Material Military Power:
A Country-Year Measure of Military Power, 1865-2019

Mark Souva
msouva@fsu.edu
Florida State University

Forthcoming at Journal of Peace Research

Abstract

Military power is central to diplomacy and much of international relations, yet common quantitative measures have limited surface validity. This limitation stems from focusing on latent power and only indirectly incorporating major weapon systems. I contend that weapons are central to military power and present a new measure of country military power based primarily on armaments. The measure includes major naval, air and land weapons as well as nuclear weapons and ballistic missile capability. I examine the surface, content, and context validity of the measure and compare it to existing measures. I show that this measure of material military power (MMP) has more surface and context validity than alternative measures. I find that MMP better predicts war outcomes, better accounts for militarized threats, and performs well as a control variable for country power.

Introduction

Military power animates much of world politics. Political leaders use their country's military power to threaten, conquer, and defend from attack. Several critical concepts in international relations are a function of military power, such as the balance of power between countries, polarity, shifting power, the probability of victory in war, and arms races. Our understanding of much of world politics depends greatly on how we measure military power.

Common large-N measures of military power focus too much on demographic and economic aspects. Owing to demographic factors, the Correlates of War Composite Indicator of National Capabilities (CINC) ranks China as the top military power in the world from 1999 through 2016 (Singer et al., 1972), but the United States has significantly more aircraft carriers, advanced fighter aircraft, nuclear attack submarines, and missile cruisers than China in this period (Saunders and Souva, 2019; Crisher and Souva, 2014). A recent measure proposed by Beckley (2018) addresses some concerns with the COW CINC measure, but it has its own limitations, largely because it is a measure of latent and not actual military power. The Beckley (2018) measure considers Japan and Germany as militarily stronger than Britain, France, and Russia in this period despite the former not having and the latter having aircraft carriers, nuclear attack submarines, ballistic missiles, and nuclear weapons.

This research presents a new measure of military power for all country-years from 1865-2019. The measure, material military power (MMP), differs conceptually from common quantitative measures of military power in two important ways. First, MMP focuses on the military. The most common large-N measures focus more on economic and population indicators. Second, the primary focus of MMP is a country's weapons systems. As such MMP includes multiple military components not found in extant measures. The focus on weapons systems also makes MMP a measure of actual and not latent military power.

MMP has considerable surface and context validity. For example, based on MMP the United States has the strongest military in the world since 2000. The MMP measure also indicates that France, Great Britain, and Russia are more militarily powerful than Japan and Germany in the twenty-first century. With respect to context validity, I find that MMP better predicts war outcomes and the making of militarized threats than broader measures of national power or an indicator based on military expenditures. MMP performs on par with most measures of national power or military expenditures as a control variable in extant models.

In addition to its comparatively strong validity, MMP has a broader range of relevant uses than other measures of military power. Researchers can use MMP to identify the strongest military powers in the world, describe the distribution of world power, forecast changes in a country's power, create dyadic power ratios, and explain the occurrence and outcomes of crises and militarized conflicts.

Measures of Military Power

Military power is the source of power that actors use to make violent threats against others or to inflict damage on others or their property. Measuring military power is difficult because the concept is complex. Like the broader concept of power, military power has multiple "bases" (Dahl, 1957, 203). The most important bases of military power are weapons, troops, training, tactics, logistical resources (e.g. transportation and fuel), strategy, and organization. Each is difficult to measure and varies in quantity and quality and the integration of them represents the immediate actual base of military power.

The Correlates of War (COW) Composite Indicator of National Capabilities (CINC) (Singer et al., 1972) is a common measure of military power, but it is really a broader measure of national power: the stock of resources a country possesses that allow it to influence others. CINC includes military factors, but economic and demographic factors significantly influence this indicator. For example, owing to its reliance on demographic factors, CINC greatly overstates the military strength of populous countries like China in the late nineteenth century and early twentieth (Beckley, 2018).

Recently, Carroll and Kenkel (2019) created a measure for 'p', the probability of winning a dispute. Their measure DOE, dispute outcome expectations, comes from a machine learning model that maximizes out of sample predictions of dispute outcomes. DOE explains dispute outcomes better than CINC and performs better than CINC as a control variable (Carroll and Kenkel, 2019). Notwithstanding its benefits, DOE has limitations. First, it is a dyadic measure; as a result, it is not useful for assessing country military power, militarization, or predicting changes in a country's power. Second, it cannot be used to explain dispute outcomes as those outcomes were used to create the measure. Explaining and predicting conflict outcomes, however, is of significant interest to many. Third, DOE is a function of economic, military, and demographic factors. This makes it a useful composite measure, but it also makes it difficult to know which of these is most important in a particular application.

Beckley (2018, 9,14) also introduces a new measure of national power. Based on the concept of "net resources", it is elegantly simple and straightforward. It is the product of a country's gross domestic product (GDP) and its gross domestic product per capita (GDP per capita). He shows that this net resource (NR) measure of power better predicts war outcomes than CINC or GDP and that when a control variable for power is present in an empirical model, the NR measure generally improves model performance compared to CINC and GDP. Anders et al. (2020) introduce a similar measure, Surplus Domestic Product

(SDP). While Net Resources and Surplus Domestic Product have significant advantages over CINC for understanding latent power, they are less relevant for understanding a country's current military power. Net Resources and SDP are essentially measures of country wealth. "Wealth provides the basis for international power, but it is not synonymous with power" (Mastanduno et al., 1989, 463). Below I show that these indicators are less useful than MMP for explaining the use of military threats. Further, as measures of latent power they cannot tell us how militarized states are, a robust correlate of conflict (Bremer, 1992).

The empirical bet of this research is that by focusing more directly on major weapons systems we can create a better measure of country military power.

Measuring Material Military Power

"Diplomacy without armaments," Frederick the Great noted, "is like music without instruments" (Blainey, 1988, 108). Armaments are not the only aspect of military power. Troops, training, tactics (Grauer and Horowitz, 2012), logistics, and civil-military relations (Narang and Talmadge, 2018), are also relevant. Nevertheless, weapons systems are both critical to military power (McNeill, 1984; Parker, 1996) and easier to measure.[1] Giergerich et al (2017) also suggest that for a "basic judgement" about a country's military power, "it should suffice to examine its core capability portfolio."

