Disaggregated Defense Spending: Introduction to Data and the Case for Systematic Use

Article *in* SSRN Electronic Journal · January 2022

DOI: 10.2139/ssrn.4241307

4 authors:

Jordan Becker
United States Military Academy West Point

Seth Benson
United States Military Academy West Point

John Paul Dunne
University of Cape Town

Edmund J. Malesky
Duke University

Disaggregated Defense Spending: Introduction to Data and the Case for Systematic Use

Jordan Becker, Seth Benson, John Paul Dunne, Edmund Malesky

Abstract: Theoretical and empirical research on causes and consequences of defense spending is plentiful. Most of this research uses "top line" defense spending data, either as a share of GDP or as a raw monetary figure. Empirical research has been limited, however, by the "blunt" nature of this data, which does not help explain *what* countries are spending on. We introduce a dataset that provides information on disaggregated defense spending from 35 NATO and EU members over as many as 51 years. We discuss the main features of this data in the paper, and the replication files will enable other scholars to automate accessing it in the future. In addition to automating the extraction of NATO and European Defence Agency (EDA) data on overall military expenditures, we make data on equipment, personnel, operating, and infrastructure spending available in a single data set. We illustrate the utility of the disaggregated defense spending dataset by replicating canonical and newer analyses using both the overall data and its disaggregated components. Findings differ depending on which type of spending is considered. We find that differences in the relationship between national wealth and defense spending depend on the category of spending considered, as does the tendency toward "free-riding." These exercises shed new light on seminal theories on burden sharing and the political economy of security. Our initial analysis suggests that disaggregating defense spending is likely to improve analysis of old and emerging research questions of considerable policy importance, and points to several opportunities to do so.

Not all defense expenditures are alike. Scholars (McInnis & Fata, 2022) and practitioners (NATO, 2022) agree that *how* countries spend on defense is as important – if not more important – than *how much* they spend. NATO and the EU divide defense spending into four categories: equipment, personnel, operating and maintenance (O&M), and infrastructure. While use of these four categories in existing research is limited, scholars have already demonstrated that both demand for defense spending (Bove & Cavatorta, 2012; Becker, 2019a) and the economic effects of that spending (Becker & Dunne, 2021) differ across categories. By replicating canonical and recent studies with disaggregated data, we help clarify this phenomenon more systematically. By automating the process of gathering disaggregated data from the two sources that make it available (EDA, 2018; NATO, 2021), we offer a way for future researchers to conduct research with the most current data going forward without manually transferring the data from NATO and EDA websites into their datasets, which is time consuming and error-prone.

This special data feature proceeds as follows: first, we introduce the data set and describe the years for which disaggregated defense spending is available for each of the 35 countries for which data is available at all. Second, we articulate a conceptual and theoretical case for disaggregation, highlighting its utility in addressing questions that policymakers are interested in as well as areas in which scholars have identified shortcomings with the use of overall military spending as a share of GDP. Third, we explain our technique for automating the data collection effort by scraping figures directly from the NATO and EDA websites. Fourth, we explore the utility of disaggregation by replicating canonical and more recent studies. Specifically, we replicate Olson & Zeckhauser's seminal 1966 study on defense spending in an alliance as a Public Good, Sandler & Hartley's 1999 exposition of the Join Product Model as an update to Olson & Zeckhauser's analysis, and George & Sandler's 2018 spatial model of defense spending. We find considerable heterogeneity in results by category of defense expenditure across all three replications. Our concluding section offers some thoughts on future uses of the disaggregated defense spending dataset.

Data Availability

Table 1, below, provides an overview of the years for which data is available by country. Panel A represents availability of all four disaggregated spending categories, and panel B represents availability of equipment spending alone – NATO allies made this data available earlier, and EDA members continued reporting it after discontinuing reporting of disaggregated data for other categories in 2017. Equipment is the one category for which both NATO (2014) and the EU (2016) have explicitly established a target at 20% of overall defense spending.

Table 1: Disaggregated Data Availability, by country

		Panel A: All Categories			Panel B: Equipment		
	Country	First Year	Latest Year	Total Years	First Year	Latest Year	Total Years
1	Albania	2009	2021	13	2009	2021	13
2	Austria	2005	2017	13	2005	2021	17
3	Belgium	1970	2021	52	1965	2021	57
4	Bulgaria	2004	2021	18	2004	2021	18
5	Canada	1970	2021	52	1965	2021	57
6	Croatia	2009	2021	13	2009	2021	13
7	Cyprus	2005	2017	13	2005	2021	17
8	Czech Republic	1999	2021	23	1999	2021	23
9	Denmark	1970	2021	52	1965	2021	57
10	Estonia	2004	2021	18	2004	2021	18
11	Finland	2005	2017	13	2005	2021	17
12	France	1999	2021	23	1999	2021	23
13	Germany	1970	2021	52	1965	2021	57
14	Greece	1970	2021	52	1965	2021	57
15	Hungary	1999	2021	23	1999	2021	23
16	Ireland	2005	2017	13	2005	2021	17
17	Italy	1970	2021	52	1965	2021	57

18	Latvia	2004	2021	18	2004	2021	18
19	Lithuania	2004	2021	18	2004	2021	18
20	Luxembourg	1970	2021	52	1965	2021	57
21	Malta	2006	2016	11	2005	2021	17
22	Montenegro	2010	2021	12	2010	2021	12
23	Netherlands	1970	2021	52	1965	2021	57
24	North Macedonia	2020	2021	2	2020	2021	2
25	Norway	1970	2021	52	1965	2021	57
26	Poland	1999	2021	23	1999	2021	23
27	Portugal	1970	2021	52	1965	2021	57
28	Romania	2004	2021	18	2004	2021	18
29	Slovakia	2004	2021	18	2004	2021	18
30	Slovenia	2004	2021	18	2004	2021	18
31	Spain	1987	2021	35	1984	2021	38
32	Sweden	2006	2016	11	2005	2021	17
33	Turkey	1970	2021	52	1965	2021	57
34	United Kingdom	1970	2021	52	1965	2021	57
35	United States	1980	2021	42	1965	2021	57

The Conceptual and Theoretical Case for Disaggregation

Existing research has highlighted significant variation in defense spending across countries and over time (Dunne & Smith, 2020), particularly within NATO (Hartley & Sandler, 1999; Kuokštytė, Kuokštis & Miklaševskaja, 2021). But, as Figure 1, below, illustrates with ten NATO allies who reported all four disaggregated categories in both 1975 and 2021, differences in the composition of defense budgets are even more stark. The figure also provides some visual insight as to where allies stand regarding the Wales Pledge on Defense Investment – allies agreed at their 2014 Wales Summit to "aim to move towards" spending 2% of GDP on defense and 20% of overall defense spending on equipment.

The red dotted line at the 2% mark represents NATO's overall defense spending guideline, and the red dotted line at .5% represents the figure implicit in the 20% (of overall defense spending) guideline for equipment spending – a country meeting both the 2% and the 20% guidelines would be spending at least .5% of GDP on equipment.

