



TRANSRESEARCH CONSORTIUM

WORKING PAPER SERIES

2021 number 2

The Strength of Nations: Constrained Indicators, and the Salience of Asymmetry in Conflict Relationships

Kelly Marie Gordell, William R. Thompson & Thomas J. Volgy

University of Arizona, Indiana University & University of Arizona

The **TransResearch Consortium (TRC)** is a registered 501(c)(3) nonprofit research institution. Its mission is to encourage in-depth research that leads to new ideas for solving problems facing society at the local, national and global levels. Focusing especially on impending transitions driven by power relations and demographic shifts, TRC is dedicated to increasing and enhancing understanding policy impact on political, social and institutional dynamics across alternative global futures. TRC is a global research and practitioner community with multi-university collaborative efforts, including Claremont Graduate University, Portland State University, and La Sierra University.

Working papers describe research in progress by the author(s) and are published to elicit comments and scientific responses contributing to constructive dialogs. The views expressed in TRC Working Papers are solely those of the author(s).

Introduction

Like a tsunami, the four years of the Trump administration caused incredible disruptions both in domestic and international politics. And unsurprisingly, it had challenged IR scholars to rethink their approaches to studying the relationships between states. Although the administration focused on “America first” and emphasized strengthening the state in its external relationships, between extensive personnel turnover and through the weakening of critical institutions in the name of “draining the swamp”, it had sapped American power substantially as state institutions frayed from incessant meddling by political operatives.¹ For IR scholars the Trump administration’s actions should serve as a compelling reminder that beyond the size of a state’s economy and the amount it spends on its military, the competence of its government is equally important for determining state strength and power. In this sense the lessons of the last administration cut to the heart of much of IR scholarship that centers on the utilization of state strength and power as explanations regarding interstate relations.

At the heart of international politics are varying conceptions of power and power relationships between states. Realists, structural realists, power transition theorists, hegemonic theorists, theorists of hierarchy and students of global leadership have all identified power relationship as crucial to an understanding of global politics (Kadera and Sorokin, 2003, Beckley 2018), although how these relationships matter have been the subject of considerable, unresolved controversy. For example, the debate over whether or not a relative balance of power between two states is more conducive to peaceful relations than a preponderance of power on the part of one of them is still unresolved (e.g. see Bremer 1992, Gortzak et al. 1997, Gartzke 1998, Geller 2000,

¹ For example, see Diamond, Rein and Eilperin (2021). While the assault on American institutions was extensive, highly evident were attacks on intelligence agencies, the State Department, and the upper echelons within the Department of Defense.

Lemke 2002, Lemke and Werner 1996, Moul 2003, Reed et al. 2008). The literature on rivalries has also generated unresolved issues about whether it is power equivalence or power predominance that facilitates rivalries (e.g. Vasquez 1996, Kline et al. 2006, Thompson 2001). For that matter, a perennial question of the US-USSR Cold War was which side had more power and therefore on “first base” in the rivalry. Something very similar is emerging in the ongoing US-China rivalry. At the system level, the corollary to the dyadic power preponderance question is over whether multipolar, bipolar, or unipolar power systems are more conducive for a peaceful and stable global political system.

Underlying all of these controversies are difficult questions regarding the conceptualization and measurement of power. In the quantitative literature, power is typically assessed as state capabilities (instead of the more nuanced conception of power) and typically measured in one of three ways: as economic size, as coercive capabilities, or, some combination of the two, as illustrated by the extensive and reflexive usage of the Correlates of War (COW) CINC scores.²

The purpose of this project is to demonstrate that an appropriate assessment of state capability (what we refer to as state strength) needs to go beyond what we describe as “raw capabilities”, and take into account factors that may limit or enhance the actual use of those capabilities. Doing so may then shed additional light on whether or not asymmetric power relationships are more conducive to limiting conflicts than balanced power relationships between states.

² There are of course several exceptions to this generalization. One is Lee and Thompson (2017) who use a complex measure of global reach in order to differentiate between global versus regional powers. See also Beckley (2018), Carroll and Kenkel (2019), Markowitz and Fariss (2018), and Moyer et al.(2019).

Imagine a contest between two warriors. One is very large and the other is small. Knowing nothing else we would expect the very large combatant to triumph. But what if the small combatant is more skilled in martial arts, more clever in general, or simply has access to better technology? What if the larger individual is ill, injured, or obese? The point is that we often need to look beyond raw capabilities and condition those base capabilities according to pertinent advantages and disadvantages in their employment. Toward that end, our work is organized in the following way: first, we discuss what we consider to be appropriate modifications to state capabilities; second, we apply these modifications and provide illustrations of how they change estimates of the relative strength of states and rivals, arguing that these modifications are more valid than raw capabilities; third we apply these modified measures to several aspects of conflict processes and outcomes in order to determine if using these measures creates better predictions. In fact they do appear to work better than relying on non-qualified capability information. This finding opens up a variety of possible applications, some of which are pursued in this paper, albeit briefly. We repeatedly find substantially stronger relationships between relative capability and conflict behavior when modified capability measures are employed instead of unmodified capability indices. The main implication is that if we think relative capabilities are important considerations in world politics, modified measures seem to provide more and better information on their comparative significance than the “raw” indicators do. Whether that claim holds up will have to await many more applications.

State Capabilities and State Strength

State capabilities³ are typically assessed in the literature one of three ways. One method is to calculate economic size as a measure of state capability. Typically, power transition theorists use this approach. A second approach calculates military assets, usually in the form of military spending by the state, although some calculate the size of armed forces or certain weapons capabilities (e.g. nuclear weapons or missiles). Realists and structural realists typically rely on this approach.

A third method and one that is the most often utilized in the quantitative literature⁴ generates a composite CINC score, based on six dimensions of capabilities that include total population, urban population, iron and steel production, energy consumption, military personnel, and military expenditures.⁵ CINC scores seek to integrate various aspects of national economies with military capacity to generate a composite measure of state capabilities. While this is a more comprehensive approach to state capacity than either of the first two, it suffers from a serious flaw: intended to measure state capabilities across a long time frame (1816-2012), it privileges certain dimensions of state capabilities that may have been historically important, but may no longer be so in the post-World War II era.⁶ As a result, CINC scores create substantial distortions from what scholars commonly believe to be extant state strength. For example, using CINC scores, China's strength surpasses that of the U.S. by 1995 (see Figure 1). We know of no IR scholars who would accept this claim; in fact most scholars would argue that China remains far

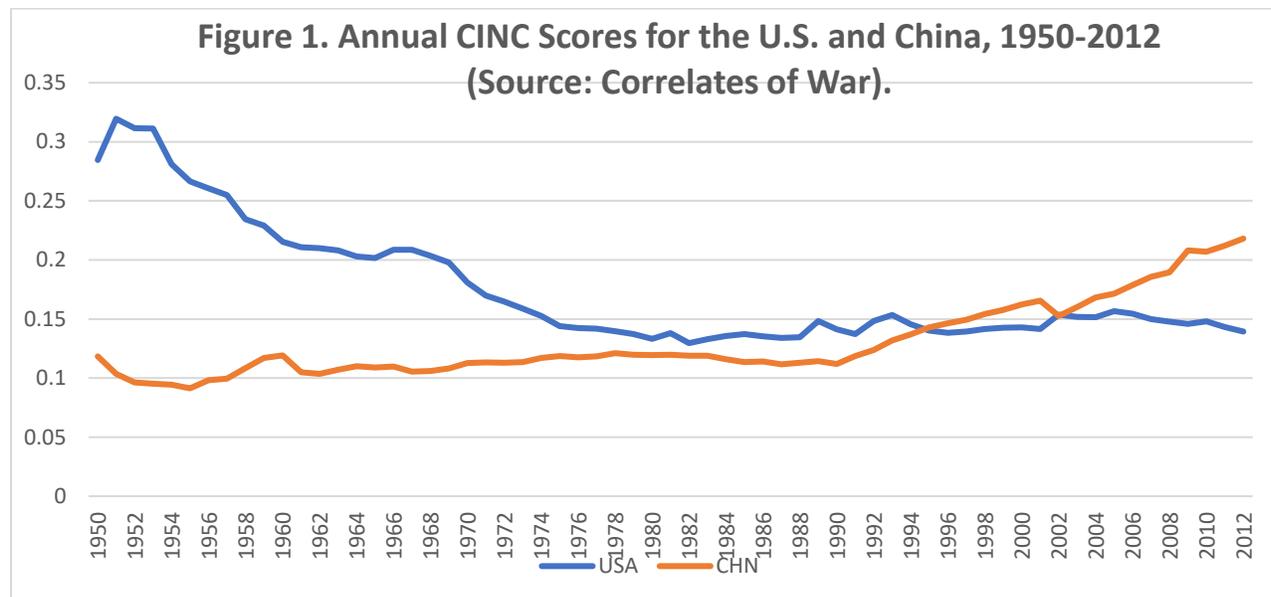
³ We avoid the use of the term "power" in what follows; we recognize that power is far more complex a concept than either state capabilities or state strength.

⁴ For example, Carroll and Kenkel (2018, p, 579) note that the overwhelming empirical international relations research in "top five journals" use CINC scores to measure capability ratios between states.

⁵ See The Correlates of War Project, at: <http://www.correlatesofwar.org/data-sets/national-material-capabilities>.

⁶ In particular, giving equal weight to the size of the population or the armed forces without qualification bestows great capability scores on states that have large populations in which most of the people engage in subsistence agrarian strategies for survival or military forces that possess a large number of under-armed and poorly trained personnel. Carroll and Kenkel (2019, p. 583) find that the best predictor of the outcome of conflicts among these dimensions is primary energy consumption, a finding which we discuss below.

behind the U.S. today in terms of military capabilities, the sophistication of its economy, its ability to govern, or its technological capabilities.⁷



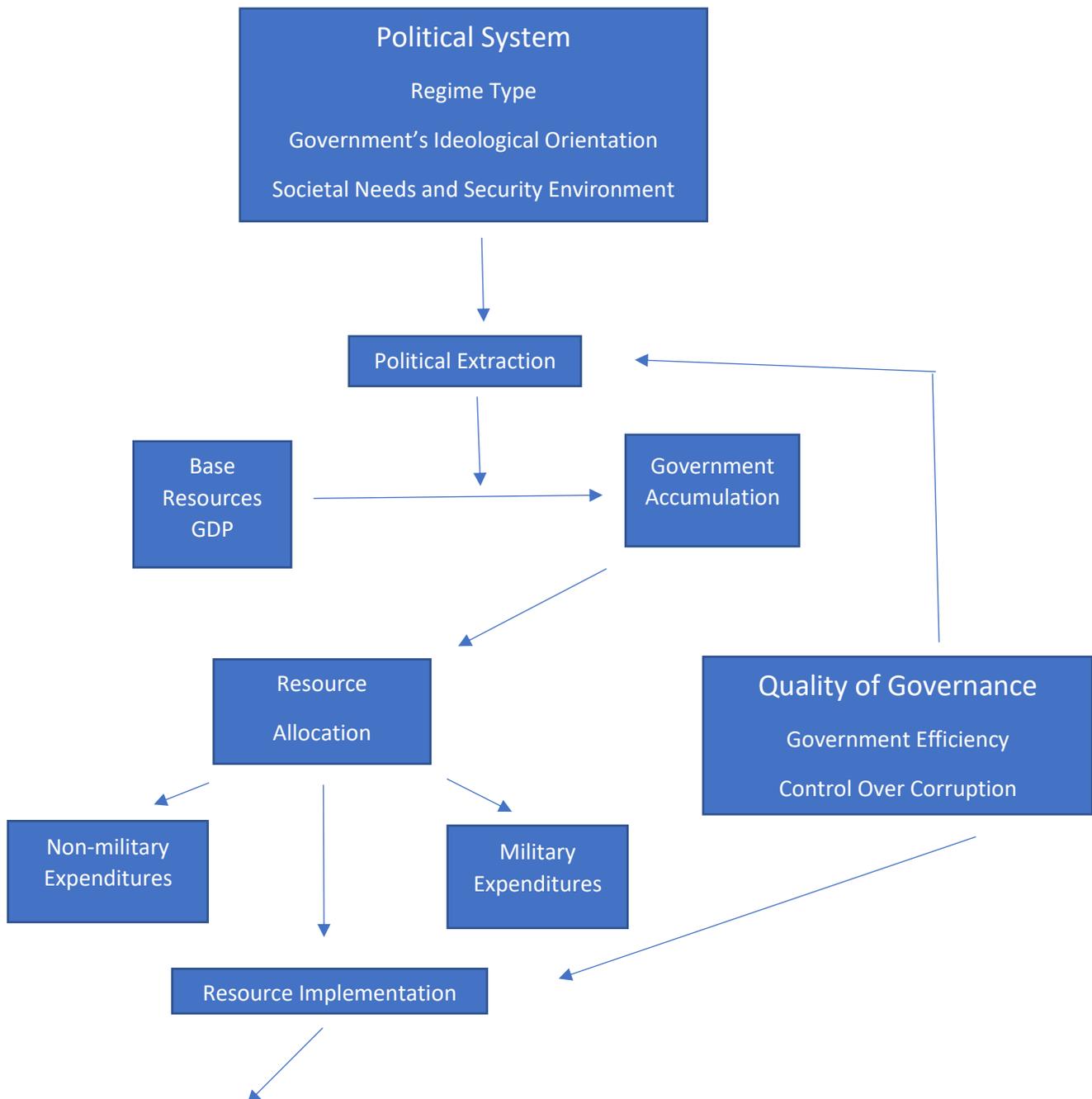
Nevertheless, aspects of the CINC approach do make conceptual sense for assessing state strength. We recognize that a state's economy is crucial to its strength. Weak economies typically do not allow policy makers to develop a strong military, a complex foreign policy infrastructure (which is quite costly),⁸ or the range of non-military tools that afford policy makers a broad set of non-coercive options (such as foreign aid) with which to pursue their objectives. However, focusing only on the size of a state's economy, measured as its GDP, is also problematic as a stand-alone indicator of state strength. We believe this to be the case for two reasons. First, it represents only the bulk size of a state's economy (often reflecting the size of a state's population