Material military power (MMP) is a function of a country's naval, air, land, and nuclear weapons and ballistic missiles. Naval power is measured as a country's annual share of world naval warship tonnage. The measure encompasses aircraft carriers, battleships, destroyers, cruisers, and submarines, all major sur-

[1] Gebeike and Magid (2010) create an indicator based on weapons systems, but it only covers 33 countries for two years.

face vessels and submarines with at least one-thousand tons displacement. Data come from Crisher and Souva (2014) and covers the period 1865-2019. To measure air power, I use the indicator created by Saunders and Souva (2019). Their measure is based on the sum of a country's fighter and attack aircraft weighted by generation. Thus, a fifth-generation stealth fighter contributes considerably more to their measure than a third-generation MiG-21. Data covers the period 1965-2019 and is measured in annual world shares.

Land power 1 is the sum of a country's tanks, armored personnel carriers and fighting vehicles. I focus on mobile armor for two reasons. First, in ground warfare since World War II "the most important single weapon is the tank" (Van Creveld, 2010, 273). Tanks and other armored vehicles are central to maneuver warfare (Reiter and Meek, 1999, 374-5), which may be associated with an increase in conflict initiation (Mearsheimer, 1985; Reiter, 1999). Second, mobile armor is also the focus of recent research on army force structure (Sechser and Saunders, 2010) and counterinsurgency success (Lyall and Wilson, 2009). Thus, this data may also be useful to those research programs. Data come from The Military Balance (for Strategic Studies, 2020) and were recorded at five-year increments for the period 1975-2020. I linearly interpolate values between the five-year increments.

For the pre-1975 era, I create a proxy measure of land power (Land Power 2). I assume that a country's total military expenditures equals its spending on naval weapons, air weapons, land weapons, and personnel. I then use a three-step algorithm to create a measure of land power.[2] First, I calculate a country's share of non-land weapons expenditures. This value is a ratio in which the numerator is the sum of a country's world share of naval power, air power, and military personnel; the denominator is the sum of naval power, air

[2] To reduce the number of missing values on military expenditures and personnel between 1865 and 2016 I use linear interpolation and, in some cases, back-fill or front-fill one year. See Appendix for more details.

power, personnel and military expenditures. Personnel and expenditure data come from the National Material Capabilities data version 6, (Singer et al., 1972). Second, I subtract this value from one. The resulting number represents the percentage of a country's military expenditures that go to land weapons. Third, I multiply the percentage from step two by that country's annual share of military spending. The resulting number is a country's annual share of world land weapons expenditures.[3]

Missile power is a five-category ordinal measure of the maximum range of a country's ballistic missiles. Countries that do not have ballistic missiles receive a score of zero. Countries who only have short range ballistic missiles (less than 1000km) score one on this indicator. The possession of medium range ballistic missiles (1000-3000km inclusive) gives a country a score of two, intermediate range (3001-5500km) is scored as three, and countries with intercontinental ballistic missiles (more than 5500km) receive a score of four. As with the other components, I transform the ordinal variable into annual world shares. Ballistic missile data come from Mettler and Reiter (2013), which I update through 2019 using data from Arms Control Association (Davenport, 2017) and country and missile reports from the Center for Strategic and International Studies (2021) and the Nuclear Threat Initiative (2021).

To create a measure of nuclear weapons power, I first created a four-category ordinal measure of the approximate number of nuclear warheads a country possesses, where zero is equal to no nuclear weapons, one means a state has at least one but fewer than two hundred nuclear weapons, two means a state has between two hundred and five hundred fifty inclusive, and three means a state has over five hundred fifty nuclear weapons. An ordinal measure is better than a binary indicator for possession of nuclear weapons as it provides some variation over time and allows us to distinguish between the superpowers, the only

[3]The correlation between Land Power 1 and 2 for the 1975-2016 period is 0.71

countries with more than five hundred and fifty nuclear weapons, and others as well as between countries with a few nuclear weapons versus countries with a moderate number.[4] I then transform the ordinal variable into annual world shares of nuclear weapons. Data on nuclear weapon stockpiles come from the Bulletin of Atomic Scientists and covers the period 1945-2019 (Kristensen and Norris, 2013; Zala, 2019). I use Bell and Miller (2015)'s coding for the first year a country has nuclear weapons.

Because of the relationship between nuclear weapons and ballistic missiles[5] and to reduce the variance and skewness of these components I take the average of these two components before including it in the final calculation. Practically speaking, this gives each of these components half the weight of the other components. The averaging of these two components is especially helpful prior to the mid 1970s when fewer than ten countries have these weapons. If one does not average these components prior to the final calculation, one will likely overstate how much power a country has. For example, from 1945-1948, the US is the only country with nuclear weapons, giving it a one hundred percent share on this component. Similarly, the Soviet Union is the only country with ballistic missiles from 1947-1950. Taking the average of nuclear weapons and ballistic missiles balances out these factors in this time period. Other analysts are free to choose alternative aggregation protocols with the data supplied here.

Finally, I use the four indicators just described to create a country-year measure of military power for the period 1865-2019.[6] MMP is a country's annual average of naval, air, ballistic missile/nuclear weapons, and land power.[7] Table

[4]I do not create a continuous indicator because we do not know the exact number of nuclear weapons each state possesses and above some threshold there are likely decreasing returns.

[5]Approximately 89 percent of country-years with nuclear weapons also have ballistic missiles, but only 31 percent of ballistic missile country-years overlap with nuclear weapons possession.

[6]The COW state system data ends in 2016. I assume all countries in existence in 2016 continue to exist through 2019.

[7]Only Land Power 2 has missing values, due to missing data on military expenditures or personnel. When land power 2 is missing, I use the average of the remaining components.

1 summarizes the indicators in MMP for each time period. The dataset includes each sub-component as well as the composite MMP indicator. Users may create alternative aggregate indicators based on these components.

[Table 1 about here.]