Figure 1: Disaggregated Defense Spending – 1975 and 2021

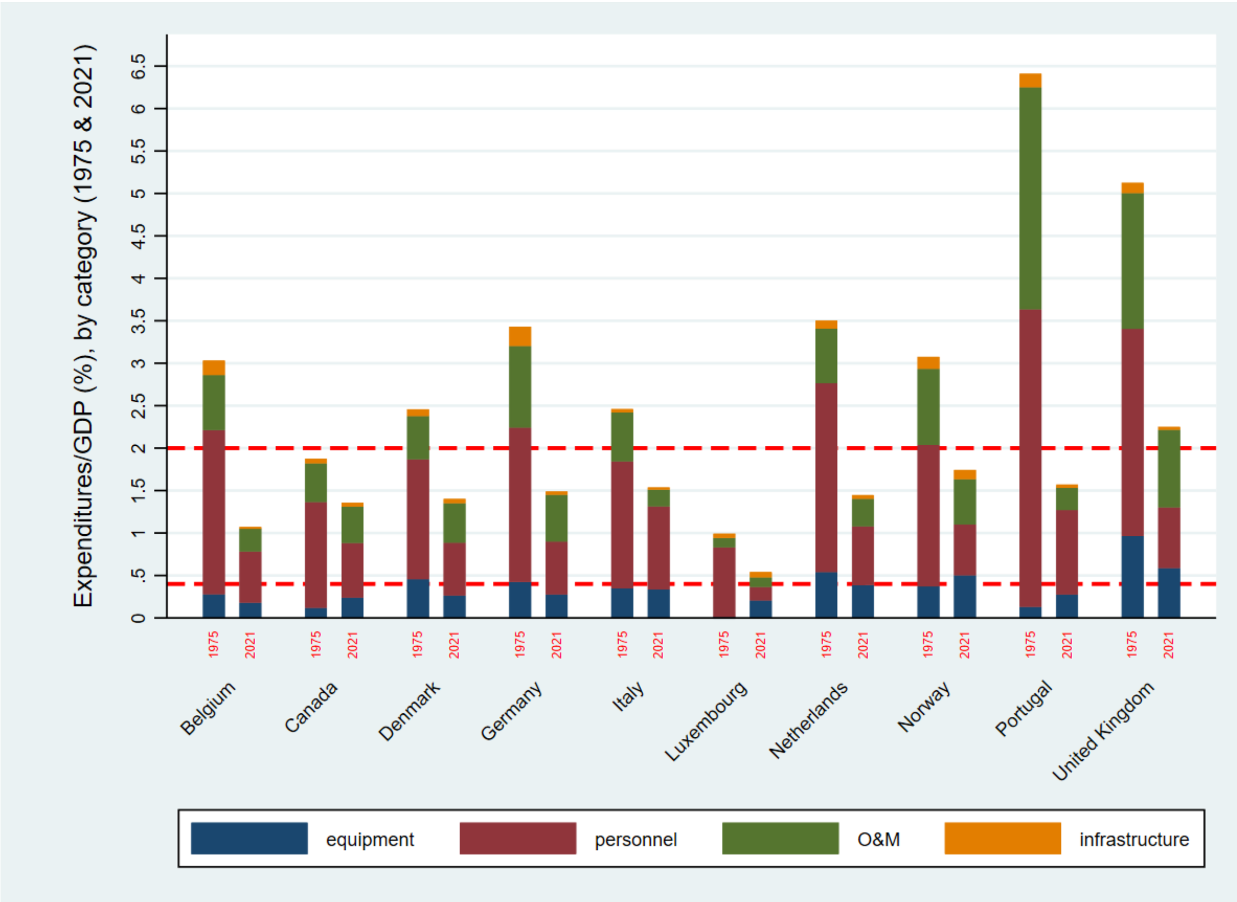
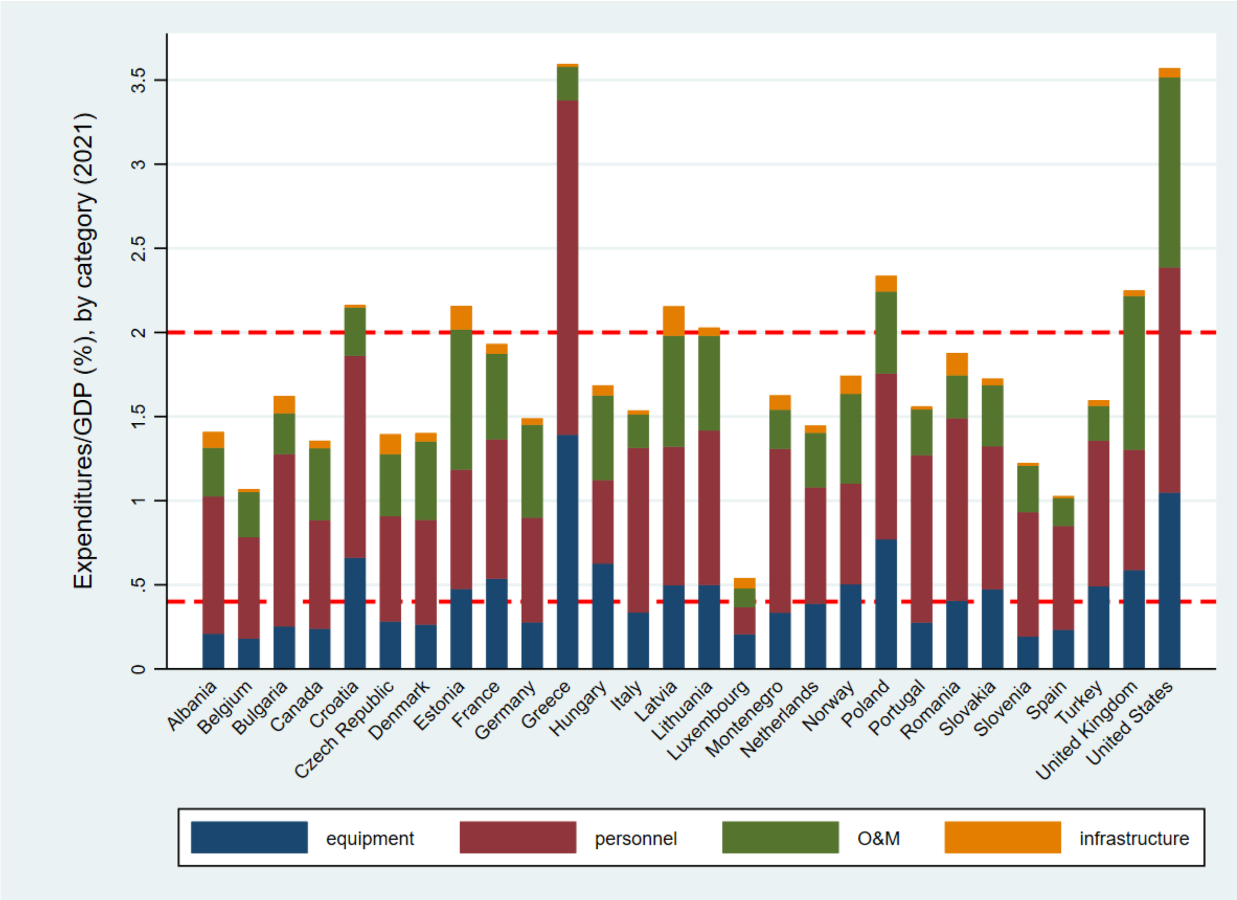


Figure 2 demonstrates that this cross-country variation is significant within the wider sample of countries for which 2021 data is available. Simply visualizing the disaggregated data allows scholars to gain an appreciation of, for example, Wales Pledge compliance at a glance (NATO allies agreed at their 2014 Wales Summit to "aim to move towards" spending 2% of GDP on defense, but also 20% of defense budgets on equipment).

Figure 2: Disaggregated Defense Spending, 2021



The significant difference in not just topline defense spending, but in the composition of defense budgets points intuitively to questions about why such differences exist.

These differences are important theoretically and practically as well. From a theoretical perspective, scholars have consistently pointed to the information about alliance commitments and capabilities that are masked by topline defense spending and pointed to more fine-grained metrics to mitigate these limitations (Cooper & Stiles, 2021). Aggregate spending data may mask differences between personnel spending that can be used to support employment (Becker, 2021), infrastructure spending that can support the stationing and movement of allied troops (Olson & Zeckhauser, 1966), equipment spending that can only have spillover effects in national economies with significant defense industries (Becker, 2019b), and O&M spending that supports military readiness and often deployments in support of shared priorities (Becker & Malesky, 2017). The metrics critical scholars propose, however, are often idiosyncratic or difficult to measure (Kunertova, 2017; Becker, Duda & Lute, 2022; McGerty et al., 2022), leaving researchers at a bit of an impasse. However, many "output metrics (Danish Foreign Ministry, 2015)" are well-predicted by disaggregated defense spending figures (Becker, 2017). Scholars are increasingly turning to such data for these reasons.