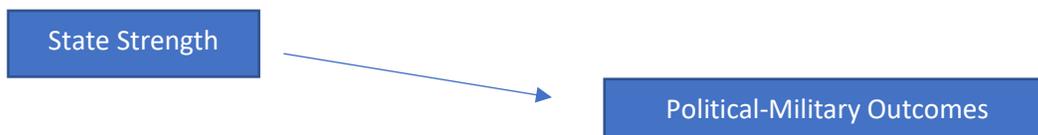
⁷ Even more perplexing, as Figure 1 shows, the use of CINC scores would indicate that China nearly demonstrates as much strength as the US in 1982, at a time of very limited Chinese capabilities. For one critique and a possible remedy for issues with CINC scores, see Kadera and Sorokin (2004).

⁸ For example, as both India and Brazil sought to dramatically increase their presence on the African continent as part of the strategy toward manifesting their "rising power" status, both states encountered major difficulties as the small size of their diplomatic infrastructure was unable to accommodate the number of ambassadorial personnel needed for the task (e.g. see Volgy and Gordell 2019). As of 2011, India's foreign ambassadorial corps was smaller than that of Singapore or New Zealand (The Economist 2011).

as well), and not the actual capabilities *available* for the state to apply to its foreign (or its domestic) policy interests and activities. Second, to measure state strength, analysts should require more extensive information about how such a raw capability might be employed or is being utilized.

Figure 2: From Political System to Political-Military Outcomes





In this sense our concept of state strength differs from simple state capability, such as the size of a state’s economy or its military spending. Both of those involve “bulk” measures of capacity. State strength, for us, is a more complex concept. It refers to three dimensions: one is the potential availability of resources for a state to use for policy purposes. This dimension is consistent with the idea of “bulk” capacity or resources. However, state strength requires two more considerations. The second is the actual *extraction* of some of those bulk resources from the economy for policy purposes; without such extraction policy makers cannot utilize those resources for the pursuit of their policy objectives. Third, the path from extraction to utilization raises the question of waste: strength is increased by the effective and efficient utilization of resources that have been extracted from a state’s economy. If substantial amounts of extracted resources are wasted or are inadequately developed or applied to policy objectives, state strength is diminished.

Figure 2 illustrates the linkages we propose between base resources and state strength. The figure also illustrates the general theoretical framework at the core of our approach. While we do not deny that dynamics external to the state (e.g., its external security environment) may impact on the willingness of state policy makers to extract resources from its economy, we suggest that the factors associated with both extraction and efficiency are critically dependent on its domestic political environment. Thus, we point to two domestic political factors⁹ that need to be

⁹ We are not suggesting that state strength is solely determined by factors within the state. State strength can also be enhanced externally, for example through territorial expansion (e.g. Russian take-over of Crimea, or the Chinese build-up of islands in the South China Sea), the infusion of massive amounts of military and economic aid, or

incorporated into any analysis that seeks to translate state capabilities into state strength. The first is the ability and willingness of the state to *extract* from its base capabilities the resources needed to advance its policy interests.¹⁰ Second, states vary extensively in their abilities to *effectively and efficiently* utilize the resources they extract for their policy objectives. These two factors constitute highly salient constraints or modifications on the raw capabilities indicated by the size of a state's economy.

As Figure 2 illustrates, the extent to which a state is willing to extract resources from its economy depends on a variety of conditions, although a large part of the extraction process is a function of the nature of the regime, the ideological orientation of the government in power, societal demands, and the nature of competition between elites in society. In both democratic and non-democratic political systems such extraction (and tax policy in general) is a basic political process that will vary with elite political beliefs regarding resistance to extraction by citizens and powerful interest groups, elite perceptions regarding the range of issues (including security challenges) needing to be addressed by the political system,¹¹ and elite assumptions about the effects of extraction on future economic growth and development. The extent to which resources should be extracted from the economy has been a key source of difference between Republican¹² and Democratic administrations in the United States; similar concerns have occupied policy

alliance relationships. For a discussion of third party involvement in disputes and alliance commitments, see Joyce et al. (2013). Nevertheless, the primary sources of state strength typically lie inside the state.

¹⁰ For this reason, some scholars use revenues rather than GDP values to assess state capacity. However valid general revenue comparisons across states are difficult to obtain over time and suffer from difficulties in estimating the extent to which centralized vs. decentralized political systems are comparable on this dimension.

¹¹ Our analysis of the data on political extraction created by Kugler and colleagues indicates that political extraction within democracies varies by whether or not there are left-wing or right-wing governments in power, while the extent of fluctuation in extraction is also correlated with democratic versus non-democratic regimes (the extraction "band" for democracies is substantially smaller than non-democracies; the latter operate across a far greater range of resource extraction).

¹² We note for instance that even the initially popular U.S. invasion of Iraq after 9/11 was not accompanied by the Republican U.S. administration's decision to extract substantial additional taxes from the American economy.

makers in China, Brazil, India, and the Russian Federation. As such, concerns about extraction vary over time across most states; a state's approach to extraction should be part of the calculus in assessing its strength.

However, even after taking into account how states differ in terms of resource extraction, the ability to efficiently and effectively utilize extracted resources should also vary substantially across states and across time. State bureaucracies (both domestic and foreign) differ from one another in terms of their efficiency in implementing policies and employing the resources at their disposal. Additionally, the extent to which a state's political system experiences substantial corruption should impact negatively on the efficient use of resources available to policy makers as they pursue their policy objectives. While we find that conventional measures¹³ of bureaucratic efficiency and control over corruption are fairly highly correlated, there is often a sufficient difference in the two measures to suggest that both of them should be included in any effort to gauge overall state strength.¹⁴

But why should we care about political extraction and quality of governance in assessing a state's economic strength if we have decent measures of military expenditures for states? We suggest two reasons. First, the economic strength of a state is critical for its ability to spend on its military. While numerous states may make efforts to increase their military capabilities above and beyond levels of spending that may be warranted by their economies, depending on their perceptions of security threat in their external environment, ultimately such spending is dependent on a state's economic strength, and "overspending" may have disastrous consequences, as Soviet

¹³ This point is based on our analysis of the relationship between the World Bank's indices of "government effectiveness" and "control over corruption". See <https://info.worldbank.org/governance/wgi/>

¹⁴ Thus, we are also assuming that quality of governance issues will impact both domestic and foreign policy in similar ways. We have found no valid and reliable longitudinal measures that address governance issues solely in the foreign policy arena.

spending in the 1980s suggests. In this sense economically strong states can also act quickly to reprioritize resources toward military objectives should the need arise. However, such reprioritization is likely also dependent on efficiency and effectiveness of the state's political system.

Second, economic strength remains important, separate from military strength since a state's strength is more than its military capacity. As Figure 2 illustrates, there are a large variety of non-military objectives pursued by states, needing substantial capabilities, and those capabilities are not reflected only in military spending. For example, even though the U.S. has been the richest of states in international politics, and its military budget has dwarfed that of any other state in the international system, it has found it increasingly difficult to provide sufficient funds to guarantee the safety of its embassies around the world. More recently, since the advent of the Trump administration, it has decreased spending on a range of activities outside of the military, including is diplomatic infrastructure.¹⁵

Regarding military capabilities, researchers typically use military spending as the appropriate measure. This indicator, by definition, already represents the political extraction of some resources from base economic capabilities and applied for military purposes. However, what it does not reflect are issues about its efficient use by state actors. Two states may spend the same amount on their militaries, but the one with a more efficient bureaucracy (both civilian and

¹⁵ Even former U.S. Defense Secretary Gates (2020) has argued that the U.S. has lost substantial power by cutting back on its diplomatic capabilities during the Trump administration, although this pattern had started prior to that administration.

military) is likely to generate a bigger bang for its resources, both in the production of hardware¹⁶ and the training of its military.

Likewise, depending on the level of corruption tolerated by the state, military spending may be wasted on non-military purposes, or, on outdated and ineffective equipment when there is substantial corruption. At its most extreme, high levels of corruption among civilian elites may even lead to the hollowing out of the military to prevent it from interfering with civilian authority,¹⁷ including the delivery of large amounts of personal benefits to military elites who support the government, all bundled within the state's military spending. Consequently, we suggest that a fiscal spending measure of military capabilities should be integrated with measures of bureaucratic effectiveness and control over corruption in order to reflect more realistically a state's military strength.

Comparing Indicators

Our approach to state strength calculates raw capabilities and modifies them with measures of political extraction and a state's quality of governance. We modify economic bulk (GDP) with both political extraction and quality of governance and modify military spending with measures of government efficiency and control over corruption.¹⁸ Do we generate better information about a state's strength through these manipulations? Applying these modifications (as we show below) results in assessments of state strength that produce substantially different results when comparing them to the original indicators.

¹⁶ As one example, Beckley (2018, p. 29) notes that in the conflict between Japan and China, while the Chinese may have had a sufficient number of hand grenades, 80 percent of them failed to explode...clearly an issue about the efficiency and effectiveness of military production in China versus Japan.

¹⁷ We are not the first to note these effects; see for example Chayes, 2015. For an example of how states can use corruption opportunities in other states to pursue their objectives, see Zelikow et al., 2020.

¹⁸ Note that we don't need to modify military spending with political extraction since military spending already represents one part of what the state has already extracted from its economy.

We look first at measures of economic bulk (GDP) and military spending and compare them with the alternative, modified indicators. For GDP we modify the base indicator in the following manner:

$$\text{Economic Strength} = \text{GDP} * \text{RPE} * (\text{Efficiency} + \text{Corruption}) / 2^{19}$$

RPE²⁰ is the difference between expected versus actual extraction of economic resources from a society's economy by the state. Efficiency is measured by the World Bank index of Government Effectiveness. Corruption is measured by the World Bank's Control over Corruption index.²¹

Since military spending represents resources already extracted from society, we only modify the measure by Efficiency and Corruption:

$$\text{Military Strength} = \text{Milspend} * (\text{efficiency} + \text{Corruption}) / 2$$

Data are restricted to the time-frame starting in 1996 since the World Bank measures are not available for earlier periods.

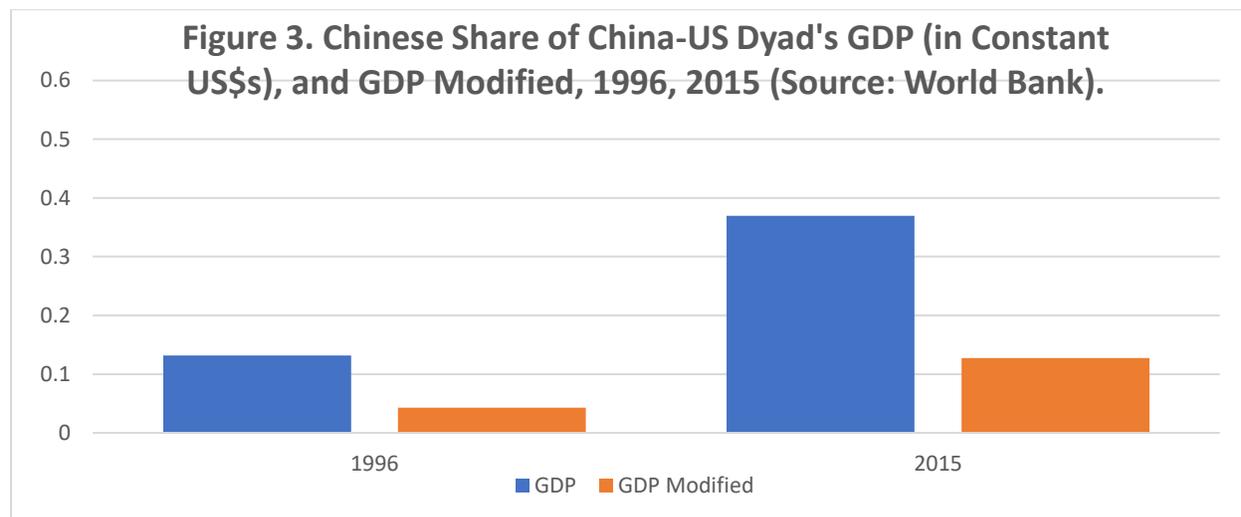
In order to illustrate the different outcomes using these alternative measures, Figure 3 shows the relative strength of the U.S. and China across two points in time, using base resources

¹⁹ Both here and for military strength we add together the two measures of Efficiency and Corruption and divide the result by two in order to obtain an average value for governmental performance, so that neither index weighs more than the other. We have no theoretical reason to weigh one dimension more so than the other.