There are several novel features of MMP. First, MMP is a large-N country-year indicator based only on military factors. Indeed, this is likely the most important difference between MMP and more common large-N measures of power. Further, as a measure of military power, the focus on existing weapons gives MMP more content validity than other common measures of power that focus more on economic and population factors. Second, MMP focuses primarily on weapons systems. Given the importance of technology to military matters, particularly the outcomes of wars, MMP should prove to be particularly useful as a measure of military power (McNeill, 1984). Third, the land power indicator introduced here differs significantly from the dominant measure of land power, which focuses on number of troops (Rasler and Thompson, 1984).

A measure should be judged based on its reliability and validity. MMP has high reliability in the same sense as the major alternative measures of military power. Armed with the data for each component one can easily recreate it. Like its competitors, MMP is based on a transparent process and its component data are readily available. In the next section I assess the validity of the MMP measure and compare it to other large-N power indicators, specifically Net Resources, DOE, and the COW military expenditures indicator. Net Resources and DOE have each proved superior to the COW CINC indicator as measures of power. Military expenditures is rarely used as an independent variable or a measure of military power in quantitative conflict research, but it is a straightforward measure of military power and does not have the major demographic drawbacks of the CINC indicator.

Surface Validity

Since 1971 when the People's Republic of China was given China's Security Council seat, MMP identifies the UN Security Council as the top five military powers. This is not the case for the Net Resources or Military Expenditure measures. The MMP and Net Resources measures show Israel as stronger than Egypt for the 1967 and 1973 wars; military expenditures does not. Net Resource and military expenditure measures ranked Kuwait as stronger than Iraq in 1990. MMP says that Iraq had considerably more military power. As these examples illustrate, MMP has credible surface validity.

The land power indicators also seem to have reasonable face validity. The Soviet Union, for example, has greater land power than the United States during the Cold War, from 1947-1990. Israel has greater land power than Egypt in 1973. Similarly, China has very little land power in the late nineteenth and early twentieth centuries. The most common measure of army strength in quantitative research is Rasler and Thompson (1984)'s army size measure, but it is only available for eight countries since 1870 and has some questionable rankings. Using this indicator, Levy and Thompson (2010) record Russia as the strongest European army from 1915-1924. The indicator proposed here suggests Germany had the strongest army from 1914-1917 and France slightly stronger than Germany in 1918, which seems more consistent with the results of World War I.

Context Validity

Military Power and War Outcomes

Blainey (1988, 113) famously wrote that "warfare is the one convincing way of measuring the distribution of power." This is only true if one has an expansive and tautological definition of power. The problem with Blainey's statement is that things like force structure, strategy, training, and civil-military relations should be viewed as distinct from military power, otherwise we cannot analytically discriminate between these concepts. Nevertheless, Blainey's statement contains a nugget of truth. A worthwhile measure of military power should have a positive correlation with victory in war. To evaluate the relationship between MMP and victory in war, I follow the same procedure as Beckley (2018). If one country has more power than the other, then it should be more likely to prevail in war. My list of wars and war participants comes from the Interstate War Data (Reiter et al., 2016).

Table 2 shows the percentage of wars MMP, Net Resources, and military expenditures correctly predict. MMP predicts about 80% of bilateral wars correctly, while Net Resources predicts 69% and military expenditures about 75%. Indeed, there are no cases in which Net Resources or military expenditures correctly predict a war that MMP incorrectly predicts.[8] Interestingly, from 1865-1945 naval tonnage alone predicts bilateral wars as well as military expenditures (71% correct) and slightly better than Net Resources (64% correct).

[Table 2 about here.]

Military Power as a Control Variable

It is common for international relations research to include an indicator for military power in an empirical model. I compare the performance of MMP as a control variable to other measures of power in thirty-one studies. I focus

[8] I focus on bilateral wars because the power comparison is clearer here than it is when we include all pairs for multilateral wars. For example, if we include multilateral wars, then we include Oman versus Iraq in 1991. It is misleading to say that Oman deserves a victory in this contest and Iraq a defeat. If we drop minor U.S. allies in multilateral wars after World War II, then MMP predicts 73% correct while Net Resources correctly predicts 71% and expenditures 70%.

on differences in Akaike Information Criteria (AIC) (Akaike, 1974). AIC is a measure of the distance between a true model and the data; thus, lower AIC scores are preferred. If a model with one measure of power has an AIC at least three points less than the model with another measure of power, then I record that measure as outperforming the other (Burnham and Anderson, 2004). If the difference in AIC is less than three points, I report no difference in model performance.

Table 3 summarizes this power-as-control-variable analysis. This analysis examines all of the studies reported in Beckley (2018) and all of the non-directed dyad year studies from Carroll and Kenkel (2019).[9]. MMP performs better than Net Resources, about the same as military expenditures, and not quite as well as DOE (Table 3). For example, in thirteen replications MMP has a lower AIC score than Net Resources, while Net Resources only has a lower AIC score in six replications. When compared to military expenditures, MMP models have a lower AIC score in six cases, while expenditures has a lower AIC in six cases. DOE has an AIC score three or more points lower than MMP thirteen times, while MMP performs better in six cases. While DOE generally improves model performance over MMP, in the seven non-directed dyad replications, MMP outperforms DOE in three cases with DOE only significantly lowering AIC in two cases. When choosing a measure of power to control for researchers should think about what base or aspect of power is most relevant for their research. In some cases researchers will want a measure of latent power and should choose DOE or Net Resources. In other cases, one may want a more direct measure of military power, in which case MMP or military expenditures are more applicable.

[Table 3 about here.]

[9] I use Beckley's replication data and file with two changes. In replicating Weeks (2012) I focus on Model 2 in Table 1, the same model as Carroll and Kenkel (2019). In replicating Grauer and Horowitz (2012) I focus on Model 3. Beckley (2018) examined Model 4, but it has about 23% fewer cases than Model 3.