As importantly, NATO and the EU have deemed disaggregated defense spending important. NATO formally identified equipment spending as a target prior to identifying it as such at a 2006 Defense Ministerial Meeting (NATO, 2022) and then at the level of heads of state and government at the 2014 Wales Summit (NATO, 2014). The EU followed NATO in identifying equipment spending as a target (European Council, 2016). Subsequently, NATO Secretary General Stoltenberg's "3C's" (2019) of burden-sharing (Cash, Capabilities, and Contributions) point to the utility of overall, equipment, and O&M spending as policy-relevant burden-sharing metrics.

In short, both theoretical and policy debates point strongly toward the importance of disaggregating defense spending as much as possible, and rigorous and replicable social science demands what NATO (2021) calls "a consistent basis of comparison of the defence effort of Alliance members based on a common definition of defence expenditure" – both over time and across countries. While further disaggregation (particularly focused on the acquisition of critical capabilities) is probably desirable, currently available data disaggregated into the four constituent categories used by NATO and the EU marks a significant improvement over simply using topline defense spending figures. Providing access to this data in a simple, automatable, and replicable manner is the primary contribution of this paper and its associated replication files.

Simplifying and facilitating disaggregation

This section briefly discusses our approach to automating the process of collecting disaggregated defense spending data. Both NATO and the European Defence Agency (EDA) make some form of disaggregated defense spending available. NATO has been publishing data on defense expenditures since 1963, and its report that year included data for allies as far back as 1949. The EDA began publishing analogous data in 2005. It ceased publishing data for personnel, Operating and Maintenance (O&M), and infrastructure spending in 2017, but has continued to publish data on equipment spending. Consequently, equipment spending is available for all countries that were members of NATO or the EU through 2021, while the other three categories are only available for NATO allies after 2017.

Prior to 2015, disaggregated defense spending was only available as Adobe Portable Document Format (.pdf) files from NATO and EU websites. In 2016, both began making it available in Microsoft Excel (.xlsx) format, with NATO data from 2009 to the present and EDA data from 2005[1] now publicly available in .xlsx format. However, the data must be retrieved from multiple different files and is not formatted to be readily analyzed from year to year. Using the python "requests" library, we retrieved each of the relevant NATO and EDA .xlsx files from the organizations' respective web pages. Then, using the "pandas" library, we transformed the data stored in each sheet into a single table capturing cleaned and stored data for every country-year. We then combined these to store each variable available for every country-year pair. Variable names and sheet formatting that changed over time on the NATO and EDA websites were corrected for in the

[1] The EDA stopped making personnel, O&M, and infrastructure spending publicly available in 2017; from 2018 to 2021 only equipment spending, among the four categories of disaggregated spending, is available. NATO has continued making all four categories available.

process. The most recent information for each country-year pair is stored so that updated estimates are used in our dataset, and the replication files associated with this paper enable other scholars to gather this data directly themselves – the only requirement is to update the web address for the most recent year in the Stata do file, which uses the Pystata module for this web-scraping task.

The remainder of this paper is dedicated to exploring some basic replications of top defense spending research using disaggregated data gathered as indicated above. It should help other researchers consider how they may best use the disaggregated data in their own work.

Exploring the Empirical Utility of Disaggregation with Replications

The Public Good Model

The utility of the sort of compositional analysis discussed above can be demonstrated with replications of important past defense spending research that has been limited to using aggregated data. We begin this exercise with Olson and Zeckhauser's seminal (1966) study. In that study, Olson and Zeckhauser found that while larger allies spent significantly more on "main forces" (overall defense spending) than did smaller allies, this disparity was not present with infrastructure spending. The finding points toward disaggregating expenditures systematically and over time but doing so was not possible with the data available in 1966.

Table 2 replicates the analysis in Olson and Zeckhauser's table entitled "NATO Statistics: An Empirical Test," and compares them to the same analysis conducted with data available through 2020, the most recent year for which all covariates are available.

Expressed formally, Olson and Zeckhauser's model would read as:

$$MILEX = f(GDP\ Rank) \tag{1}$$

Table 2: Replication of "NATO Statistics: An Empirical Test"

			Mean, 1949-2020	1964 (Olson&Zeck.)		Mean, 1970-2020		
Country	GDP Rank	GDP/Cap Rank	Milburden Rank	Milburden Rank	Equipment Rank	Personnel Rank	O&M Rank	Infrastructure Rank
United States	1	4	1	1	2	13	2	11
Germany	2	8	8	6	7	8	5	4
United Kingdom	3	6	3	3	3	12	2	9
France	4	9	4	4	4	8	9	5
Italy	5	10	10	10	8	4	11	12
Canada	6	8	11	8	9	8	3	7
Netherlands	6	5	8	7	6	7	8	6
Belgium	8	8	11	12	12	3	9	7
Turkey	8	13	5	5	5	9	7	5
Norway	10	2	8	11	5	11	5	3
Denmark	10	3	11	13	8	7	6	9
Greece	11	11	3	9	8	4	11	10
Portugal	12	12	6	2	11	3	10	10
Luxembourg	13	1	14	14	10	4	12	4

Columns 3 and 4 demonstrate a relatively close similarity between rankings in overall military burden in 1966 and the long-term average of that statistic from 1949 to 2020. Only Greece and Portugal differ by more than three ranks, and many countries do not differ at all. Disaggregated spending, however, tells a very different story, with countries varying wildly from their ranks in overall defense spending. This further points to the importance of disaggregated analysis.

Visualizing the correlations between GDP and aggregate and disaggregated defense spending over time helps clarify the differences captured in Table 2 and helps address the question of why the relationship that Olson and Zeckhauser identified failed to hold in the years that followed their analysis (Cornes & Sandler, 1984). This phenomenon led scholars to develop the Joint Product Model of defense spending, which evaluates the *extent* to which defense is public. We analyze the implications of disaggregation to the Joint Product Model in a later section of this paper.

Figure 3: Annual Cross-Country Bivariate Regression Coefficients, GDP and Disaggregated Milex