²⁰ Relative political extraction refers to the extraction of resources from the economy by government, based on the difference between "expected" and actual extraction, controlling for level of development. For a discussion of the concept, measurement, and sources, see Kugler and Tammen (2012). RPE data are available at: <http://transresearchconsortium.com/data> For an extensive discussion and use of RPE, see Arbetman and Kugler (2018).

²¹ The World Bank's index of Government Effectiveness is designed to capture perceptions of the "quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies." Its measure of control of corruption is designed to capture "perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests." See <https://info.worldbank.org/governance/wgi/Home/Documents>. For a discussion and defense of the validity of these indices, see Kaufmann et al. 2010. The World Bank data are available at: <https://info.worldbank.org/governance/wgi/>

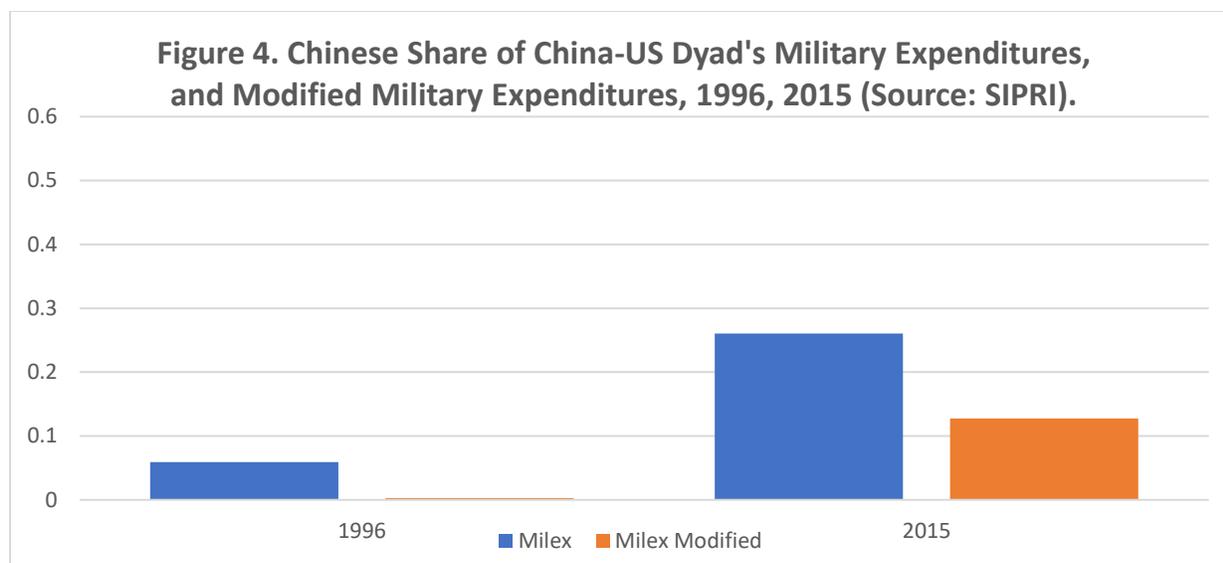
and the modifications placed on those resources. The base resource measure (GDP in constant 2010 dollars) indicates that Chinese GDP within the dyad grew from a bit over a 10 percent share in 1996 to a 37 percent share by 2015.²² The modified GDP measure reflects a dramatically different picture: this indicator shows the Chinese share at around 5 percent in 1996, and at around 12 percent by 2015, or roughly only one third of the share indicated by the base resource measure.



The difference between the military expenditures indicator and the modified version shows an equally dramatic difference between the two approaches.²³ As Figure 4 illustrates, US military expenditures dwarfed those of China in 1996; by 2015, the Chinese share of the dyad's military spending was at around 25 percent. The modified version of military spending (taking into account government efficiency and control over corruption) however, suggests a much smaller Chinese share of the dyad at around a twelve percent in 2015.

²² We use 2015 for these comparisons since the available RPE component of the measures cover states only through 2015.

²³ Military expenditures data are from SIPRI.



Does the modified version of our capabilities indicators constitute a more valid assessment of the relative strength in the Chinese-U.S. dyad? That jury is still out, although there is substantial evidence that China is both weaker and more fragile than it seems (e.g. Stephens 2019).²⁴ The list of concerns over China’s strength include: a) the difficulty of moving out of the “middle income trap” (World Bank 2012), reflected in its slowest growth in over a quarter century (da Costa 2019); b) the departure of over \$1.2 trillion in capital over the past decade (Harada and Manabe 2019); c) the souring of foreign investors (Bird 2019); d) extensive corruption especially among the elite (Barboza 2012, Jennings 2018); e) its high profile Belt and Road Initiative mired in what Stephens describes as a “swamp of corruption, malinvestment, and bad debt” (see also Standish 2019); and f) a ruling elite that may be substantially out of its prime.²⁵

²⁴ An unobtrusive indicator of government effectiveness is hinted at by problems that have been encountered in measuring the size of China’s GDP in the first place. Wildau (2019; Chen et al, 2019) attribute a 12% exaggeration of GDP size and a 2% overstatement of GDP growth to a lack of capacity to correct for local, inflated figures.

²⁵ Note Stephens’ (2019) description: “[the formal celebrations of] the 70th anniversary of the People’s Republic looked like something out of the late Brezhnev era: endless military pomp and gray old men.”

These different estimates resulting from modifications to the base measures are not restricted to major powers. We provide a second example from the ongoing rivalry in the Iranian-Israeli dyad. Figure 5 focuses on these states' relative economic capabilities. As the figure illustrates, raw GDP data suggests that the Iranian economy was substantially larger than that of Israel, both in 1996 and 2015 (67 and 63 percent shares of the dyad, respectively). However, the modified GDP measure indicates that Iranian GDP was roughly the equivalent of Israel's in 1996, and by 2015 was only 40 percent of the dyad's.

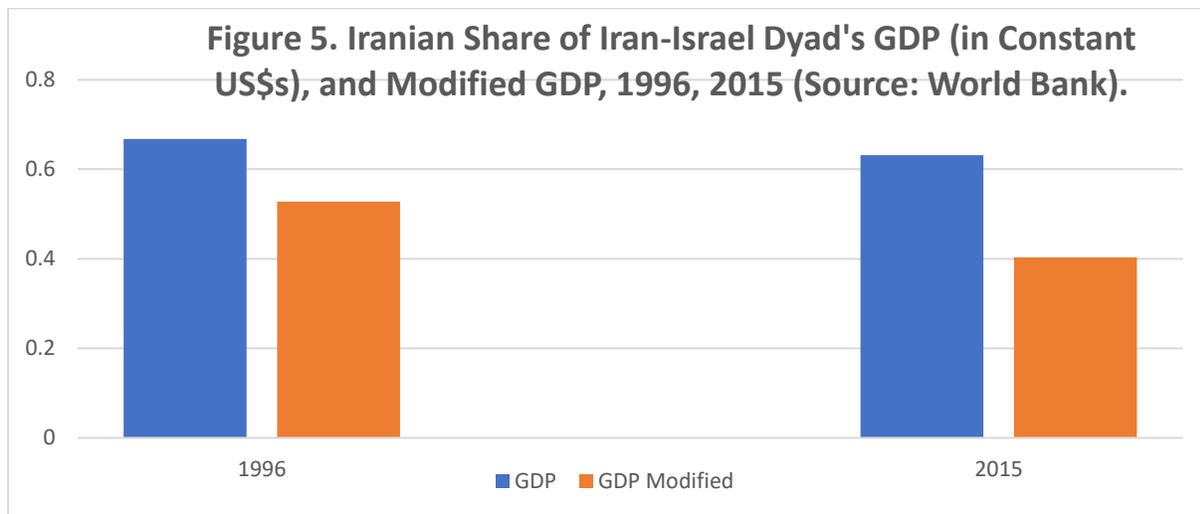
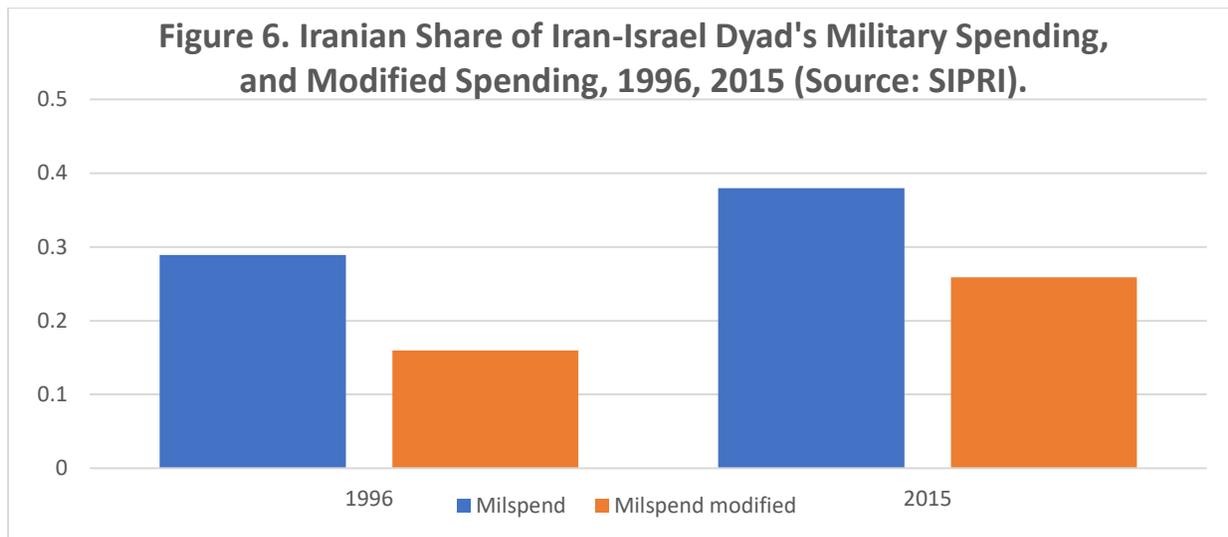


Figure 6 indicates shares of Iranian military capabilities in the dyad. A simple measure of relative military spending for 1996 shows an Iranian share of less than 30 percent, with the share increasing to nearly forty percent by 2015, and beginning to edge toward a rough parity between the two states. The modified military strength measure however shows an even more imbalanced picture favoring Israel: it indicates Iranian capacity at roughly 15 percent in 1996 (showing dominant Israeli military strength) and while this measure also shows increasing Iranian capabilities, they barely exceed 25 percent in 2015 (still showing dominant Israeli military strength). These data are consistent with the nature of Iranian-Israeli interactions; despite deep

expressions of Iranian hostility towards Israel, Iran's military activities toward Israel have been manifested almost exclusively by proxy, while Israel has directly attacked Iranian capabilities in Lebanon, Iraq, Syria, and inside Iran²⁶ without a reciprocal, direct response from Iran.



As these cases indicate, relative state strength appears to be dramatically different, depending on whether or not basic or modified capabilities are utilized to assess these relationships. We have not listed here some two dozen other cases of dyads—both within and outside of rivalries—that had led us to a similar conclusion. All too frequently and unsurprisingly, dynamics associated with political extraction, government efficiency, and control over corruption appear to qualify capabilities when assessing state strength. It would appear to be the case that these modifications create indicators that are more valid measures of state strength than the unmodified measures of GDP and military spending.

In Search of a Proxy for Government Effectiveness and Control of Corruption

²⁶ For a most recent case, see Fassihi et al., 2020.

There is, however, a problem with our measurement strategy: while we suggest that the information provided on government effectiveness and system-wide corruption are essential to the identification of state strength, their use is restricted by data availability. The World Bank measurements are available only for year starting in 1996. This is a severe limitation since these measures cannot be used to compare interstate relationships across a larger time horizon, and consequently do not allow researchers to compare between Cold War and post-Cold War eras.