Military Power and Militarized Threats

As a final application, I examine the relationship between military power and militarized threats. As noted previously, diplomacy often involves the threat of military force. As the practitioner Frederick the Great and the scholars Blainey (1988) and Schelling (1966) recognized, the threat and limited application of military power is central to much of international relations. Dahl made a similar point. Power "must be exploited in some fashion if the behavior of others is to be altered. The means or instruments of such exploitation are numerous; often they involve threats or promises to employ the base in some way and they may involve actual use of the base" (Dahl, 1957, 203). If MMP is a valid measure of military power, then it should have a robust correlation with the making of military threats. To assess this expectation, I posit the following logistic regression model of militarized threat initiation:

$$\begin{aligned} \text{Threat Initiation}_{it} = \alpha + \beta_1 * (\text{Sum Power States A and B})_{it} \\ + \beta_2 * (\text{State A's Share of Dyadic Power})_{it} \\ + \beta_k * \delta_k it + \quad {}_{it} \end{aligned}$$

The unit of analysis is the directed-dyad year. I examine politically relevant dyads and two different dependent variables. The first is the initiation of a dyadic militarized interstate dispute (Maoz et al., 2019). The second is the initiation of a militarized compellent threat (Sechser, 2011). (In the appendix I describe the operationalization of each variable as well as the data sources.) We are interested in whether a model with military power measured using MMP performs better, worse, or about the same as a model with other measures of military power. As before, the measure of performance is the Akaike Information

Criteria (AIC).

Table 4 summarizes this analysis of threat models. Models with MMP outperform models with Net Resources or DOE for each measure of threat. Models with MMP and Military Expenditures perform about the same when threat is measured with the militarized compellent threat data, but models with MMP perform better than military expenditures when threat is measured with the dyadic militarized interstate dispute data. While not shown in Table 4, MMP (sum of State A and B and A's share of dyadic power) is statistically significant and positive in all models. As relative military power increases a country is more likely to initiate a militarized threat.

[Table 4 about here.]

Discussion and Conclusion

Military power is a central concept in international relations, yet it is difficult to measure. In large-N research, researchers often employ broad measures of power that emphasize economic and demographic features. I propose a measure of military power based primarily on weapons systems. The resulting measure, called MMP for Material Military Power, incorporates data on naval warships, fighter aircraft, tanks and armored fighting vehicles, ballistic missiles, and nuclear weapons.

As a measure of country military power, MMP has better surface validity than alternative large-N measures. MMP, for example, identifies the United States as the world's strongest military power today and identifies the members of the Security Council as the top military powers in the world. As a measure of military power, its substantive components are more valid than broad measures of national power. Further, MMP correctly predicts a higher percentage of bilateral wars than economic measures of power or a measure based only on military expenditures. When used as a control variable, models with MMP perform slightly better than models with Net Resources but not quite as well as DOE (Carroll and Kenkel, 2019). When including a control for power, researchers will have to think carefully about what aspects of power they want to control for. Finally, I find that MMP performs better than Net Resources, DOE, and military expenditures in models of threat initiation.

In conclusion, MMP will be especially useful for understanding variation in military power across countries and over time, comparing military to economic power, forecasting changes in military power, understanding the effects of force structure and the relationship between military power and conflict processes.

References

Akaike, H. (1974). A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike*, pp. 215–222. Springer.

Anders, T., C. J. Fariss, and J. N. Markowitz (2020). Bread before guns or butter: Introducing surplus domestic product (sdp). *International Studies Quarterly 64*(2), 392–405.

Beckley, M. (2018). The power of nations: Measuring what matters. *International Security 43*(2), 7–44.

Bell, M. S. and N. L. Miller (2015). Questioning the effect of nuclear weapons on conflict. *Journal of Conflict Resolution 59*(1), 74–92.

Blainey, G. (1988). *Causes of war.* Simon and Schuster.

Bremer, S. (1992). Dangerous dyads: conditions affecting the likelihood of interstate war, 1816-1965. *Journal of Conflict Resolution 36*(2), 309–341.

Burnham, K. P. and D. R. Anderson (2004). Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research 33* (2), 261–304.

Carroll, R. J. and B. Kenkel (2019). Prediction, proxies, and power. *American Journal of Political Science 63*(3), 577–593.

Center for Strategic and International Studies (2021). Missile threat. https://missilethreat.csis.org/missile/.

Crisher, B. B. and M. Souva (2014). Power at sea: a naval power dataset, 1865–2011. *International Interactions 40*(4), 602–629.

Dahl, R. A. (1957). The concept of power. *Behavioral science 2*(3), 201–215.

Davenport, K. (2017). Worldwide ballistic missile inventories. *Arms Control Association*.

for Strategic Studies, I. (1975-2020). *The Military Balance*. Institute for Strategic Studies.

Grauer, R. and M. C. Horowitz (2012). What determines military victory? testing the modern system. *Security Studies 21*(1), 83–112.

Kristensen, H. M. and R. S. Norris (2013). Global nuclear weapons inventories, 1945–2013. *Bulletin of the Atomic Scientists 69*(5), 75–81.

Levy, J. S. and W. R. Thompson (2010). Balancing on land and at sea: do states ally against the leading global power? *International Security 35* (1), 7–43.

Lyall, J. and I. Wilson (2009). Rage against the machines: Explaining outcomes in counterinsurgency wars. *International Organization 63*(1), 67–106.

Maoz, Z., P. L. Johnson, J. Kaplan, F. Ogunkoya, and A. 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.

Mastanduno, M., D. A. Lake, and G. J. Ikenberry (1989). Toward a realist theory of state action. *International Studies Quarterly 33*(4), 457–474.

McNeill, W. H. (1984). *The pursuit of power: Technology, armed force, and society since AD 1000*. University of Chicago Press.

Mearsheimer, J. (1985). *Conventional deterrence*. Cornell University Press.

Mettler, S. A. and D. Reiter (2013). Ballistic missiles and international conflict. *Journal of Conflict Resolution 57*(5), 854–880.

Narang, V. and C. Talmadge (2018). Civil-military pathologies and defeat in war: Tests using new data. *Journal of Conflict Resolution 62*(7), 1379–1405.

Nuclear Threat Initiative (2021). Country reportst. https://www.nti.org/countries/.

Parker, G. (1996). *The military revolution: Military innovation and the rise of the West, 1500-1800*. Cambridge University Press.

Rasler, K. A. and W. R. Thompson (1984). *The great powers and global struggle, 1490-1990*. University Press of Kentucky.