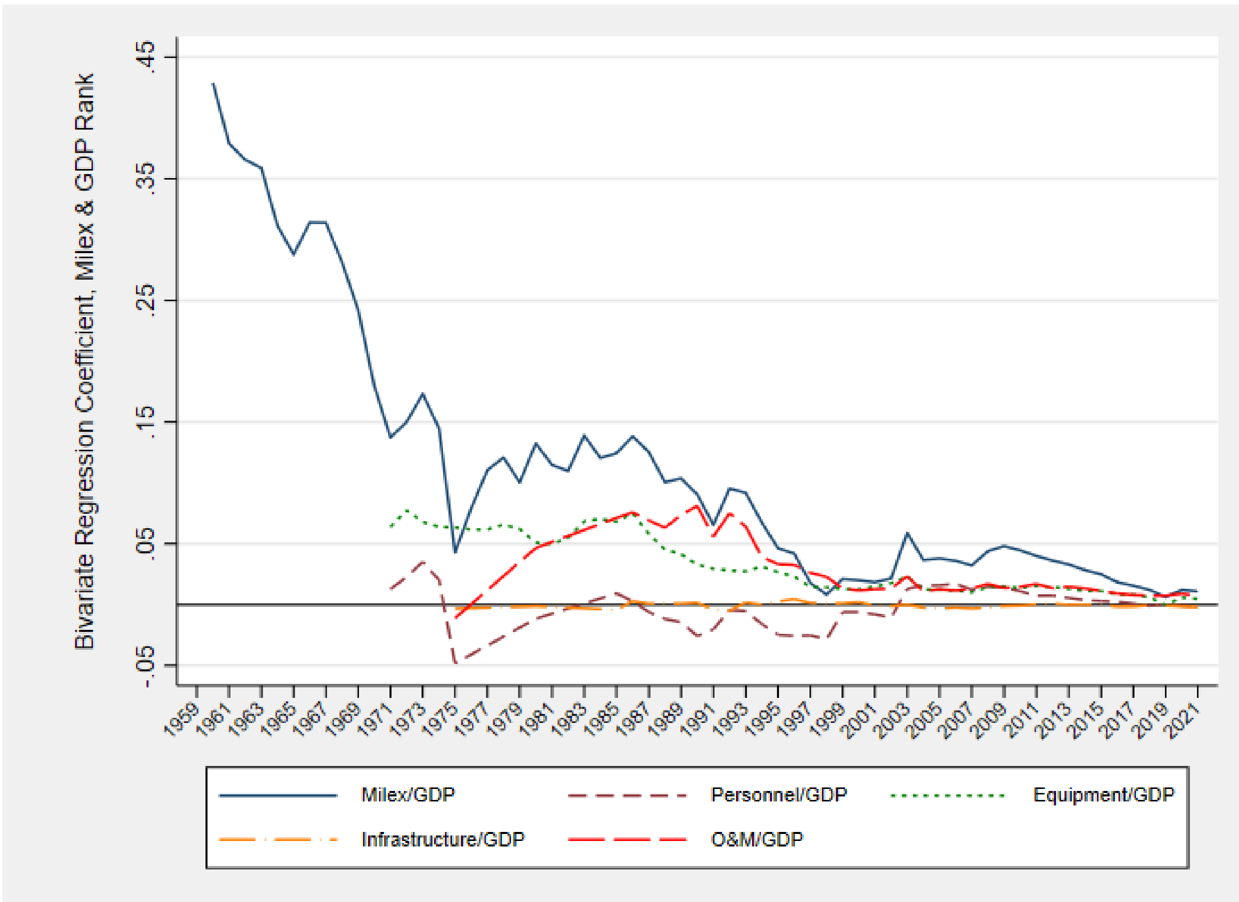
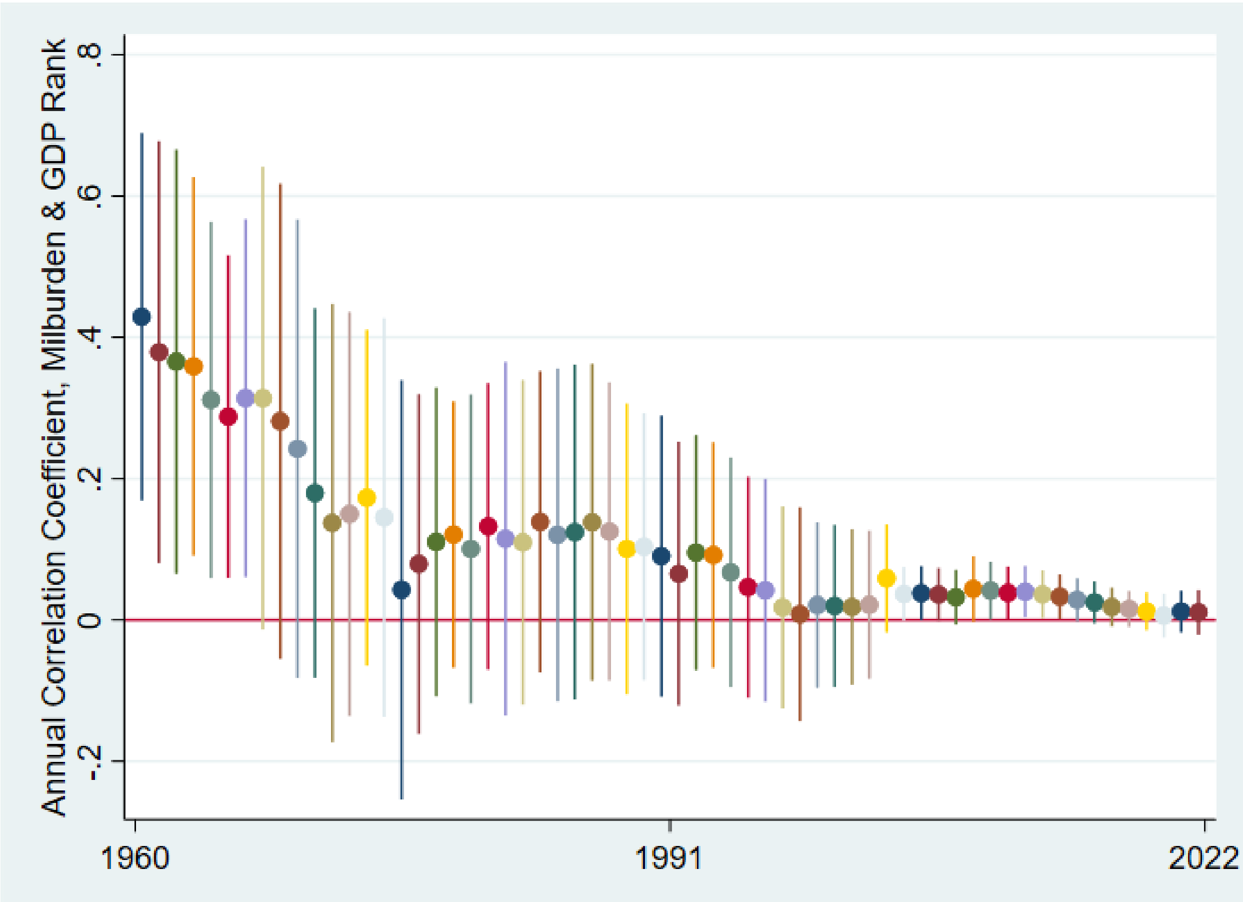


Figure 3 shows that even when we take the disaggregated components of milex as a share of GDP, their rank correlation with GDP differs significantly over time. Since about 2000, O&M and equipment have begun tracking much more closely with overall defense spending – this suggests that topline defense spending has actually become a *better* metric for burden-sharing, if we adhere to NATO Secretary General Stoltenberg's "3C's" approach emphasizing cash (overall spending), capabilities (equipment investment), and contributions to operations (O&M spending). At the same time, poorer countries have begun to spend relatively more on personnel and infrastructure, both of which have fluctuated on either side of zero correlation over time.

Figure 3 does not, however, indicate whether correlations are statistically significant during each year. Figure 4 plots 95% confidence intervals around the correlation coefficient between overall defense spending and GDP by year. Two observations are apparent. First, the figure confirms the observation (Sandler & Hartley, 1999) that the relationship uncovered by Olson and Zeckhauser failed to hold soon after their work was published. However, for a period in the mid-2000s, the relationship appeared again – as the United States focused on the "Global War on Terror" the early Cold War dynamic of increased publicness of defense spending arose again. Second, confidence intervals shrink significantly over time – this is due to the increasing number of observations as data became available for more countries as they joined the EU, NATO, or both.

Figure 4: Correlation coefficients by year, Defense Spending/GDP & GDP Rank (95% confidence intervals)



However, it is still not clear which *type* of defense spending drove these correlations over time. There are good theoretical reasons to imagine that wealth may shape the *kinds* of defense investment countries make – for example, richer countries may elect to invest in more capital-intensive militaries, while seeking to project power and influence abroad using military power as a tool of influence.

We therefore begin by visualizing the correlation between GDP rank and disaggregated defense expenditures, starting with equipment investment in Figure 5 below.