Therefore, we have searched for a measure that a) would be more broadly available, and b) is highly correlated with the dynamics involving both effectiveness and corruption. This search was manifested in an inductive exercise seeking to yield a high empirical correlation between those variables and a theoretically appropriate proxy. The proxy candidate we propose is GDP per capita.²⁷

GDP per capita (GDP/capita) is typically used as an indicator of wealth within a state. While it is a measure of economic performance, it is nevertheless linked to quality of governance. Higher levels of wealth are correlated with a state's ability to provide public policies and the implementation of public goods efficiently, while control over corruption creates substantial amounts of predictability for economic activity that, in turn, leads to greater wealth. In these contexts there should be a strong theoretical link between the measures we used to modify our base resource variables and GDP/capita.²⁸

²⁷ For similar conceptualizations of state strength using GDP/capita, see Bairoch 1976, Beckley 2018, Miller 2017, and Fearon and Laitin, 2003. We recognize that apart from using GDP/capita as a measure of wealth, it has also been used as a proxy for a state's development (e.g. see Souva and Prins, 2006). Thus, its multiple usage should be taken into account for any complex analysis of interstate conflicts.

²⁸ It is tempting to argue that GDP/capita also becomes a proxy for resources available for foreign policy pursuits: as a state's wealth increases, there could be substantially more funding available for policies other than those involved with socio-economic welfare functions. For example, Anders et al. (2020) argue for modifying base resources by a "surplus domestic product" that is net of the level of poverty in society. While controlling for such societal needs is important, we avoid this approach for two reasons. First, and even in democracies, there is

Does it work empirically as well? It appears to be the case. In order to make that assessment we performed the following experiment. Recall that we modified base capabilities with the following formulas:

$$\text{Economic strength} = \text{GDP} * \text{RPE} * (\text{Government effectiveness} + \text{Control over corruption}) / 2$$

and

$$\text{Military strength} = \text{Milspend} * (\text{Government effectiveness} + \text{Control over corruption}) / 2$$

We next substitute the variable GDP/capita for government effectiveness and control over corruption and compare results for the 1996-2018 time-frame (consistent with available World Bank data).²⁹ While the relationship between the proxy measures and the World Bank indicators is less than perfect, it does appear to be extremely high. Comparing across dyads and across time the analysis yields a correlation in excess of .9; when focusing only on dyads that are identified as being engaged in rivalries (e.g. Colaresi et al 2007, Sakuwa and Thompson 2019), the correlation between the proxy-based indicators and the unmodified measures when we aggregate both military and economic capabilities is also in excess of .9. Such high correlations between these alternative measures gives us confidence that we can use GDP/capita to substitute³⁰ this proxy

considerable variation in the extent to which policy makers are willing to address societal socio-economic needs. This variation is likely to be further exacerbated in non-democratic regimes. Second, even in wealthy and democratic polities, there appears to be substantial variation in both economic inequality and the extent to which such inequalities are being addressed by policy makers. Simply focusing on levels of poverty or even levels of inequality within a state will not necessarily reflect the preferences of policy makers to shift substantial resources to these issues.

²⁹ By multiplying GDP by GDP per capita, we inadvertently duplicate Beckley's (2018) proposed improvement for relying on GDP as a stand alone capability measure. This was not intentional because Beckley's argument is predicated on a both a different theory and a different focus. He argues that the gross resources encompassed by GDP imply production, welfare, and security costs that are necessary to assemble and maintain the gross resources. What is critical in international politics are the net resources (gross resources minus costs) which he claims to capture by multiplying GDP by GDP per capita. Whether this operational assertion is defensible is an interesting question but not really germane to what we are attempting to do. His GDP per capita proxy is intended to represent different processes than the quality of governance that we have in mind.

³⁰ One exception may be proxying GDP/capita for states that are primarily dependent on oil production and oil exports. This exception is discussed below. The correlation between these measures for OPEC members is approximately .72, lower than for the larger sample of "all states".

variable for the World Bank indices. Thus we utilize the proxy measure to analyze a broader sampling of time that would include interstate interactions prior to the post-Cold War era.

The Iranian-Israeli dyad provides an illustration of the utility of the proxy measure. Figure 7 compares alternative measures of Iranian economic strength in the Iran-Israel dyad for 1996 (the start of the World Bank data series) and 2015. As the figure illustrates, removing the World Bank governance indices and replacing them with the GDP/capita proxy results in nearly identical scores for Iran's share of the dyad's economic strength, and both show similar dramatic decreases from the base measure of GDP size.

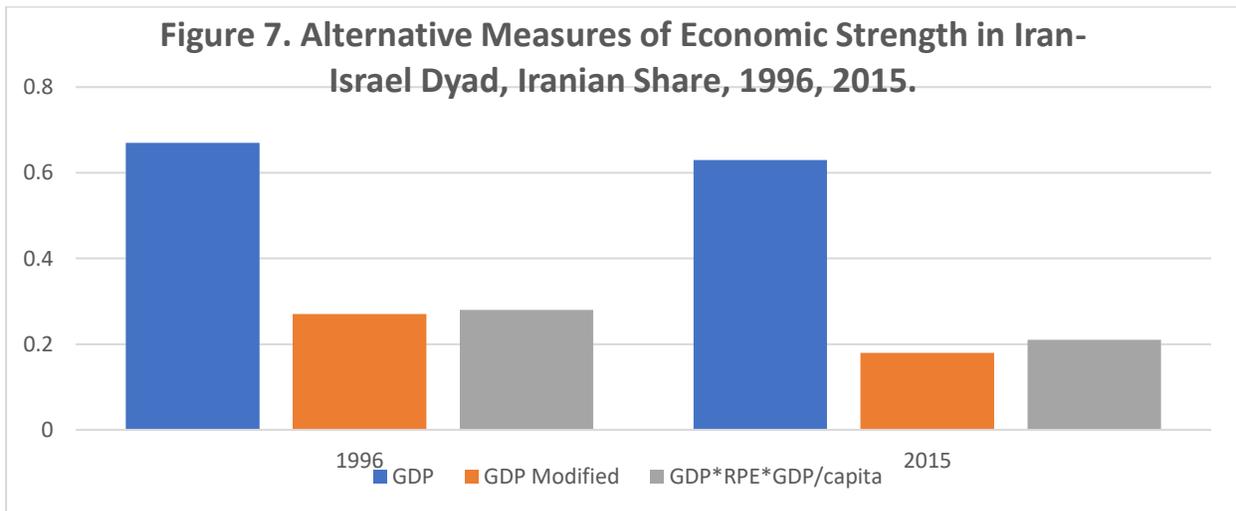
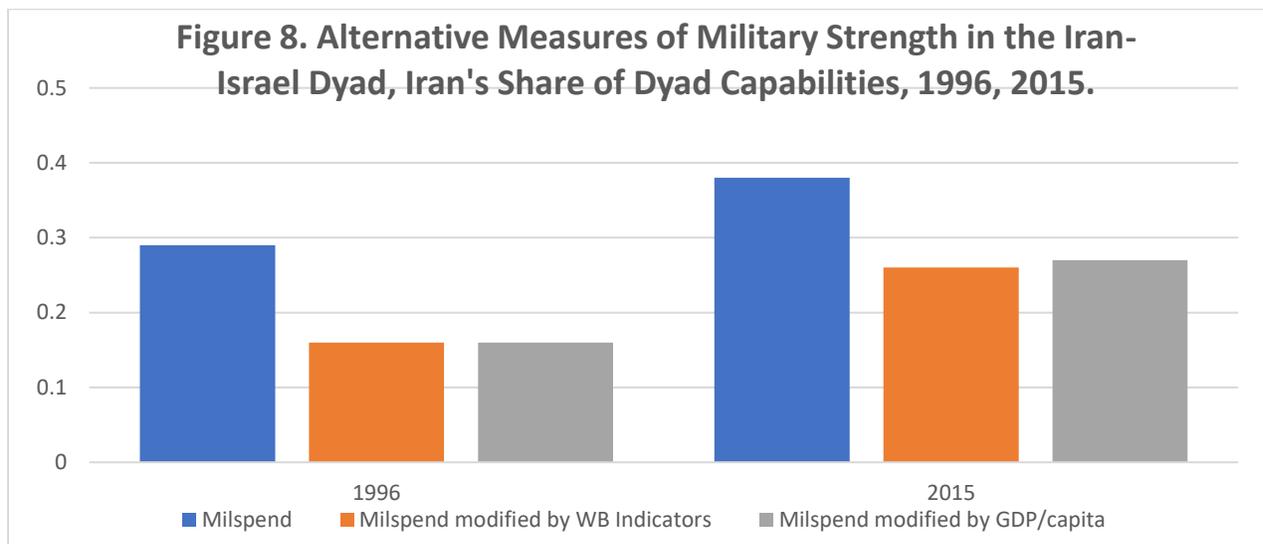


Figure 8 compares Iranian shares of military strength in the dyad across the three types of measures for 1996 and 2015. Note that the GDP/capita proxy version of the measure is—as in the previous figure—virtually identical to the measure that uses the World Bank indicators, providing some confidence that it is measuring a highly similar phenomenon.

These patterns are consistent with the high correlations we find between the proxy-based indicator and the indicator that includes estimates for quality of governance and control over corruption. However, we note that the correlation is not perfect and that there may be indeed

outliers that fail to reflect this close relationship between the World Bank measures and GDP/capita. We list one case in Appendix A (the Saudi Arabian-Iranian dyad) where the proxy measure shows significantly smaller relative strength than the measure it is meant to proxy. We expect such distortions to be minimal, yet researchers should exercise caution in simply assuming that the proxy will perfectly capture government performance.³¹



Do These Modifications Matter? Does Quality of Governance Matter?

One way to try to assess the salience of quality of governance issues for evaluating state strength is to focus on a variety of conflict processes and outcomes that are typically associated with relative state strength. We can then assess the extent to which our modified measures do

³¹ We assumed that this distortion in the Saudi Arabian-Iranian dyad may have been due to the GDP/capita-based proxy measure inadequately reflecting governance issues for states that are heavily dependent on oil resources for their economy. Note that the correlation between the World Bank based modified measure of state strength and the version using GDP/capita is .72 for oil producing OPEC states. Certainly, as Figures 7 and 8 illustrate, the proxy does well in the Iranian-Israeli dyad. It is plausible that Appendix A, focused on the Iranian-Saudi dyad, indicates a distortion between the original modifications and the proxy due to the fact that the Saudi economy is the most dependent one on oil resources among all of the OPEC states. For a similar argument and evidence regarding state capacity and oil rich states, see Savoia and Sen (2015). For an expansion of this argument to states dependent on land rent extraction for their economies, see Markovitz et al. 2020.

better than the base measures in predicting to salient dependent variables associated with conflict processes and/or outcomes.

In order to do so, we first recalibrate both sets of measures so that state strength is the combined product of both economic and military capabilities. For the base measures, focusing on dyads, we create a relative strength measure that identifies the share of dyad member (a), according to the following formula:

$$\text{Base Strength} = \frac{\text{GDP}_a + \text{Milex}_a}{(\text{GDP}_a + \text{Milex}_a) + (\text{GDP}_b + \text{Milex}_b)}$$

We then contrast the prediction of the base strength measure to the dependent variable with our modified measure of state strength:

$$\text{Modified Strength} = \frac{\text{GDP}_a * \text{RPE}_a * \text{GDP}/\text{CAPITA}_a + \text{Milex}_a * \text{GPD}/\text{CAPITA}_a}{(\text{GDP}_a * \text{RPE}_a * \text{GDP}/\text{CAPITA}_a + \text{Milex}_a * \text{GPD}/\text{CAPITA}_a) + (\text{GDP}_b * \text{RPE}_b * \text{GDP}/\text{CAPITA}_b + \text{Milex}_b * \text{GPD}/\text{CAPITA}_b)}$$

In order to gauge the relative salience of the modified measure, we look at a number of conflict processes and outcomes, focused on militarized interstate disputes (MIDs)³² between politically relevant dyads. We observe the performance of our modified measures using both simple descriptive graphs and more complex regression models.

