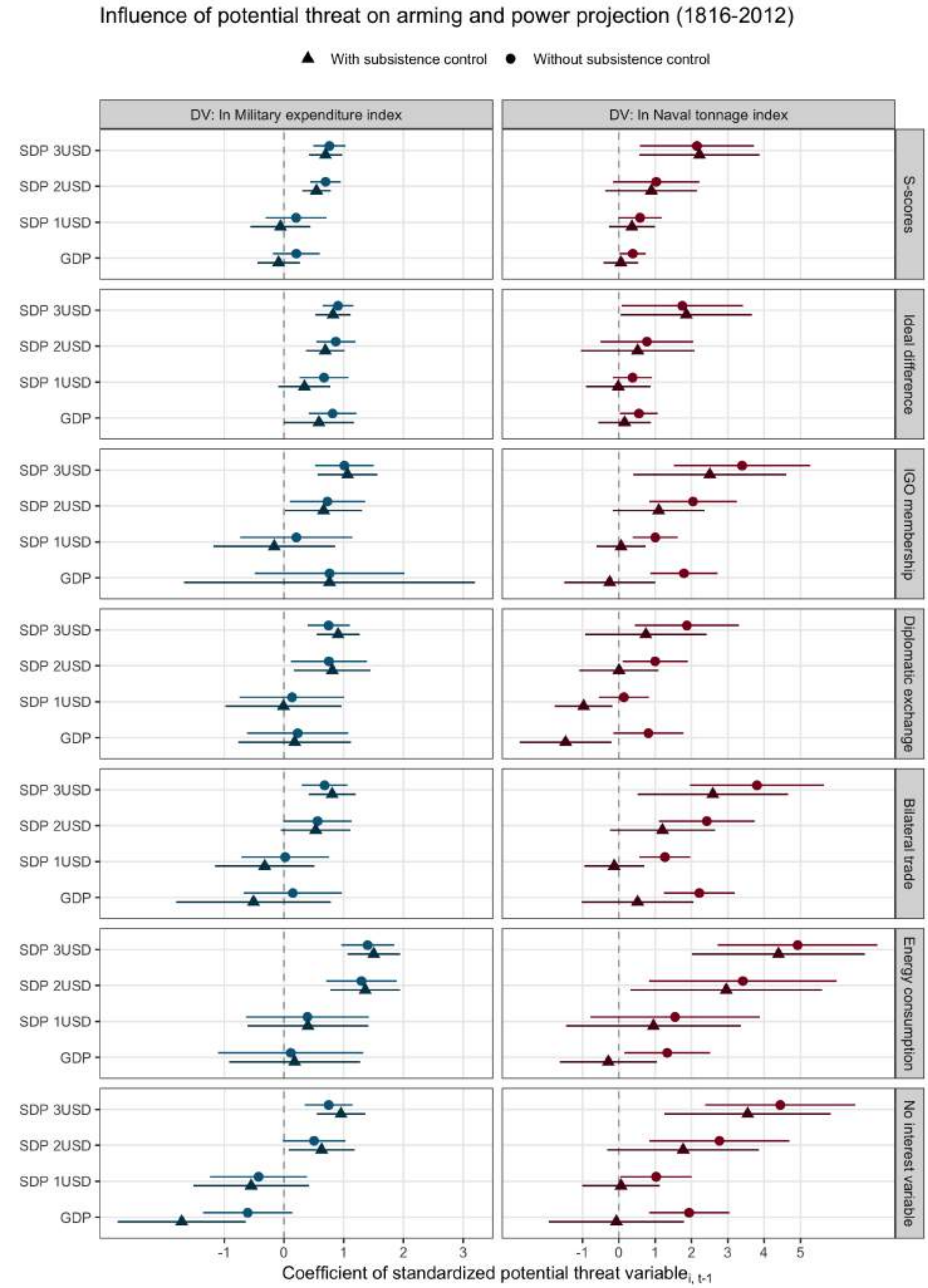


Figure 27: The graph plots the coefficients and 95% confidence intervals of standardized potential threat variables for regression models of two dependent variables — the military expenditure index and naval tonnage index — on potential threat and control variables. Preference compatibility is measured 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 $t = 1$, σ_{gdp} and σ_{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 τ_k: $\{\tau_{gdp}, \tau_{pop}, \tau_{gdppc}\}$. These τ_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 α_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	σ_{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	σ_{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”	α_j	$\sim$	$\mathcal{N}(\bar{y}_{itj}, 4)$
Model j slope “discrimination parameter”	β_j	$\leftarrow$	1
Model uncertainty for all GDP items	τ_{gdp}	$\sim$	$\mathcal{G}(0.001, 0.001)$
Model uncertainty for all population items	τ_{pop}	$\sim$	$\mathcal{G}(0.001, 0.001)$
Model uncertainty for all GDP per capita items	τ_{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.