Reiter, D. (1999). Military strategy and the outbreak of international conflict: Quantitative empirical tests, 1903-1992. *Journal of Conflict Resolution 43* (3), 366–387.

Reiter, D. and C. Meek (1999). Determinants of military strategy, 1903–1994: A quantitative empirical test. *International Studies Quarterly 43*(2), 363–387.

Reiter, D., A. C. Stam, and M. C. Horowitz (2016). A revised look at interstate wars, 1816–2007. *Journal of Conflict Resolution 60*(5), 956–976.

Saunders, R. J. and M. Souva (2019). Command of the skies: An air power dataset. *Conflict Management and Peace Science*, 0738894219863348.

Schelling, T. C. (1966). *Arms and Influence*, Volume 190. Yale University Press.

Sechser, T. S. (2011). Militarized compellent threats, 1918–2001. *Conflict Management and Peace Science 28*(4), 377–401.

Sechser, T. S. and E. N. Saunders (2010). The army you have: The determinants of military mechanization, 1979–2001. *International Studies Quarterly 54*(2), 481–511.

Singer, J. D., S. Bremer, J. Stuckey, et al. (1972). Capability distribution, uncertainty, and major power war, 1820-1965. *Peace, war, and numbers 19*, 48.

Van Creveld, M. (2010). *Technology and war: From 2000 BC to the present*. Simon and Schuster.

Weeks, J. L. (2012). Strongmen and straw men: Authoritarian regimes and the initiation of international conflict. *American Political Science Review* , 326–347.

Zala, B. (2019). How the next nuclear arms race will be different from the last one. *Bulletin of the Atomic Scientists 75*(1), 36–43.

Table 1: Components of Material Military Power (MMP) By Time Period

Time Period	MMP
1975-2019	Naval Tonnage, Air Power, Land Power 1, Ballistic Missiles/Nuclear Weapons
1965-1974	Naval Tonnage, Air Power, Land Power 2 Ballistic Missiles/Nuclear Weapons
1945-1964	Naval Tonnage, Land Power 2 Ballistic Missiles/Nuclear Weapons
1865-1944	Naval Tonnage, Land Power 2

Notes: MMP is the mean of annual world shares of the specified components for each time period.

Land Power 1 is based on world shares of mechanized armor vehicles.

Land Power 2 is an estimate of land power.

Table 2: Percentage of Bilateral Wars Correctly Predicted: MMP, Net Resources, and Military Expenditures

Time Period	MMP	Net Resources[1]	Military Expenditures[2]	N
1865-2007	80	69	75	36

[1] From Beckley (2018)
[2] From NMCv6 (Singer, et al 1972)

Table 3: Models with MMP as a Control Variable versus Net Resources and DOE[1]

Competing Measures	MMP AIC < 3 (MMP wins)	MMP AIC > 3 (MMP loses)	AICs within 3 (Tie)
MMP vs Net Resources[2]	13	5	13
MMP vs DOE[3]	6	12	6
MMP vs Military Expenditures[4]	6	6	19

[1] I use the same models and data sets as Beckley (2018) and Carroll and Kenkel (2019)
[2] From Beckley (2018)
[3] From Carroll and Kenkel (2019). There are fewer replications with DOE because it can only be used in dyadic designs.
[4] From National Material Capabilities data, version 6, Singer et al (1972).

Table 4: AIC Values for Models of Threat Initiation: MMP versus Other Power Measures

DV and Sample	MMP	Net Resources[1]	DOE[2]	Mil Exp[5]
MCT, 1918-2001 Politically Relevant	2492	2530	2502	2491
Dyadic MID, 1946-2011 Politically Relevant	13992	14164	14181	14006
Dyadic MID, 1865-1945 Politically Relevant	6486	6555	6603	6503

[1] From Beckley (2018)
[2] From Carroll and Kenkel (2019)
[3] From Sechser (2011)
[4] From Maoz et al. (2019)
[5] From Singer et al. (1972)

Appendix for Material Military Power: A Country-Year Measure of Military Power, 1865-2019

This appendix includes the following:

- a discussion of the measures and data sources for the control variables in the directed-dyad threat model,
- a discussion of how to identify bilateral wars,
- a discussion of how I filled-in some missing values on the COW military expenditure indicator,
- a table of summary statistics for MMP and its component indicators,
- tables with the specific AIC scores for the control variable replications,
- figures showing the correlation coefficients between MMP and component indicators,
- a figure showing the correlation between MMP, Net Resources, and Military Expenditures.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

Measurement and Data Sources for Control Variables in Directed-Dyad Threat Model

For the military threat analysis reported in the paper in addition to the power measures, control variables include contiguity, a binary variable equal to one if two states share a land border or are separated by less than 400 miles of water (data from (Stinnett et al., 2002)), distance is the natural log of the great circle distance between state capitals, Democracy A and Democracy B are binary variables equal to one if a state's score on the Polity democracy-autocracy index is greater than five (data from (Marshall et al., 2019)), joint democracy is a binary variable equal to one if both states are democratic, preference similarity is measured as alliance portfolio similarity kappa indicator ((H ̈age, 2011)) (data from (Chiba et al., 2015)), peace years, peace years squared, and peace years cubed.

Discussion of War Outcomes

To examine the relationship between military power and war outcomes, I focus on bilateral wars because the comparison is more straightforward. To identify bilateral wars, I use the IWD larger war id variable. Beckley (2018) identifies bilateral wars by first dropping World War II and then examining whether IWD lists a second adversary or an ally of the initiator. This leads him to identify four additional wars as bilateral: Mecklenberg Scherwin versus Bavaria in the Seven Weeks War, Germany versus Belgium in World War 1, Greece versus Turkey in the First Balkan War, and Egypt versus Israel in the 1948-49 Arab-Israeli War. While none of these are bilateral wars, when I use this methodology, MMP correctly predicts 75% of the wars correctly, Net Resources predicts 70%, and military expenditures predicts 70%.