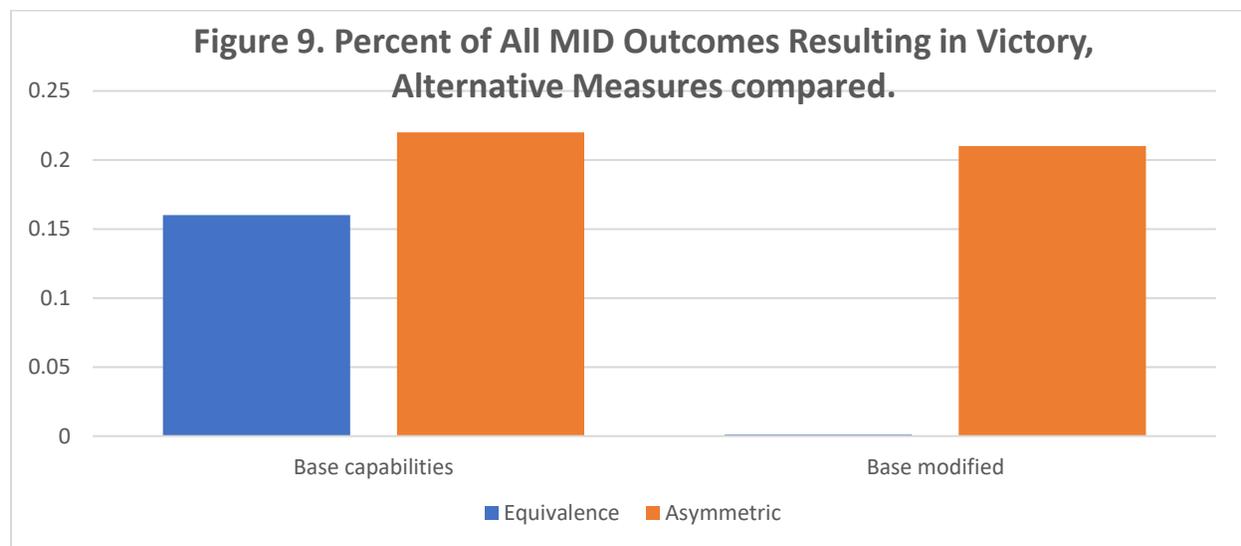
We first focus on the *outcomes* of militarized interstate disputes (MIDs) between states.³³

In this part of the analysis we differentiate between dyads that have asymmetrical capabilities

³² The latest version of the militarized interstate dispute database is discussed in Maoz et al. 2019. For recent critiques of the MIDs database, see Fordham and Sarver 2001, Gibler et al. 2016., and Palmer et al. 2020.

³³ The Correlates of War project initially codes MIDs according to nine different possible outcomes (excluding -9 for missing data). They differentiate between victories for either side in a MID (1 & 2), yields by either side in a MID (3 & 4), stalemates (5), compromises (6), released (7), unclear (8), and joins ongoing war (9). For our purposes, we focus on less ambiguous outcomes. To do this, we created three possible outcomes: Stronger State Victories, Weaker State Victories, and Stalemates. Stronger State Victories includes those MIDs that ended in a victory for the stronger state of the dyad and yields by the weaker state; Weaker State Victories includes those MIDs that

(where one member of the dyad has at least 80 percent of the dyadic strength) versus those that are more balanced. We assume that states in dyads with roughly equivalent capabilities or similar degrees of state strength, all else being equal, would more likely yield a stalemated outcome for their MIDs than for asymmetrical dyads. Those states with asymmetrical strength should have a much higher rate of generating a positive outcome from their MIDs. If our state strength measure is a more valid indicator of strength than the base measures then we should find fewer outcomes contrary to our expectations than when we use the unmodified measures.

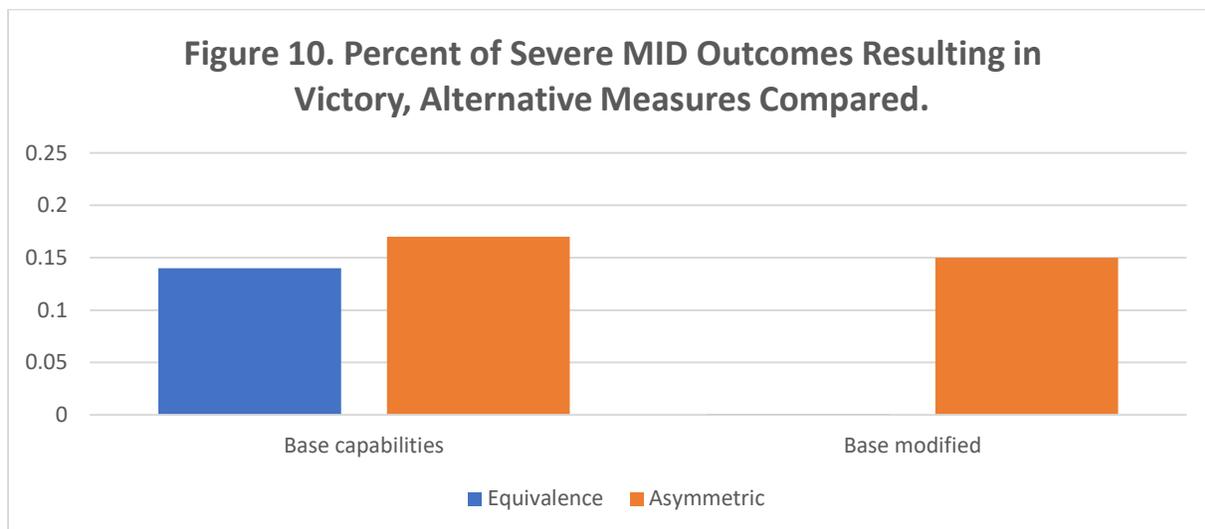


The harsh reality is that most MIDs end in stalemate; relative strength may only make a difference on the margins. Figure 9 reflects our findings. Using the base measure (combining the GDP and Milex indicators) the difference between roughly balanced and asymmetrical dyad outcomes is in the predicted direction with approximately 16 percent of MIDs leading to a winning outcome for one side over the other in roughly balanced dyads, compared to about 21 percent for asymmetrical dyads. However, our modified state strength measure indicates that

ended in a victory for the weaker state and yields by the stronger state; and Stalemates include those MIDs that ended in Stalemates and Compromises. We exclude all other MID outcomes from our assessment.

these differences are far more dramatic and more consistent with the idea that relative equivalence in strength creates stalemate. Using the modified measure, and other things being equal, we find no victories for roughly equivalent dyads while asymmetric dyads conclude their MIDs with one side victorious about one in five times.

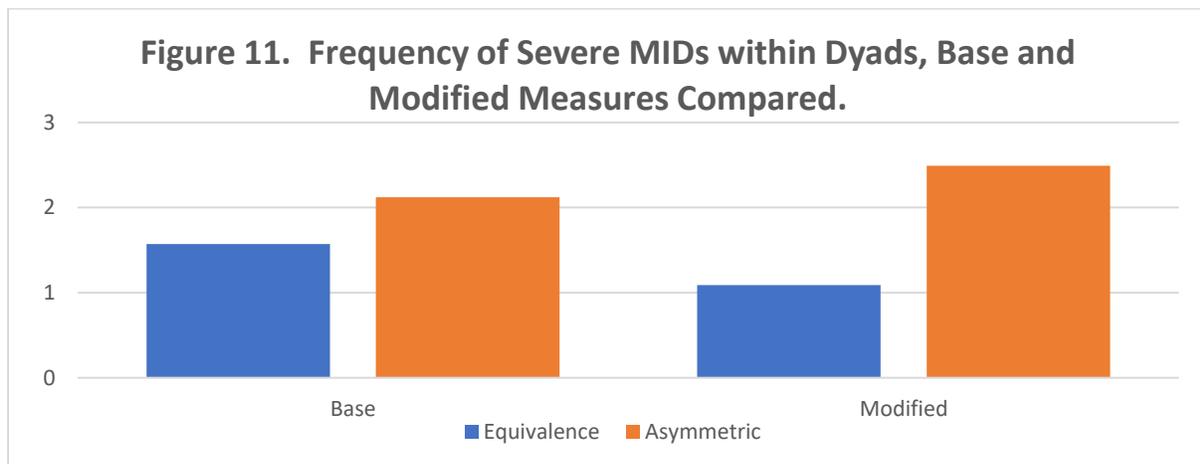
Of course, other things are rarely equal. We should not expect state strength alone to be an outstanding predictor of conflict outcomes.³⁴ This is precisely the reason why we differentiate between state strength and state power. The latter is about constraining or changing the behavior of the “other”. Strength is an important ingredient for power, but it is only one, important component. Therefore, we should not expect that even a nuanced measure of state strength would predict by itself to conflict outcomes between states. Still, strength matters and our qualified measure produces a more theoretically interesting outcome than the non-qualified version does.



³⁴ The literature on conflict has underscored the limited predictive capability of state strength, regardless of how it is measured. Sullivan (2007) identifies several critical components driving success and failure between stronger and weaker states, noting that major powers since 1945 had failed to attain their political objectives in nearly 40 percent of cases where they used military intervention. Bueno de Mesquita (2000) reports that over the last 200 years weaker states prevailed in wars over 40 percent of time.

The results do not change substantially if we refocus the analysis to observing just those dyads that engage in severe (categories of 4 and 5) MIDs. Figure 10 illustrates that applying our comparison to this subset of MIDs yields outcomes that are highly similar to those in Figure 9. The only significant difference between the two figures is around the success rate of asymmetrical dyads, which in the case of severe MIDs is reduced from approximately 20 percent to a more modest 15 percent.

Next, we focus on the *frequency* with which MIDs occur within dyads. Since most MIDs end in stalemate, it is not a stretch to assume that a dyad will engage in more than one MID.³⁵ However, given that non-asymmetric dyads seldom (according our base indicators) or virtually never (according to our modified measures) succeed when engaging in MIDs, we would expect that repeated numbers of MIDs would occur primarily among dyads with asymmetrical state strength.³⁶ This should be even more the case with severe MIDs since those carry the greatest risks of escalation to even larger conflicts in the form of wars.



³⁵ In fact, Goertz and Diehl (1993), and Kline et al. (2006) include repeated MIDs as part of their definition of dyads in rivalries.

³⁶ Although we would expect that dyads engaged in rivalries would likely generate more MIDs than dyads not in rivalries.

Figure 11 shows the mean rate of severe MID occurrence for dyads with rough equivalence in state strength, and the frequency with which asymmetrical strength dyads engage in severe MIDs, comparing the base measure with the modified measure. As the figure indicates, using the base measures, roughly equivalent strength dyads average approximately 1.6 MIDs per dyad compared to a rate of 2.12 MIDs for asymmetric dyads. Our modified measure, however, indicates a starker contrast: now roughly equivalent dyads engage in an average of around only one (1.12) MID versus 2.5 MIDs for dyads that contain asymmetrical state strength. Thus, the modified measure is associated with a dramatic decline for cases of multiple severe MIDs for roughly equivalent dyads, along with a substantially greater frequency of severe MIDs within asymmetrical dyads, consistent with our argument.

These findings may seem puzzling to rivalry analysts who argue that rivalries either involve states with roughly equal capabilities or at least persist longer if capabilities are roughly symmetrical. Rivalries do not generate all MIDs but they are responsible for a disproportionate number. How is it that then that more asymmetrical dyads account for many more MIDs, on average, than symmetrical dyads?

To pursue this question farther we conducted a partial replication of a recent research effort that probes the relationships between territorial disputes, rivalries and MIDs, and we specifically focused on the models the authors utilized for exploring MID initiation by dyads (Sakuwa and Thompson 2019, Table 10, p. 217). We choose this work to replicate for several reasons, including: a) it is the most recent publication to report the most utilized base model for predicting MID initiation within dyads; b) it highlights alternative models focusing on the two strongest predictors in the literature regarding interstate conflicts (territorial disputes and rivalries); and c) we find the work to be relatively straightforward for replication.

Sakuwa and Thompson first create what they appropriately label the “industry standard” for predicting MID initiation (the base model). Then they refine the model by inserting a variable reflecting whether or not dyads are in rivalry (“including rivalry”). Finally, they offer a third model which disaggregates the variables of dispute and rivalry to reflect various dispute type-rivalry relationship (“including dispute categories”).³⁷

We make two changes to their work. First, Sakuwa and Thompson utilize the entire 1816-2001 time-frame; our data cover primarily the period beginning with 1960, and so we restrict our replication effort to the 1960-2001 period. Second, in order to weigh relative capabilities, we restructure their capability variable so that it becomes a percentage measure of the stronger state’s spending divided by total spending in the dyad. We structure our modified capability variable in the same manner.³⁸ Note that unlike the analysis above, instead of bifurcating dyads based on whether or not they have asymmetrical capabilities, our measure of relative strength now uses a more specific range of relative strength in the dyad, ranging from perfect balance (.5) to a near perfect imbalance (.9).

If our assertions are correct with respect to the need to modify capabilities with quality of governance concerns, we should witness two outcomes in our regression equation. First, we should expect our modified measure of state strength will perform better in predicting MID initiation than just the relative capability scores within a particular dyad. Second, if such relative strength is important for predicting conflicts within dyads, the improved predictive capability of

³⁷ In the original article there is a fourth model which in effect explores predictors of MID initiation for only a small subset of dyads that are engaged in rivalries; we do not replicate that model since a) they find virtually no predictor variables of interest and b) the resulting small sample makes comparison with the other three models difficult.

³⁸ We also add AIC and BIC scores to the table in order to identify whether our modified version or the original version produces a better fit with the dependent variable of interest.

the modified measure should also impact on the two variables most often identified in the literature as being associated with conflicts in dyads: territorial disputes and rivalries.