Figure 5: Correlation coefficients by year, Equipment Spending/GDP & GDP Rank (95% confidence intervals)

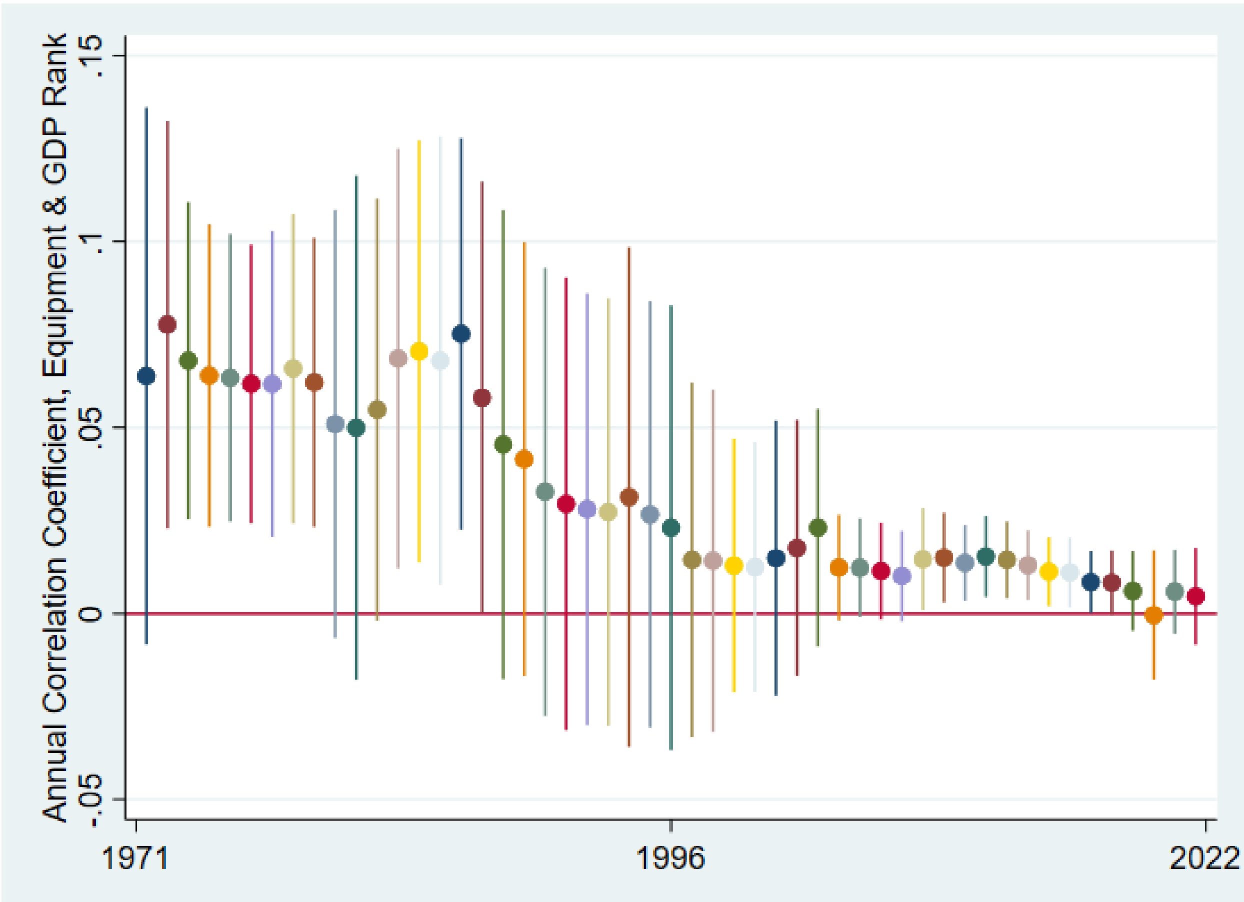


Figure 5 shows that even after the positive correlation between overall defense spending and GDP disappeared in the late 1960s, richer countries continued to spend significantly more on equipment through most of the 1970s. Only in the mid-1980s did a durably insignificant relationship between equipment spending and GDP rank establish itself, and, slightly differently from top line spending, there was a brief period in the 2010s during which there was a significant and positive correlation between GDP rank and equipment spending. These are meaningful differences both across time and spending categories. Because NATO and the EU both explicitly target equipment spending, the gradual decline in correlations between GDP rank and equipment spending since the mid-2010s indicates an improvement in that dimension of burden-sharing – this would not be apparent in topline data alone.

Figure 6 replicates the exercise for O&M spending. Like Figure 5, it shows that positive cross-country correlations occurred at different times than for topline defense spending, encompassing only the early 1990s and the height of the Iraq and Afghanistan wars. That the correlation is statistically insignificant through the 1990s and early 2000s suggests that the burden of deploying to and participating in operations in the Balkans was borne relatively equally between rich and poor allies.

Figure 6: Correlation coefficients by year, O&M Spending/GDP & GDP Rank (95% confidence intervals)

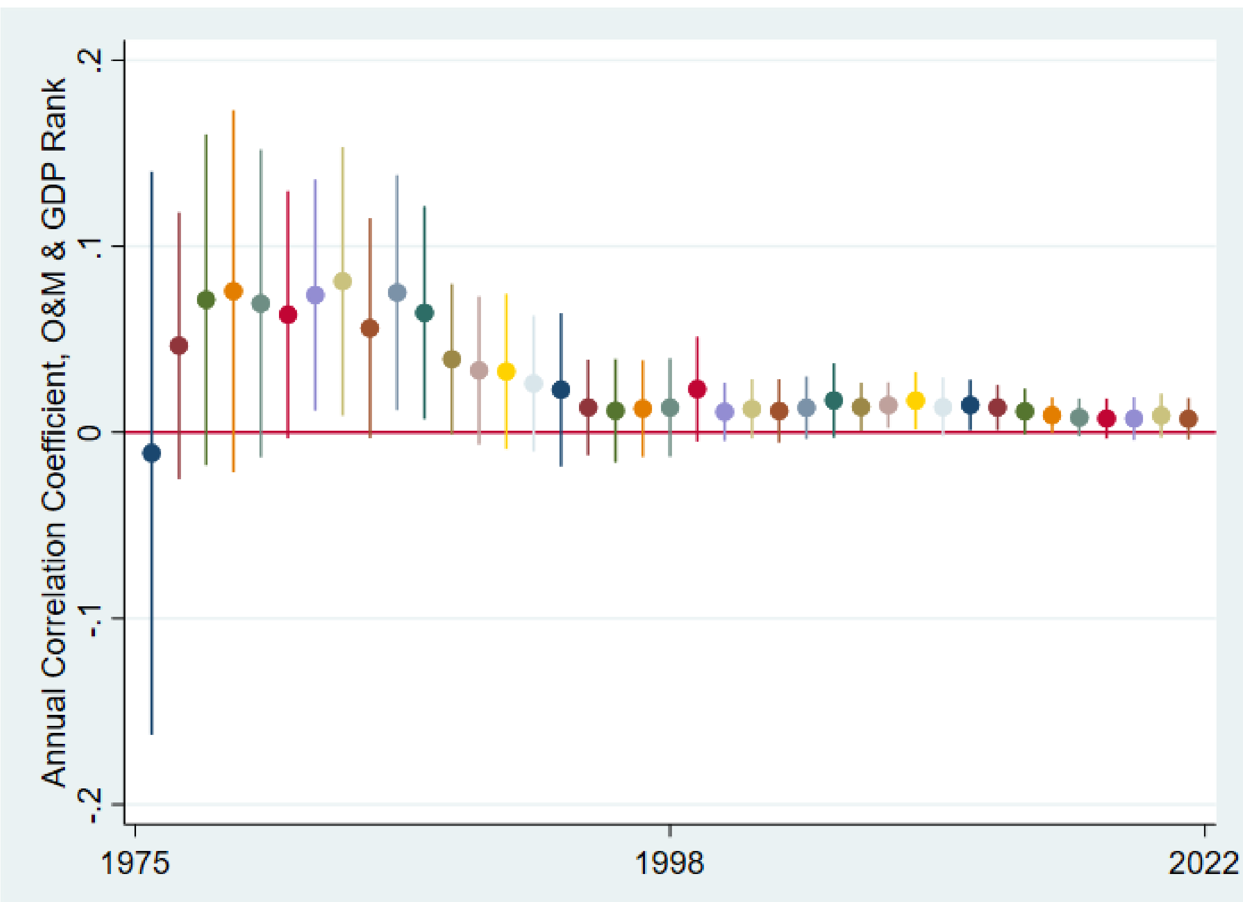
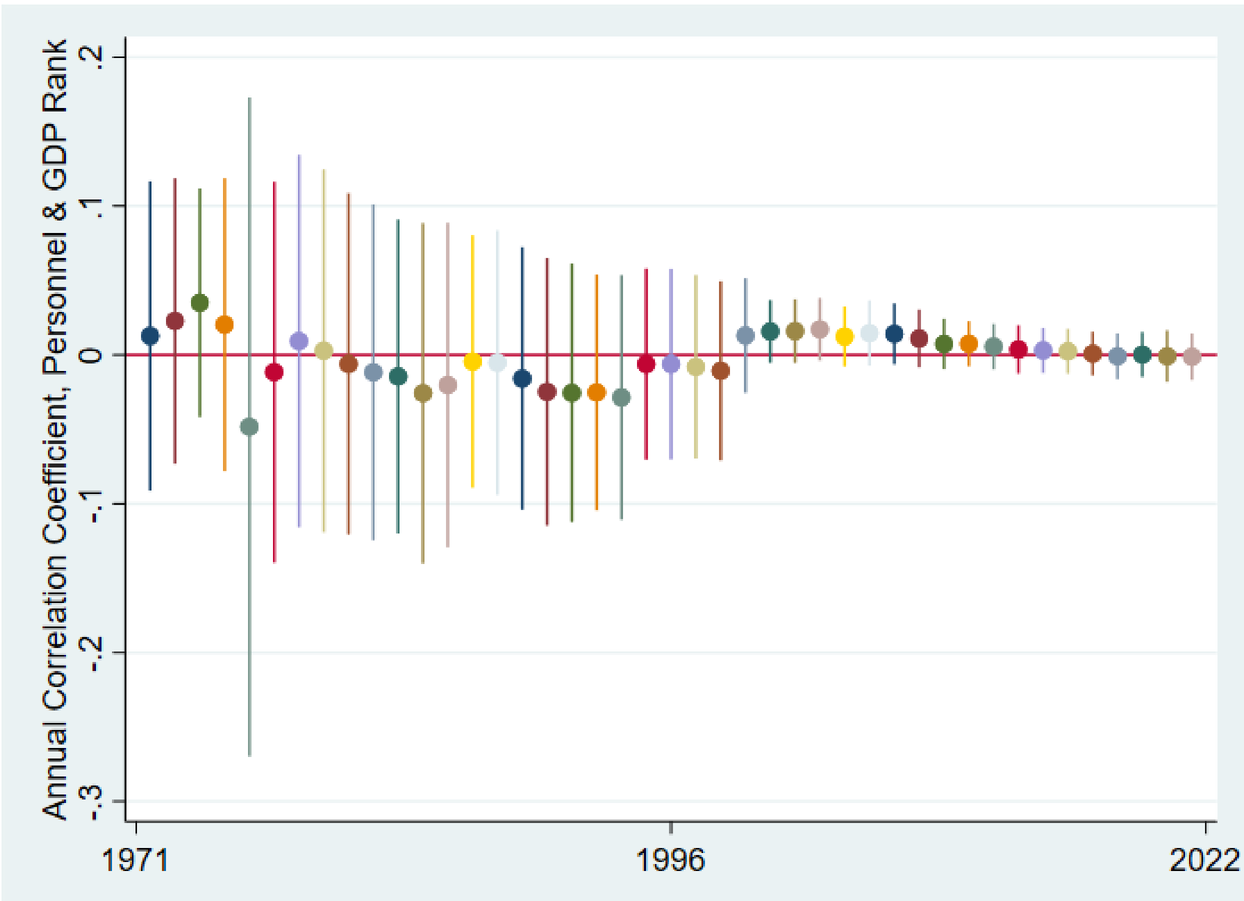