Filling-in Missing Military Expenditure Data

The NMC data has 1400 missing values for military expenditures between 1865 and 2016, and 330 missing values on military personnel. Carroll and Kenkel (2019) use multiple imputation to fill-in these series. Beckley (2018)'s net re- sources measure is also based on data generated in part from multiple imputa- tion. My approach is closer to Gleditsch (2002). First, I change six odd values in the NMC data, discussed below. Second, I use linear interpolation. This reduces the number of missing expenditure values to 585. In a number of cases, there is a missing value for the first year a country is in the system. Given the high correlation between expenditures from one year to the next, I fill forward or backward one-year. This leaves me with 506 country years between 1865 and 2012 that are missing expenditure data. The six odd military expenditure cases that I change are North Korea (1948 and 1958), Thailand (2001), and Gabon (2002-2004). North Korea in 1948 is listed as having zero military expenditures, despite having 200,000 military personnel. I recode the 1948 value to missing. For the next nine years the value of North Korea's expenditures is listed missing. Then in 1958 they are listed as having 2,210,054. In 1959 the value is missing and in 1960 it is 200,000. There is no reason to believe North Korea's military expenditures were eleven times higher in 1958 than in 1960. Since 1959 is also missing, I recode the 1958 value to missing. This gives a continuous series of missing values for North Korea from 1948 through 1959. In 2001 it appears that an extra zero was entered for Thailand's military expenditures. From 2002-2004 it appears that a zero was left off of Gabon's expenditures. The SIPRI data show no major changes in expenditures for Thailand or Gabon for the years in question, further suggesting that the NMC values are the result of a data entry error (PeaceResearchInstitute), 2019). Finally in ten cases the NMC records zero for expenditures but not for military personnel for the first year the country enters the state system. Since personnel have to be paid for, I replace the zero expenditure value with the value in the subsequent year. For military per- sonnel, after interpolation there are 113 missing values. After filling forward and backward one year, there are 78 missing values. Iceland has a missing value for military personnel for 2000 and 2001, zero for all other years. I change the missing value to zero for these two years.

Appendix References

Allen, Michael A & Matthew DiGiuseppe (2013) Tightening the belt: Sovereign debt and alliance formation. *International Studies Quarterly 57*(4): 647–659.

Beckley, Michael (2018) The power of nations: Measuring what matters. *Interna- tional Security 43*(2): 7–44.

Carroll, Robert J & Brenton Kenkel (2019) Prediction, proxies, and power. *American Journal of Political Science 63*(3): 577–593.

Carter, David B & Paul Poast (2017) Why do states build walls? political economy, security, and border stability. *Journal of conflict resolution 61*(2): 239–270.

Carter, Jeff, Michael Bernhard & Glenn Palmer (2012). Social revolution, the state, and war: How revolutions affect war-making capacity and interstate war out- comes. *Journal of Conflict Resolution 56*(3), 439–466.

Chiba, Daina, Jesse C Johnson & Brett Ashley Leeds (2015) Careful commitments: Democratic states and alliance design. *The Journal of Politics 77*(4): 968–982.

Clay, K Chad & Andrew P Owsiak (2016) The diffusion of international border agreements. *The Journal of Politics 78*(2): 427–442.

Colgan, Jeffrey D & Jessica L. Weeks (2015) Revolution, personalist dictatorships, and international conflict. *International Organization*, 163–194.

Dreyer, David R (2010) Issue conflict accumulation and the dynamics of strategic rivalry. *International Studies Quarterly 54*(3): 779–795.

Early, Bryan R (2012) Alliances and trade with sanctioned states: A study of us economic sanctions, 1950–2000. *Journal of Conflict Resolution 56* (3): 547–
572.

Findley, Michael G, James A. Piazza & Joseph K. Young (2012) Games rivals play: Terrorism in international rivalries. *The Journal of Politics 74* (1): 235–248.

Fordham, Benjamin O (2008) Power or plenty? economic interests, security concerns, and American intervention. *International Studies Quarterly 52* (4): 737–758.

Fuhrmann, Matthew & Todd S Sechser (2014) Signaling alliance commitments: Hand- tying and sunk costs in extended nuclear deterrence. *American Journal of Political Science 58* (4): 919–935.

Gartzke, Erik (2007) The capitalist peace. *American journal of political sci- ence 51* (1): 166–191.

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

Grauer, Ryan & Michael C. Horowitz (2012) What determines military victory? testing the modern system. *Security Studies 21* (1): 83–112.

Häge, Frank M (2011) Choice or circumstance? adjusting measures of foreign policy similarity for chance agreement. *Political Analysis*: 287–305.

Haynes, Kyle (2012) Lame ducks and coercive diplomacy: Do executive term limits reduce the effectiveness of democratic threats? *Journal of Conflict Resolution 56* (5): 771–798.

Horowitz, Michael C & Alan C. Starn (2014) How prior military experience influ- ences the future militarized behavior of leaders. *International Organization*: 527–559.

Huth, Paul K, Sarah E Croco & Benjamin J Appel (2013) Bringing law to the table: Legal claims, focal points, and the settlement of territorial disputes since 1945. *American Journal of Political Science 57* (1): 90–103.

Kinne, Brandon J & Nikolay Marinov (2013) Electoral authoritarianism and credible signaling in international crises. *Journal of Conflict Resolution 57* (3): 359– 386.

Kroenig, Matthew (2013) Nuclear superiority and the balance of resolve: Explaining nuclear crisis outcomes. *International Organization*: 141–171.

Lupu, Y (2016) Why do states join some universal treaties but not others? an analysis of treaty commitment preferences. *Journal of Conflict Resolu- tion 60* (7): 1219–1250.

Marshall, Monty G, Ted R Gurr, & Keith Jaggers (2019) Polity iv project: Political regime characteristics and transitions, 1800-2018. *Center for Systemic Peace*.

Moon, Chungshik & Mark Souva (2016) Audience costs, information, and credible commitment problems. *Journal of Conflict Resolution 60* (3): 434–458.

Narang, Vipin & Caitlin Talmadge (2018) Civil-military pathologies and defeat in war: Tests using new data. *Journal of Conflict Resolution 62* (7): 1379–1405.

Owsiak, Andrew P (2012) Signing up for peace: International boundary agree- ments, democracy, and militarized interstate conflict. *International Studies Quarterly 56* (1): 51–66.

Park, Johan & Michael Colaresi (2014) Safe across the border: The continued signif- icance of the democratic peace when controlling for stable borders. *Interna- tional Studies Quarterly 58* (1): 118–125.