Specifically, we should expect that our modified capability measure should, compared to the variable in the Sakuwa and Thompson models, yield substantially higher coefficients. At the same time, to the extent that AIC³⁹ scores reflect relative goodness of fit between competing empirical models, our versions of the three models should demonstrate *consistently* lower AIC scores across all three models, demonstrating a better fit through the use of our modified measures. Additionally, the salience of our modified measure should be demonstrated also through a diminished impact of disputes and rivalries on the dependent variable.

Table 1 shows the results from the replication of Sakuwa and Thompson's three primary models of MID initiation and offers a comparison with our modified strength variable. We have added both AIC and BIC values for each model. The models are presented in terms of three panels, with each panel containing the alternative specification for relative state strength in the dyad. As per our previous arguments, our modified alternative in these panels uses military expenditures modified by GDP/capita as the proxy for quality of governance constraints.

The comparison between alternative specifications of military strength across all three types of models consistently reflect our initial expectations. In every one of the three comparisons our modified strength measure demonstrates substantially higher coefficients than the basic military expenditure measure. Furthermore, the second panel of each model, as we had

³⁹ And/or BIC scores.

expected, yields AIC values that are substantially smaller than in the initial model, suggesting a consistently better “fit” is created by using our modified measure of military capabilities.⁴⁰

Table 1. Comparison of Unmodified and Modified Military Expenditure Variables for Predicting MID initiation, logit regressions (Source: Sakuwa and Thompson, 2019).

VARIABLES	Base Model		Including Rivalry		Including Dispute Categories	
	Unconstrained	Constrained	Unconstrained	Constrained	Unconstrained	Constrained
Dispute	1.261** (0.091)	1.207** (0.113)	0.676** (0.102)	0.675** (0.125)		
Rivalry			1.526** (0.105)	1.483** (0.128)		
Disputing – not rivalry					1.373** (0.126)	1.388** (0.150)
Not disputing - rivalry					2.388** (0.139)	2.520** (0.176)
Disputing - rivalry					2.237** (0.119)	2.181** (0.149)
Contiguity	2.884** (0.088)	2.704** (0.110)	2.634** (0.093)	2.470** (0.115)	2.413** (0.099)	2.217** (0.124)
Milspend	-1.885** (0.226)		-1.408** (0.238)		-1.449** (0.237)	
Milspend modified		-3.924** (0.610)		-3.528** (0.627)		-3.531** (0.625)
Alliance	0.0929 (0.120)	0.0601 (0.134)	0.235 (0.121)	0.185 (0.137)	0.181 (0.121)	0.146 (0.137)
Joint democracy	-0.031** (0.006)	-0.025** (0.007)	-0.023** (0.006)	-0.019** (0.007)	-0.021** (0.006)	-0.018* (0.007)
Major power	1.885** (0.091)	1.394** (0.105)	1.998** (0.091)	1.554** (0.106)	1.927** (0.091)	1.515** (0.106)
Peace years	-0.331** (0.018)	-0.362** (0.023)	-0.294** (0.018)	-0.325** (0.023)	-0.290** (0.018)	-0.322** (0.023)
Spline1	-0.001** (0.000)	-0.002** (0.000)	-0.001** (0.000)	-0.001** (0.000)	-0.001** (0.000)	-0.001** (0.000)
Spline2	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)
Spline3	-0.000	-0.000*	-0.000	-0.000	-0.000	-0.000*

⁴⁰ As a robustness test, we reran the logit regressions using the SIPRI-based measure of military spending instead of COW military expenditure measure used in the initial study. The results are shown in Appendix B; our constrained capability version of relative strength continues to create stronger predictions to the dependent variable and both AIC and BIC scores continue to suggest a better fit than for the unmodified measure.

Constant	(0.000) -3.576** (0.205)	(0.000) -2.482** (0.377)	(0.000) -4.150** (0.218)	(0.000) -2.902** (0.390)	(0.000) -4.157** (0.217)	(0.000) -2.956** (0.389)
N	377,899	259,609	377,899	259,609	377,899	259,609
AIC	9333.9	6294.4	9125.3	6162.8	9059.9	6109.1
BIC	9453.2	6409.5	9255.4	6288.5	9200.8	6245.2
χ^2	4850.69**	3087.55**	5061.31**	3221.09**	5128.70**	3276.85**

Standard errors in parentheses. ** $p < 0.01$, * $p < 0.05$. Bold values highlight key comparisons in the text.

Are these differences between the basic versus the modified measures of military strength of substantive significance? We offer a set of predicted probabilities, based on Table 1, that compare first the predictions between the two approaches for measuring the relative strength in the dyads sampled on MID initiation; and second, the dampening effect of these measures on territorial disputes and rivalries.

Figure 12. Predicted Probabilities of Unmodified vs. Modified Relative Strength Measures on MID Initiation.

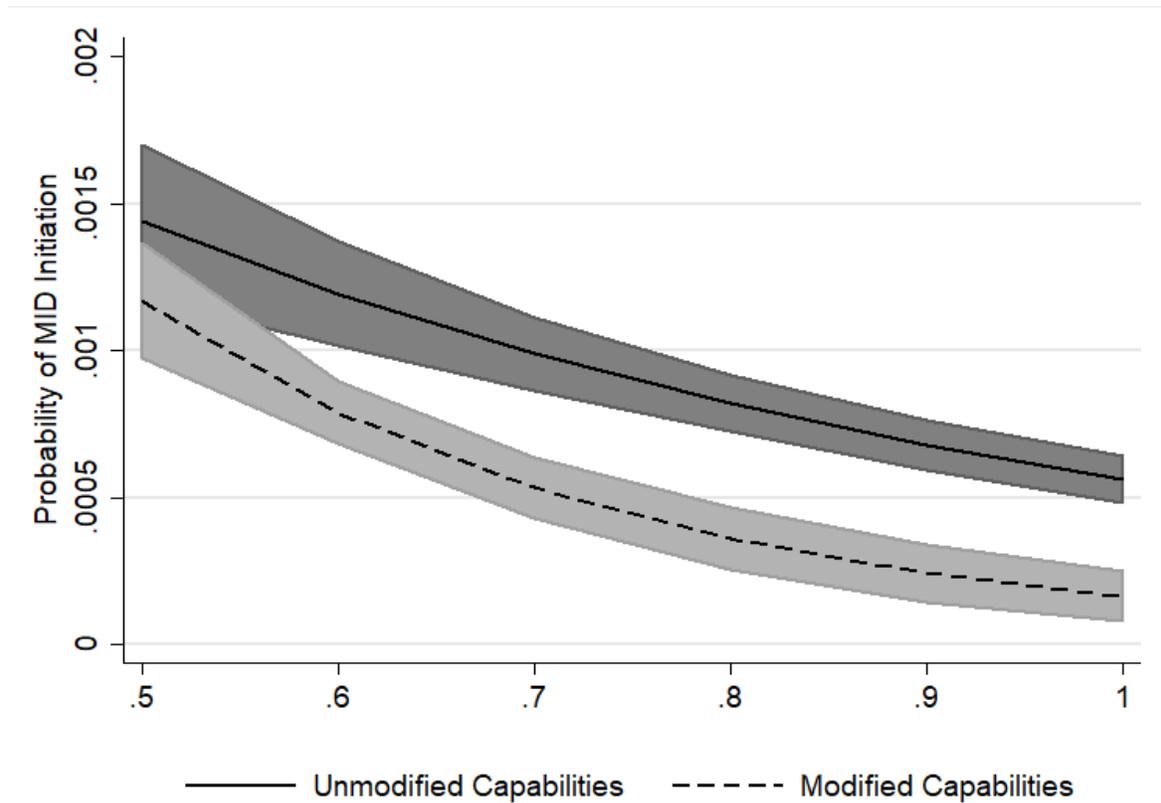
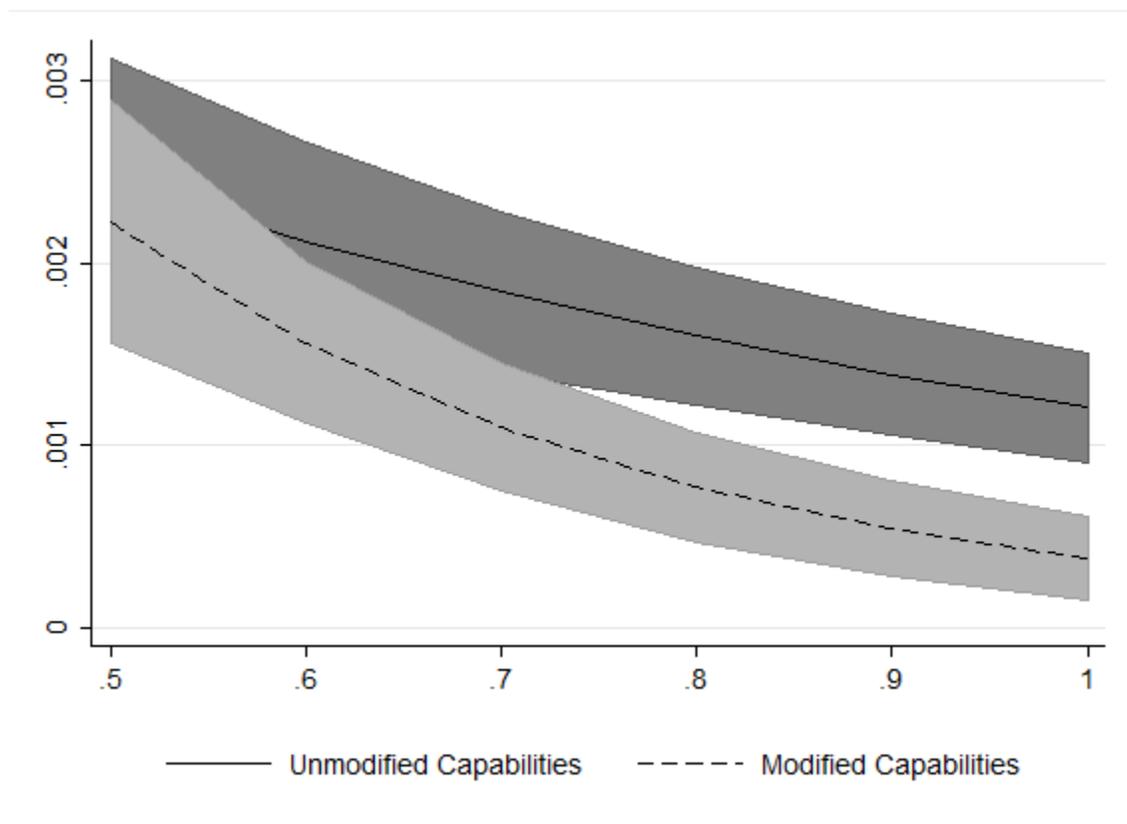


Figure 12 demonstrates the predicted probabilities on MID initiation by the two different measures of state strength. As the asymmetry in the dyad increases, the modified relative strength variable diminishes the probability of the dyad experiencing a MID initiation substantially more than the simpler measure of military spending. For example, in near-perfect balance the modified relative strength measure reduces the probability of MID initiation by roughly 19 percent compared to the base measure. This difference rises to 64 percent difference between the two measures at high asymmetry.

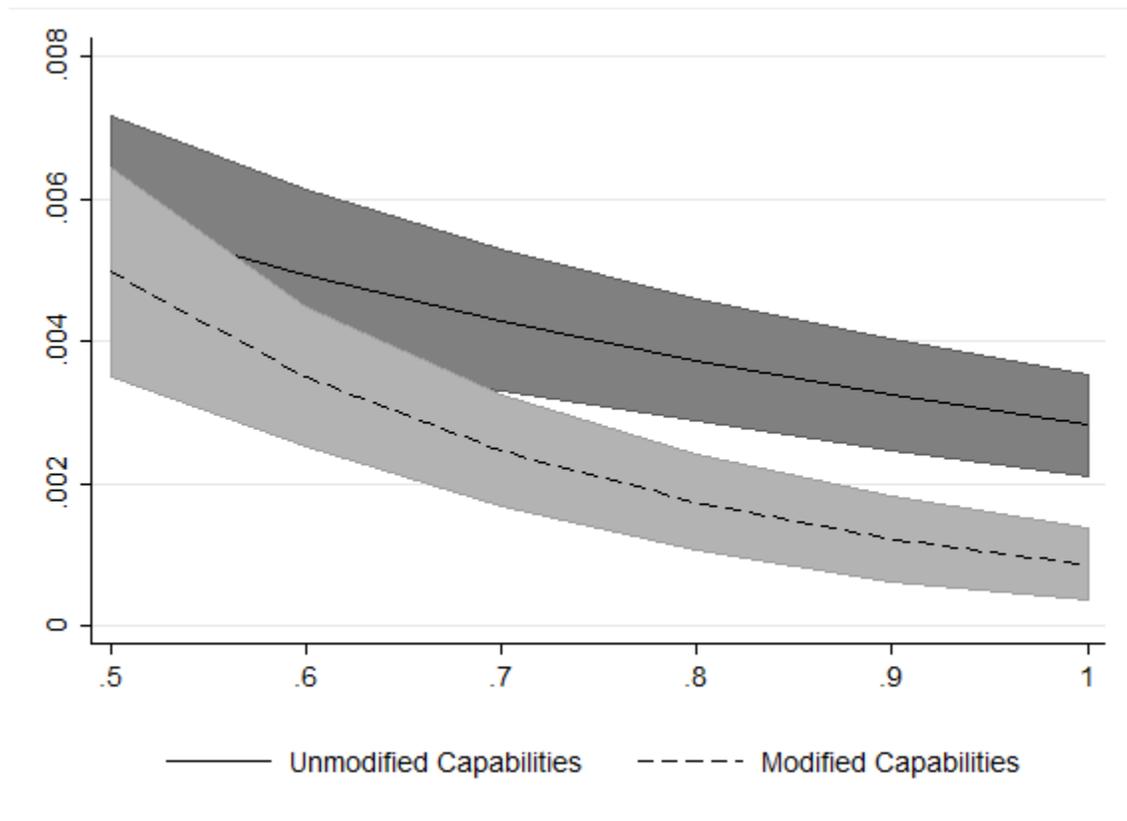
Figure 13. Predicted Probabilities of Territorial Disputes on MID Initiation Across Unmodified and Modified Measures of Relative State Strength.