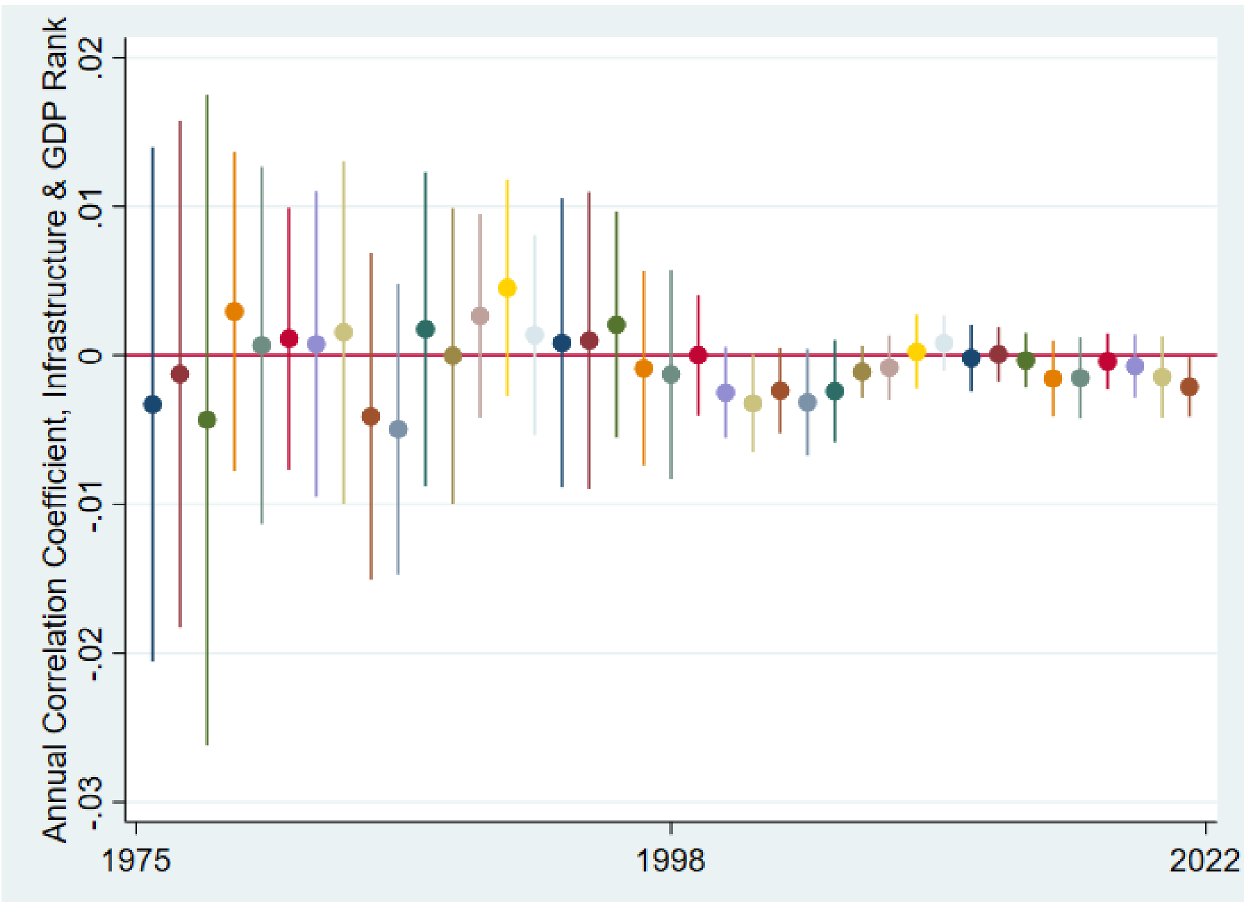
Figure 7 demonstrates that there has never been any statistically significant correlation between personnel spending and GDP. That richer countries, whose militaries are more capital intensive, do not spend relatively *less* on this category is perhaps theoretically surprising and merits investigation.

Figure 7: Correlation coefficients by year, Personnel Spending/GDP & GDP Rank (95% confidence intervals)



Finally, Figure 8, below, shows that while from 1975 to 2019 there was no statistically significant correlation between infrastructure spending and GDP, that changed in 2020, with poorer countries tending to spend more on infrastructure. This could suggest that poorer countries adjacent to Russia are investing in infrastructure to house military units from larger allies, particularly if it persists into the future.