Peace Research Institute, S. S. I. (2019) Yearbook 2019 armaments, disarmament and international security.

Powell, Jonathan M (2014) Regime vulnerability and the diversionary threat of force. *Journal of Conflict Resolution 58*(1): 169–196.

Salehyan, Idean (2008) No shelter here: Rebel sanctuaries and international conflict. *The Journal of Politics 70*(1): 54–66.

Shelef, Nadav G (2016) Unequal ground: Homelands and conflict. *International Organization*: 33–63.

Singer, J David, Stuart Bremer & John Stuckey (1972) Capability distribution, uncertainty, and major power war, 1820-1965. *Peace, war, and numbers 19* (48):9.

Sobek, David, M Rodwan Abouharb & Christopher G Ingram (2006) The human rights peace: How the respect for human rights at home leads to peace abroad. *The Journal of Politics 68*(3): 519–529.

Stinnett, Douglas M, Jaroslav Tir, Paul F Diehl, Phil Schafer & Charles Gochman (2002) The correlates of war (cow) project direct contiguity data, version 3.2. *Conflict Management and Peace Science 19*(2): 59–67.

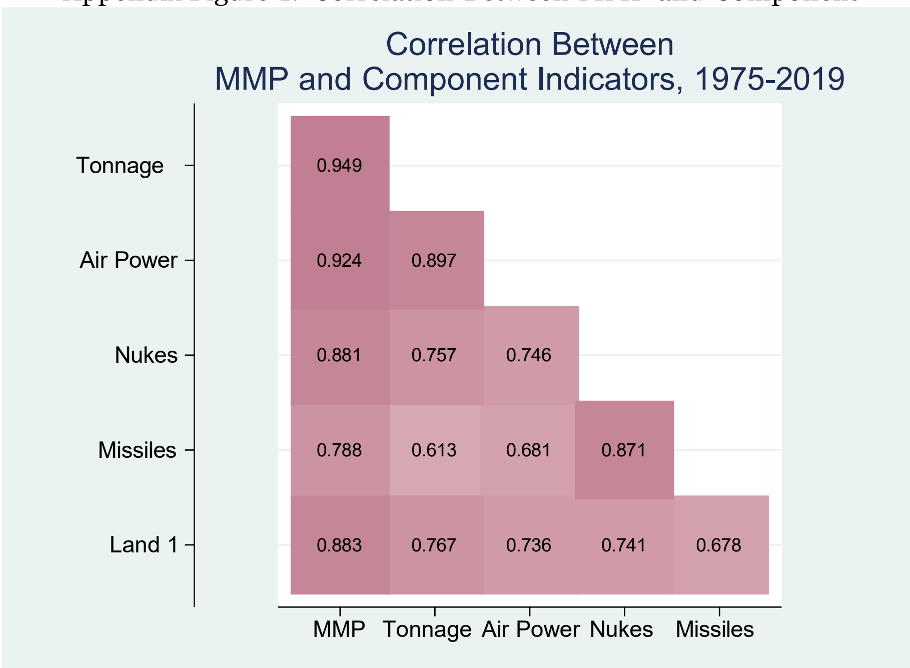
Way, Christopher & Jessica L Weeks (2014) Making it personal: regime type and nuclear proliferation. *American Journal of Political Science 58* (3): 705–719.

Weeks, Jessica L (2012) Strongmen and straw men: Authoritarian regimes and the initiation of international conflict. *American Political Science Review 106 (2):* 326–347.

Weisiger, Alex & Karen Yarhi-Milo (2015) Revisiting reputation: How past actions matter in international politics. *International Organization 69* (2): 473–495.

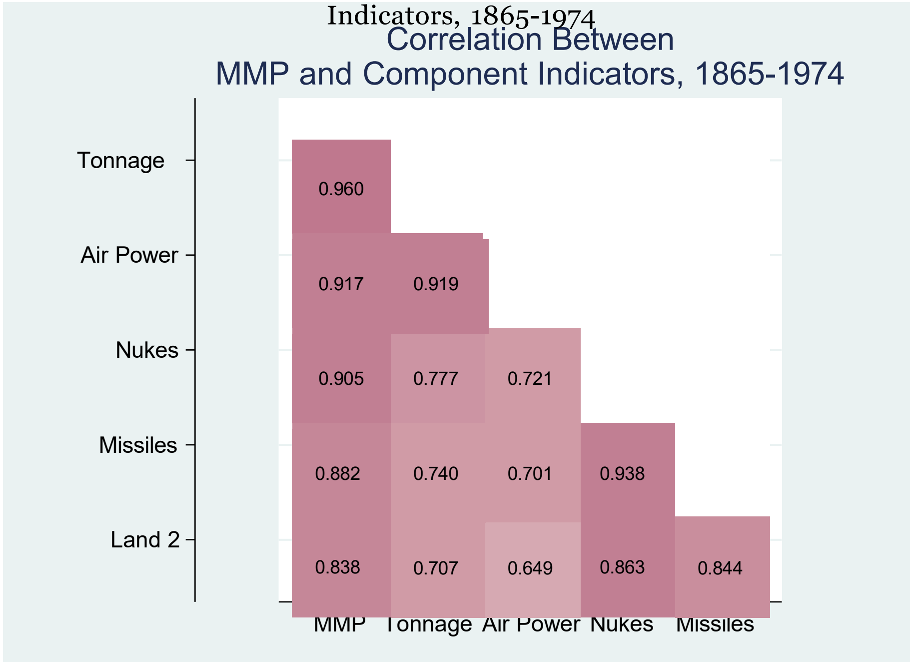
Wright, Thorin M & Paul F Diehl (2016) Unpacking territorial disputes: Domestic political influences and war. *Journal of Conflict Resolution 60* (4): 645–669.