Figures 13 and 14 show the relative effects of the two strength measures when territorial dispute or rivalry occurs in the dyad. In both instances, the effect of disputes and rivalries are

substantially more diminished by the modified state strength measure than by the unmodified state strength measure. For example, the probability of experiencing a MID in the dyad when there is a territorial dispute is 61 percent lower at high asymmetry using the modified strength measure compared to the original measure. This difference is roughly similar when observing dyads in rivalry. As these figures also show, the differences between the two strength measures become more significant as asymmetry in the dyad increases.

Figure 14. Predicted Probabilities of Rivalries on MID Initiation Across Unmodified and Modified Measures of State Strength.



More substantively, though, none of these findings clarify the puzzle about asymmetric dyads being more likely to engage in serious MIDs than symmetric dyads. It turns out that the tendency is applicable to both dyads involved in rivalries and ones that do not involve rivalries.

However, there is a difference. The tendency (using the modified military expenditure indicator) to engage in MIDs is even more pronounced in asymmetrical rivalry dyads (75%) than in asymmetrical non-rivalry dyads (67%). This finding does not seem to violate the intuitive expectation that symmetrical dyads are less likely to clash in MIDs. If neither side has a perceived advantage, why risk escalated conflict? Yet it also suggests that asymmetry hardly precludes rivalry and that asymmetry may be a good clue as to who is most likely to fight whom.

Conclusion

The persistent struggle to measure power differences reflects, among other things, our inability to operationalize influence success easily and the corresponding tendency to fall back on raw material differences to predict who is likely to prevail in international political contests. We offer no antidotes for the former problem. The later problem, however, can be addressed by developing better formulations of state strength. In this analysis, we qualify economic (GDP) and military expenditure information in terms of extraction performance and state capacity indexes. It makes some discernible difference in comparing members of dyads. It also helps explain why apparent asymmetries in dyadic relationships are not always what they seem. Even so, asymmetry may be a more interesting characteristic than is often assumed.

We note as well some caveats regarding the use of our approach to measuring state strength. First, the most obvious: while GDP/capita appears to be a good proxy for tempering economic and military capabilities by quality of governance, it has some obvious limitations. One is that it cannot be used in models that also rely on other independent variables that highlight the explanatory roles of economic development and/or wealth as conditioning factors effecting conflict (or cooperative) processes. In such instances, researchers would need to find another proxy for conditioning state strength.

A second caveat is that GDP/capita, while functioning as a good proxy for quality of governance in general, may not function well for states where it is not an actual correlate of economic development and therefore, of quality of governance. We have in mind especially states with heavy reliance on oil production and oil exports as a large part of their economy. However, other states that depend heavily on a single resource extraction or land-based extraction for their economic well being would fit into the same category (Markowitz et al. 2020).

Third, we recognize that for certain time periods, and especially during the Cold War, estimates of basic resources (GDP), the wealth of states (GDP/capita) and even military spending for communist states with closed economies have been highly unreliable. This problem is not unique to our approach; researchers who had estimated the economic and military capabilities of these states have struggled with the same problem. For these states, an alternative specification besides wealth and economic size would be warranted (e.g. energy consumption), assuming that such alternative specifications would be more reliable indicators of basic resources and their application. Alternatives to military spending are more difficult to uncover for these states. Regardless, researchers need to be wary of the estimates created for this critical subsample of states.

Fourth, we don't wish to discourage the search for other proxies for estimating quality of state governance and control over corruption. One plausible alternative, as our reading of Carroll and Kenkel's (2019, p. 583) work suggests, is energy consumption by states as a plausible proxy for quality of governance. Another possibility for military strength would be proxying quality of governance by dividing military spending by the size of a state's military forces; this "per capita" measure may tap quality of governance issues (including technological investments in the military) as well as our proxy measure. The general point here is that while we are satisfied with

our proxy measure, we are aware that alternative proxies (both theoretically meaningful and empirically feasible) can be utilized to gauge quality of governance issues.

Finally, we suggest that substantially more work needs to be done to estimate the overall strength of states. Our approach has been to treat economic and military strength equally. We recognize that this is a somewhat arbitrary decision and future research needs to assess whether either of these capabilities should be weighed more heavily than the other, and especially so dependent upon the issue of concern.⁴¹

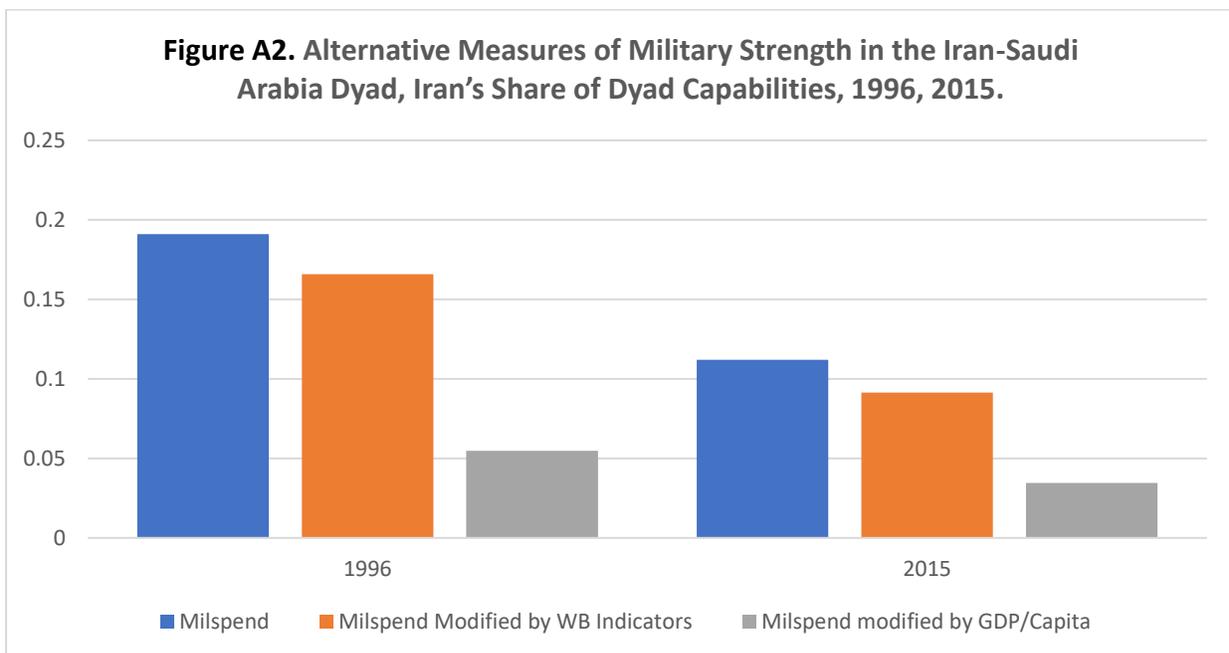
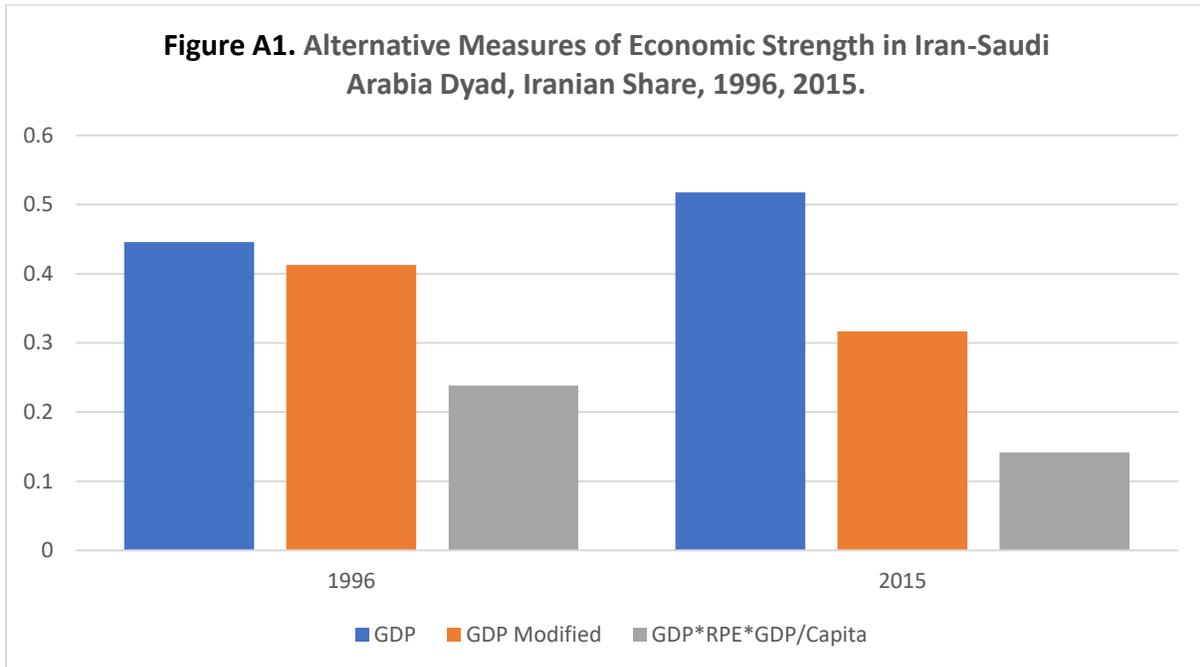
Even with these caveats in mind, we suggest that approaches to assessing state strength as a salient explanation of conflict processes and even conflict outcomes are well served by refocusing their measurement strategies on variables that also address concerns about both political extraction and quality of governance as part of the measurement process. Scholars who continue to work with CINC scores need not abandon them but should seek modifications that would condition those scores with quality of governance and extraction measures. Scholars who work more directly with GDP values or military expenditures as measures of state strength are encouraged to utilize our proxy measure or another viable alternative as long as they can control for the mediating effects of political extraction and quality of governance.

We acknowledge as well that developing qualified measures of raw capabilities is only a starting point. The next question is what difference the application of such measures might make to enduring issues about who competes with whom and/or who wins or loses conflicts in international politics. There is no reason to assume that material differences will be shown to be either more or less significant than what is recorded in previous analyses. But we might find

⁴¹ Note Sullivan's (2007) findings, showing that more powerful states vary in their relative success depending on the objectives motivating their use of coercive power.

ourselves in a better position than was the case earlier to assess the significance of symmetry and asymmetry in clashes between actors. At the same time, our probes so far indicate that qualified measures of state strength do make a significant difference to empirical outcomes. More work along these lines appears justified.

APPENDIX A: Comparison of Unmodified and Two Modified Indicators of State Strength, Saudi-Iranian Dyad.



Appendix B. Comparison of Unconstrained and Constrained Military Expenditure Variables for Predicting MID initiation, logit regressions, using SIPRI as military expenditure source.