Figure 8: Correlation coefficients by year, Infrastructure Spending/GDP & GDP Rank (95% confidence intervals)



Taken together, figures 4-8 demonstrate significant variation in the annual cross-sectional correlations between GDP rank and defense spending, both over time and across categories of defense spending. Although personnel represents on average the highest share of countries' defense spending (55%), it is equipment (16% of defense budgets on average) and O&M (24% of defense budgets on average) spending that seem to drive these differences. Richer countries tend to purchase more equipment, and to train and deploy more with that equipment than do poorer countries. Infrastructure, accounting for only 3.6% of defense budgets on average, may take on particular importance in areas in which the basing of allied forces is strategically salient, as is the case in the Baltic states and the Black Sea littoral states in recent years.

The Joint Product Model

In response to the apparent changes over time in the relationship between GDP and topline defense spending, scholars developed the Joint Product Model, which "generalizes the pure public good model because it encompasses the latter as a special instance when only a single pure public good output is derived from the defense activity. If defense provision gives rise to multiple outputs, then joint products exist…. Benefits derived from defense activities are imperfectly public among allies when these benefits are either partially excludable by the providing ally or else partially rival among the allies (Sandler & Hartley, 1999)."

We replicate canonical tests of the pure public good model alongside the Joint Product models in line with Sandler and Hartley. We begin with Table 3 below, which replicates (in Panel A) Sandler and Hartley's rendition of the pure public good model using the full sample of data available for 28 NATO allies, from the first year of availability through 2020, which is the last year for which all covariates are available. Panel B replicates the analysis from 2004 (the year of NATO's fifth – and largest – round of enlargement). Income (GDP) is positively correlated with all categories of defense spending across all models, providing support for the pure public good model.:

$$DEF = f(PRICE, INCOME, SPILLINS, THREAT, STRATEGIC) \tag{2}$$

Table 3: Pure Public Good Model, full sample and post-2004

Pure Public Good Model (Sandler & Hartley 1999, Time Periods)

	Panel A: 1975-2020					**Panel B: 2004-2020**				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	Milex/GDP	Personnel/GDP	Equipment/GDP	O&M/GDP	Infrastructure/GDP	Milex/GDP	Personnel/GDP	Equipment/GDP	O&M/GDP	Infrastructure/GDP
PRICE	-198,993.79	-2,494.89	-1,462.60	-276.34	-77.28*	-10557417.71*	-2,584.94	-3,948.71**	-2,708.93	-116.47***
	(6642654.702)	(2,285.849)	(1,477.109)	(2,256.126)	(43.502)	(5664599.555)	(2,480.096)	(1,817.048)	(2,645.443)	(44.923)
INCOME	0.03***	0.01***	0.01***	0.01***	0.00***	0.03***	0.01***	0.01***	0.01***	0.00***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.001)	(0.000)	(0.001)	(0.000)
SPILLINS	-0.05*	-0.03**	-0.01	-0.02	-0.00*	-0.15**	-0.05***	-0.03***	-0.05*	-0.00
	(0.030)	(0.012)	(0.005)	(0.013)	(0.001)	(0.063)	(0.018)	(0.012)	(0.030)	(0.001)
THREAT	-28.28	-13.95	-12.71	1.55	-1.33	-2.85	9.45	-0.32	9.73	0.49
	(52.159)	(18.466)	(11.440)	(19.706)	(1.706)	(87.427)	(29.192)	(22.077)	(18.672)	(0.775)
STRATEGIC	-6,916.96	-2,738.02	-904.87	-3,092.97	-74.10	-17,048.65*	-6,155.11**	-3,267.93**	-6,586.94	-291.32
	(7,123.886)	(2,020.193)	(1,275.635)	(3,409.482)	(156.099)	(8,845.758)	(2,451.782)	(1,538.829)	(4,341.019)	(191.869)
Constant	50,837.82	25,993.92*	8,259.48	17,608.65	1,825.76*	149,654.01**	47,807.66**	28,227.84**	49,169.45	1,900.78
	(32,083.896)	(13,319.348)	(6,331.745)	(14,906.489)	(1,022.137)	(65,978.449)	(18,985.693)	(11,979.621)	(31,175.485)	(1,280.124)
Observations	744	744	744	744	744	452	452	452	452	452
Number of ccode	28	28	28	28	28	28	28	28	28	28

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Price, on the other hand (the inverse of the correlation between milex and GDP for the entire time series for country *i*), is negatively correlated with equipment and infrastructure spending only after 2004 – countries appear to seek savings in those areas. Regarding SPILLINS (ally *i*'s defense spending subtracted from the sum of all allied defense spending in year *t*), there is considerable heterogeneity across both time and defense spending categories. Over the full time series, only personnel spending is associated (negatively) with SPILLINS. On the other hand, in the post-2004 period, overall, equipment, and personnel spending are negatively associated with SPILLINS, suggesting that easy-riding (Cornes & Sandler, 1984) may be more prevalent during that period.

The differences during the two time periods covered in Table 4 suggest that there are factors that differ over time but are similar across countries – it is reasonable to suspect the presence of time-specific shocks. This situation points toward modeling year fixed effects. Table 4 captures the results of a replication of the models used in Table 3 with year fixed effects included. The results are striking – the relationship between income and milex becomes very heterogeneous across categories, with richer allies appearing to spend more on personnel and less on O&M, and with a negative coefficient on SPILLINS across all models pointing toward systematic easy riding.

Table 4: Pure Public Good Model, Year Fixed Effects

Pure Public Good Model (Sandler & Hartley 1999), Year Fixed Effects					
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Milex/GDP	Personnel/GDP	Equipment/GDP	O&M/GDP	Infrastructure/GDP
PRICE	-118,484.08	-384.93	-868.63*	785.68	82.24
	(117,089.946)	(1,020.483)	(512.086)	(1,030.623)	(94.648)
INCOME	0.00	0.00**	0.00	-0.00***	-0.00
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
SPILLINS	-0.99***	-0.34***	-0.26***	-0.39***	-0.02***
	(0.008)	(0.019)	(0.020)	(0.009)	(0.002)
THREAT	1.42	-24.76	-7.19	17.71	-2.09
	(2.902)	(25.267)	(14.812)	(19.500)	(3.606)
STRATEGIC	9,662.45***	7,432.59***	6,680.35***	8,455.53***	28.24
	(132.610)	(1,391.717)	(721.469)	(1,789.706)	(436.472)
Constant	993,763.94***	342,903.07***	260,681.85***	387,326.08***	17,087.59***
	(8,136.414)	(18,891.682)	(19,968.331)	(8,615.412)	(2,114.855)
Observations	744	744	744	744	744
Number of ccode	28	28	28	28	28

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In short, disaggregation changes the results emerging from traditional public good analyses, both as they were initially conceived by Olson and Zeckhauser and as they were refined by later scholars. In policy terms, the heterogeneity observed in Table 3 suggests that allies may selectively free (?)ride – skimping, for example, on long-term investments in equipment, while seeking to maintain a high operational tempo in order to support major allies' out of area operational priorities – a form of "paying for protection (Ringsmose, 2009)." The results in table 4, on the other hand, suggest that richer countries tend may be fielding "banzai armies (Missiroli, 2013)" with highly paid personnel who train and deploy less than those in their poorer counterparts. These analyses are merely suggestive - only by disaggregating defense spending data into its constituent categories can scholars really explore the causes and consequences of these results.

Modern scholars, however, tend to gravitate toward the Joint Product model more than the pure public good model, for both theoretical (defense provision conceptually conforms to theoretical definitions of jointness) and empirical (empirical tests affirm this) reasons. We therefore turn to replications of the two Joint Product models outlined by Sandler and Hartley (Sandler & Hartley, 1999).