Appendix Figure 1: Correlation Between MMP and Component

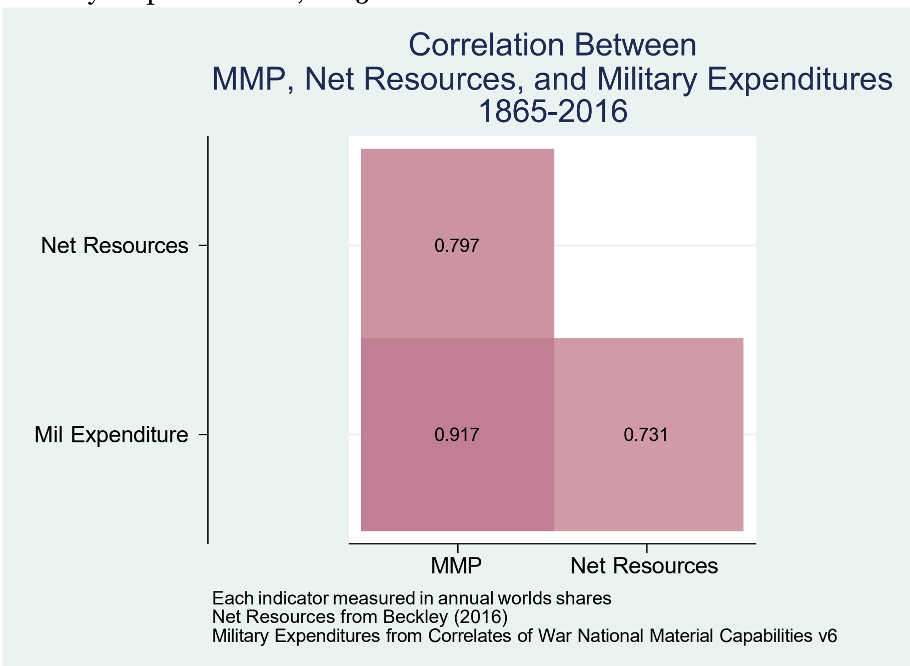


Indicators, 1975-2019

Appendix Figure 2: Correlation Between MMP and Component Indicators, 1865-1974



Appendix Figure 3: Correlation Between MMP, Net Resources, and Military Expendi- tures, 1865-2016



Appendix Table 1: Summary Statistics

Variable	N	Mean	SD	Min	Max	Years Covered
MMP	14504	.0086	.0319	0	.4857	1865-2019
Tonnage	14504	.0106	.0503	0	.7184	1865-2019
Air Power	9340	.0058	.0180	0	.3226	1965-2019
Ballistic Missiles	11037	.0066	.0359	0	1	1945-2019
Nuclear Weapons	11037	.0067	.0410	0	1	1945-2019
Land Power 1	8007	.0056	.0212	0	.3617	1975-2019
Land Power 2	6468	.0067	.0253	0	.5265	1865-1974

Appendix Table 2: Control Variable Analysis: MMP versus Other Power Measures

Study	MMP AIC	Net Resources[1] AIC	DOE[2] AIC	Mil Exp[3] AIC
Allen and DiGiuseppe (2013), Table 4, Model 1	1531.6	1528.4		1531.7
Carter and Poast (2017), Table 2, Model 1	2502.8	2496.5	2348.3	2505.7
Carter et al. (2012), Table 2, Model 5	378.9	382.2	375.3	381.8
Clay and Owsiak (2016), Table 1, Model 4	504.9	496.5	502.4	504.1
Colgan and Weeks (2015), Table 2, Model 1	6039.8	6058.8		6063.0
Early (2012), Table 1, Model 6	97561.7	97575.7	97551.5	97575.3
Findley et al. (2012), Table 1, Model 1	7674.6	7648.4	7642.4	7669.8
Fuhrmann and Sechser (2014), Table 2, Model 3	2610.0	2609.4	2582.3	2609.6
Grauer and Horowitz (2012), Table 2, Model 4	97.2	97.9		99.1
Haynes (2012), Table 1, Model 1	245.4	240.8	240.2	244.6
Horowitz and Starn (2014), Table 1, Model 1	8752.1	9052.6		8739.5
Huth et al. (2013), Table 4	78.9	76.3	79.7	80.5
Kinne and Marinov (2013), Table 2, Model 1	1723.3	1722.5	1708.5	1720.6
Kroenig (2013), Table 3, Model 5	64.3	64.8	66.8	63.2
Lupu (2016)	21.3	21.4		20.4
Moon and Souva (2016), Table 1, Model 2	263.4	261.9	259.2	264.7
Narang and Talmadge (2018), Table 2, Model 5	90.8	92.1		90.9
Powell (2014), Table 3 Full Model	670.8	669.6	653.1	669.0
Fuhrmann and Sechser (2014), Table 1, Model 2	261.2	260.3	265.6	262.8
Shelef (2016), Table 3, All MIDs	249.5	249.3	249.2	249.7
Way and Weeks (2014), Table 1, Plus Capabilities	425.8	434.6		432.4
Weeks (2012), Table 1, Model 2	15481.9	15530.1	15520.9	15500.0
Weisiger and Yarhi-Milo (2015), Table 1, Model 6	13883.5	13918.6	13917.4	13886.9
Wright and Diehl (2016), Table 2, Model 3	1135.3	1135.6	1072.8	1135.0

[1] From Beckley (2018)
[2] From Carroll and Kenkel (2019)
[3] From National Material Capabilities, version 6; Singer et al. (1972)

Appendix Table 3: Control Variable Analysis: MMP versus Other Power Measures

Study	MMP AIC	Net Resources[1] AIC	DOE[2] AIC	Mil Exp[3] AIC
Fordham (2008), Table 2, Third Column	462.5	614.6	472.0	455.9
Dreyer (2010), Table 2, Model 2	3007.7	3006.8	2983.4	3004.2
Salehyan (2008), Table 1, Model 1	2989.2	2986.6	2987.6	2989.0
Owsiak (2012), Table 3, Model 3	5040.4	5048.7	5044.7	5039.2
Park and Colaresi (2014), Table 1, Model 4	10035.1	10048.7	9952.3	10028.0
Gartzke (2007), Table 1, Model 4	3926.4	3947.3	3934.6	3943.1
Sobek et al. (2006), Table 2, Index Model	5136.2	5154.1	5137.5	5129.1

[1] From Beckley (2018)
[2] From Carroll and Kenkel (2019)
[3] From National Material Capabilities, version 6; Singer et al. (1972)