Variables	Base Model		Including Rivalry		Including Dispute Categories	
	Unconstrained	Constrained	Unconstrained	Constrained	Unconstrained	Constrained
Dispute	1.255** (0.123)	1.322** (0.138)	0.686** (0.139)	0.799** (0.154)		
Rivalry			1.464** (0.145)	1.403** (0.158)		
Disputing – not rivalry					1.374** (0.169)	1.412** (0.183)
Not disputing - rivalry					2.400** (0.198)	2.444** (0.232)
Disputing - rivalry					2.207** (0.162)	2.280** (0.180)
Contiguity	2.789** (0.120)	2.690** (0.137)	2.543** (0.127)	2.468** (0.144)	2.323** (0.136)	2.226** (0.155)
Milspend	-1.988** (0.301)		-1.675** (0.312)		-1.732** (0.310)	
Milspend constrained		-2.345** (0.921)		-2.401** (0.927)		-2.623** (0.915)
Alliance	-0.103 (0.147)	-0.002 (0.150)	0.046 (0.149)	0.0921 (0.152)	0.008 (0.149)	0.058 (0.152)
Joint democracy	-0.028** (0.008)	-0.021* (0.008)	-0.020* (0.008)	-0.0155 (0.008)	-0.0187* (0.008)	-0.014 (0.008)
Major power	1.664** (0.123)	1.275** (0.127)	1.861** (0.122)	1.541** (0.129)	1.782** (0.124)	1.498** (0.130)
Peace years	-0.357** (0.025)	-0.351** (0.028)	-0.316** (0.025)	-0.315** (0.028)	-0.310** (0.026)	-0.311** (0.028)
Spline1	-0.001** (0.000)	-0.001** (0.000)	-0.001** (0.000)	-0.001** (0.000)	-0.001** (0.000)	-0.001** (0.000)
Spline2	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)
Spline3	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
Constant	-3.173** (0.267)	-3.400** (0.550)	-3.661** (0.282)	-3.594** (0.552)	-3.676** (0.280)	-3.522** (0.544)
χ^2	2740.89**	2173.97**	2843.68**	2253.78**	2881.49**	2283.49**
N	226,061	191,930	226,061	191,930	226,061	191,930
AIC	5144.5	4330.0	5043.6	4252.1	5007.8	4224.5
BIC	5258.1	4441.8	5167.6	4374.1	5142.1	4356.6

Standard errors in parentheses. ** p<0.01, * p<0.05. Bold values highlight the key comparisons in the text.

References

Anders, Theresa, Christopher J. Fariss, and Jonathan M. Markowitz. 2020. "Bread Before Guns or Butter: Introducing Surplus Domestic Product (SDP)." *International Studies Quarterly*,

Arbetman, Marina, and Jacek Kugler (eds.). 2018. *Political Capacity and Economic Behavior*. New York: Routledge.

Arbetman-Rabinowitz, Marina, and Kristin Johnson. 2007. "Relative Political Capacity: Empirical and Theoretical Underpinnings." Claremont, California (October).

Bairoch, Paul. 1976. "Europe's Gross National Product: 1800-1976." *Journal of European Economic History*, 5, 2: 273-340.

Barboza, David. 2012. "Billions in Hidden Riches for Family of Chinese Leader." *New York Times*, October 25, retrieved at: <https://www.nytimes.com/2012/10/26/business/global/family-of-wen-jiabao-holds-a-hidden-fortune-in-china.html?module=inline>

Beckley, Michael. 2010. "Economic Development and Military Effectiveness." *Journal of Strategic Studies*, 33, 1: 43-79.

Beckley, Michael. 2018. "The Power of Nations: Measuring What Matters." *International Security*, 43, 2: 7-44.

Bird, Mike. 2019. "Foreign Investors are Checking Out of the Chinese Stock Market." *The Wall Street Journal*, May 15, retrieved at: <https://www.wsj.com/articles/foreign-investors-are-checking-out-of-the-chinese-stock-market-11557912910>

Bremer, Stuart A. 1992. "Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War, 1816-1965." *Journal of Conflict Resolution*, 36, 2: 309-341.

Bueno de Mesquita, Bruce. 2000. *Principles of International Politics*. Washington, DC: Congressional Quarterly Press.

Carroll, Robert J., and Brenton Kenkel. 2019. "Prediction, Proxies, and Power." *American Journal of Political Science*, 63, 3: 577-593.

Chayes, Sarah. 2015. *Thieves of State*. New York: W.W. Norton.

Chen, Wei, Xilu Chen, Chang-Tai Hsieh, and Zheng (Michael) Song. 2019. "A Forensic Examination of China's National Accounts," *Brookings Papers on Economic Activity*, [brookings.edu/bpea-articles/a-forensic-examination-of-chinas-national-accounts/](https://www.brookings.edu/bpea-articles/a-forensic-examination-of-chinas-national-accounts/).

Colaresi, Michael, Karen Rasler, and William R. Thompson. 2007. *Strategic Rivalries in World Politics: Position, Space and Conflict Escalation*. Cambridge, UK: Cambridge University Press.

Cole, Wade M. 2015. "Mind the Gap: State Capacity and the Implementation of Human rights Treaties." *International Organization*, 69, 2: 405-441.

da Costa, Ana Nicolaci. 2019. "China's economic slowdown: How bad is it?" BBC News, September 26, retrieved at: <https://www.bbc.com/news/business-49791721>

de Soysa, Indra, John R. Oneal, and Yong-Hee Park. 1997. "Testing Power-Transition Theory Using Alternative Measures of National Capabilities." *Journal of Conflict Resolution*, 41, 4: 509-528.

Diamond, Donald, Lisa Rein, and Juliet Eilperin. 2021. "Trump Leaves Behind a Damaged Government." *The Washington Post*, February 6. Retrieved at: https://www.washingtonpost.com/politics/biden-trump-federal-government/2021/02/06/93a22e4e-5b50-11eb-b8bd-ee36b1cd18bf_story.html

Fassihi, Farnaz, Richard Perez-Pena, and Ronen Bergman. 2020. "Iran Admits Serious Damage to Natanz Nuclear Site, Setting Back Program." *New York Times*, July 5, at: <https://www.nytimes.com/2020/07/05/world/middleeast/iran-Natanz-nuclear-damage.html> (retrieved July 27, 2020).

Fearon, James D., and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil Wars." *American Political Science Review*, 97, 1: 75-90.

Fordham, Benjamin O., and Christopher C. Sarver. 2001. "The Militarized Interstate Disputes Data Set and the United States Use of Force." *International Studies Quarterly*, 45, 3: 455-466.

Gartzke, Erik. 1998. "Kant We All Get Along? Motive, Opportunity, and the Origins of the Democratic Peace." *American Journal of Political Science* 42:1-27.

Gates, Robert M. 2020. "The Overmilitarization of American Foreign Policy." *Foreign Affairs*, 99, 4:121-132.

Geller, Daniel S. 2008. "Relative Power, Rationality, and International Conflict." In Kugler and Lemke, (eds.) *Parity and War*. Ann Arbor: University of Michigan.

Geller, Daniel S. 2000. "Material Capabilities: Power and International Conflict." In Paul Vasquez (ed.) *What Do We Know About War?* New York: Rowman & Littlefield.

Gibler, Douglas M., Steven V. Miller, and Erin K. Little. 2016. "An Analysis of the Militarized Interstate Dispute (MID) Dataset, 1816-2001." *International Studies Quarterly*, 60, 4: 719-730.

Goertz, Gary, and Paul F. Diehl. 1993. "Enduring Rivalries: Theoretical Constructs and Empirical Patterns." *International Studies Quarterly*, 37, 2: 147-171.

Harada, Issaku, and Kazuya Manabe. 2019. "Quiet capital flight dents China's sway as \$1.2 tn disappears." *Nikkei Asian Review*, June 23, retrieved at:

<https://asia.nikkei.com/Spotlight/Datawatch/Quiet-capital-flight-dents-China-s-sway-as-1.2tn-disappears>

Jennings, Ralph. 2018. "Bad for Business? China's Corruption Isn't Getting Any Better Despite Government Crackdowns." *Forbes*, March 15, retrieved at: <https://www.forbes.com/sites/ralphjennings/2018/03/15/corruption-in-china-gets-stuck-half-way-between-the-worlds-best-and-worst/#7037aeee73d1>

Joyce, Kyle A., Faten Ghosn, and Res-at Bayer. 2013. "When and Whom to Join: The Expansion of Ongoing Violent Interstate Conflicts." *British Journal of Political Science*, 44: 205-238.

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

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2010. "What do the Worldwide Governance Indicators Measure?" *European Journal of Development Research*, 22:55-58.

Klein, James P., Gary Goertz, and Paul F. Diehl. 2006. "The New Rivalry Dataset: Procedures and Patterns." *Journal of Peace Research*, 43,3: 331-338.

Kugler, Jacek, and Douglas Lemke (eds.). 2008. *Parity and War: Evaluations and Extensions of the War Ledger*. Ann Arbor: University of Michigan Press.

Kugler, Jacek, and Ronald L. Tammen. 2012. *The Performance of Nations*. Lanham: Rowman and Littlefield.

Lee, Michael J., and William R. Thompson. 2017. "Major Powers vs. Global Powers: A New Measure of Global Reach and Power Projection Capacity." In William R. Thompson (ed.), *Oxford Research Encyclopedia of Politics*. New York: Oxford University Press.

Lemke, Douglas. 2002. *Regions of War and Peace*. Cambridge: Cambridge University Press.

Lemke, Douglas, and Suzanne Werner. 1996. "Power Parity, Commitment to Change, and War." *International Studies Quarterly* 40: 235-60.

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

Markowitz, Jonathan, Suzie Mulesky, Benjamin A.T. Graham, and Christopher L. Fariss. 2020. "Productive Pacifists: The Rise of Production-Oriented States and the Decline of Profit Motivated Conquest." *International Studies Quarterly*,

Maoz, Zeev, Paul L. Johnson, Jasper Kaplan, Fiona Ogunkoya, and Aaron P. Shreve. 2019. "The Dyadic Militarized Interstate Disputes (MIDs) Dataset Version 3.0: Logic, Characteristics, and Comparisons to Alternative Datasets." *Journal of Conflict Resolution*, 63, 3: 811-835.

Miller, Benjamin. 2017. "Theory of Regional War and Peace." *Oxford Research Encyclopedia of Politics*, DOI: 10.1093/acrefore/9780190228637.013.273

Moul, William. 2003. "Power Parity, Preponderance, and War between Great Powers, 1816-1989." *Journal of Conflict Resolution*, 47, 4: 468-489.

Palmer, Glenn, Vito D'Orazio, Michael R. Kenwick, and Roseanne W. McManus. 2020. "Updating the Militarized Interstate Dispute Data: A Response to Gibler, Miller, and Little." *International Studies Quarterly*, 64, 2: 469-475.

Reed, Willim, David H. Clark, Timothy Nordstrom, and Wonjae Hwang. 2008. "War, Power, and Bargaining." *Journal of Politics*, 70, 4:1203-1216.

Sakuwa, Kentaro, and William R. Thompson. 2019. "On the origins, persistence, and termination of spatial and positional rivalries in world politics: Elaborating a two-issue theory of conflict escalation." *International Area Studies Review*, 22, 3: 203-225.

Savioa, Antonio, and Kunal Sen. 2015. "Measurement, Evolution, Determinants, and Consequences of State Capacity: A Review of Recent Literature." *Journal of Economic Surveys*, 29,3: 441-458.

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

Standish, Reid. 2019. "China's Path Forward Is Getting Bumpy." *The Atlantic*, October 1, retrieved at: <https://www.theatlantic.com/international/archive/2019/10/china-belt-road-initiative-problems-kazakhstan/597853/>

Stephens, Bret. 2019. "Is China Heading for Crisis?" *New York Times*, retrieved on 10/3/2019 at: <https://www.nytimes.com/2019/10/03/opinion/china-xi-jinping.html>

The Economist. 2011. "India in Africa: Catching Up." May 28: 7.

Thompson, William R. 2001. "Identifying Rivals and Rivalries in World Politics." *International Studies Quarterly*, 45, 4: 557-586.

Vasquez, John A. 1996. "Distinguishing Rivals That Go to War From Those That Do not: A Quantitative Comparative Case Study of the Two Paths to War." *International Studies Quarterly*, 40, 4: 531-558.

Volgy, Thomas J., and Kelly Marie Gordell. 2019. "Rising Powers, Status Competition, and Global Governance: A Closer Look at Three Contested Concepts for Analyzing Status Dynamics in International Politics." *Contemporary Politics*, doi: [10.1080/13569775.2019.1621719](https://doi.org/10.1080/13569775.2019.1621719).

Wildau, Gabriel (2019) "China's Economy is 12% Smaller Than Official Data Say, Study Finds." *Financial Times*, March 6, 2019. [ft.com/content/961b4b32-3fce-11e9-b896-fe360c329e3e](https://www.ft.com/content/961b4b32-3fce-11e9-b896-fe360c329e3e).

World Bank. 2012. "Avoiding the Middle Income Growth Traps." *Poverty Reduction and Economic Management Network*, November: 98: 1-7.

Zelikow, Philip, Eric Edelman, Kristofer Harrison, and Celeste Ward Gventer. 2020. "The Rise of Strategic Corruption: How States Weaponize Graft." *Foreign Affairs*, 99, 4: 107-120.