Their first joint product model (Joint Product 1), which analyzes defense spending as a function of PRICE (the opportunity cost of defense spending), FULL (the defense expenditures of all allies plus the GDP of country *i*, SPILLINS (defined above), THREAT, and STRATEGIC (a manually-coded measure of the excludability of a given strategic approach pursued by allies in year *t*), devised by Sandler and Hartley (1999): (3)

$$ALLDEF = f(PRICE, FULL, SPILLINS, THREAT, STRATEGIC)$$

Table 4: Joint Product 1 Replication

Joint Product 1 (Full vs. Spillins), Sandler & Harltley 1999					
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Milex/GDP	Personnel/GDP	Equipment/GDP	O&M/GDP	Infrastructure/GDP
PRICE	-9.08e+11	-3.91e+09	-4.61e+09**	-1.26e+09	414853180.46***
	(6.224e+12)	(3.338e+09)	(1.950e+09)	(2.001e+09)	(1.206e+08)
FULL	0.03***	0.02***	0.01***	0.01***	0.00***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)
SPILLINS	923,975.78***	404,986.51***	304,667.96***	350,113.35***	8,915.53***
	(28,126.857)	(18,234.111)	(9,053.793)	(20,532.693)	(949.267)
THREAT	-29224116.26	226176697.20***	34478581.33	198824649.84***	4,875,172.16*
	(49970435.443)	(36022630.239)	(21909658.168)	(33737436.473)	(2668560.039)
STRATEGIC	-6.51e+09	-5.60e+09**	1.26e+10***	-1.81e+10***	8,240,610.13
	(6.747e+09)	(2.840e+09)	(2.325e+09)	(3.011e+09)	(2.153e+08)
Constant	4.83e+10	-1.17e+10	-8.28e+10***	-4.32e+10**	9.48e+09***
	(3.020e+10)	(1.910e+10)	(9.900e+09)	(2.056e+10)	(9.514e+08)
Observations	745	745	744	745	745
Number of ccode	28	28	28	28	28

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4 replicates Sandler and Hartley's analysis using the disaggregated dataset. There is notable heterogeneity in the results across spending categories. Allies for whom defense has high opportunity costs appear to substitute infrastructure for other types of defense spending. Allies appear to respond to increasing threat from Russia with personnel and O&M spending rather than equipment modernization (again consistent with Ringsmose's "paying for protection" hypothesis). Strategic excludability appears to dampen those same expenditures while encouraging equipment spending – when allies are less able to free- or easy-ride, they spend more on equipment modernization, capabilities, and capacity.

The spatial component of spillover effects

George and Sandler (2018) used "spatial autoregressive (SAR) weighting of alliance spillovers of defense based on ally contiguity (augmented by US defense spending) and inverse distance between allies." Their results were notably consistent with those arising from the use of the SPILLINS variable above. In particular, they found evidence of free riding throughout their models. How do these findings respond to disaggregation?

We began by replicating Table 2 in George and Sandler (2018), in Table 5 in this paper. Columns 3 and 6 in that table demonstrate heterogeneity in O&M and Infrastructure spending. The relationships George and Sandler observed appear to be driven, then, by equipment and personnel spending – it is in these areas that allies appear to do most of their free-riding. George and Sandler depict this with the following equation:

$$\ln(ME)_{it} = \rho \sum_{k \neq i} \ln(ME)_{kt} + \beta_1 \ln(GDP)_{it} + \beta_2 \ln(Pop)_{it} + \beta_3 Attacks_{it} + \beta_4 \ln(Russia\ ME)_t + \beta_5 MAD^* \sum_{k \neq i} \ln(ME)_{kt} + \mu_i + \varepsilon_{it}, \quad (4)$$

Table 5: Replication of Table 2 in George and Sandler (2018)

George & Sandler Table 2: NATO military expenditure, 1975–2015.

VARIABLES	(1) G&S ME	(2) G&S ME	(3) O&M	(4) G&S ME	(5) G&S ME	(6) Infrastructure
Spatial lag of ME (Alliance Membership)	-0.000528***	-0.000529***	-8.07e-05			
	(0.000160)	(0.000170)	(0.000255)			
Spatial lag of ME (contiguity + US)				-0.00234***	-0.00249***	-0.00190
				(0.000501)	(0.000504)	(0.00175)
Ln GDP	0.0524*	0.0603*	-0.0600	0.0368	0.0525*	-0.333***
	(0.0289)	(0.0310)	(0.0443)	(0.0265)	(0.0279)	(0.0935)
Ln Population	1.177***	1.163***	1.507***	1.045***	1.007***	1.873***
	(0.171)	(0.173)	(0.303)	(0.175)	(0.177)	(0.582)
Observations	771	734	734	771	741	741
R-squared	0.277	0.266	0.146	0.254	0.261	0.093
Number of ccode	27	27	27	27	27	27
Country FE	YES	YES	YES	YES	YES	YES
F test:	39.37	36.95	9.152	42.11	45.29	9.214
Kleibergen-Paap LM Test (Prob > chi-sq):	0	0	0	1.34e-10	1.45e-10	1.45e-10
Adjusted R-squared:	0.248	0.235	0.110	0.223	0.230	0.0544

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Why might this be the case? We again find suggestive evidence supporting Ringsmose's "paying for protection" hypothesis: O&M spending, particularly in the absence of underlying capabilities investments, suggests that allies supporting out of area activities in a period that the US emphasized that, while neglecting long term capital improvements. Similarly, infrastructure spending can be a way of inviting larger allies to provide such protection "onsite" by building suitable cantonment, storage, and training areas. Thus, the lack of free-riding in these two categories of expenditure aligns with, in particular, the intra-alliance strategic dynamic in the 2010s, wherein countries perceiving themselves as vulnerable to the threat from Russia deployed disproportionately in support of US priorities in Iraq and Afghanistan, and at the same time invested heavily in infrastructure to support allied "boots on the ground" on their territory.

Table 6, below, replicates George and Sandler's (2018) Table 4, again showing heterogeneity in results surrounding O&M and infrastructure spending – while allies continued to free- or easy-ride in overall military expenditures, they appeared to at least attempt to "punch above their weight (Wivel & Crandall, 2019)" in line with the hypotheses outlined above.

Table 6: Replication of Table 4 in George and Sandler (2018)

Table 4: NATO military expenditure, 1991–2015 and 1999–2015 (alliance membership weights).

	1991-2015				1999-2015		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	G&S ME	O&M	G&S ME	G&S ME	O&M	G&S ME	Infrastructure
Spatial lag of ME	-0.00032**	-0.00010	-0.00053**	-0.00053**	-0.00040	-0.00100***	-0.00048
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Attacks	0.00086***	0.00116**	0.00077	0.00079	-0.00432	0.02842***	0.00059
	(0.000)	(0.001)	(0.003)	(0.003)	(0.006)	(0.008)	(0.021)
Ln GDP	0.05366	-0.03362	0.08022*	0.08038*	-0.01170	0.83814***	0.95631***
	(0.039)	(0.059)	(0.045)	(0.045)	(0.065)	(0.019)	(0.057)
Ln Population	1.07885***	1.92920***	0.84321***	0.85058***	1.50140***	0.29659***	0.06832
	(0.175)	(0.316)	(0.244)	(0.245)	(0.348)	(0.023)	(0.065)
Ln Russia ME						-0.44234***	-1.05984***
						(0.077)	(0.223)
Constant						-6.90276***	4.96323
						(1.728)	(5.079)
Observations	515	515	404	403	403	404	404
R-squared	0.111	0.134	0.046	0.047	0.049	0.972	0.792
Number of ccode	27	27	27	27	27		
Country FE	YES	YES	YES	YES	YES	NO	NO
F test:	23.49	15.66	3.950	3.946	5.001	2013	268.6
Kleibergen-Paap LM Test (Prob > chi-sq):	0	0	0	0	0	0	0
Adjusted R-squared:	0.0537	0.0782	-0.0329	-0.0329	-0.0304	0.972	0.790

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Within budget disaggregation

While the foregoing has demonstrated the utility of disaggregating defense expenditures to evaluate compositional differences in both demand for and effects of defense spending as they pertain to overall national budgets, there are reasons to evaluate demand functions in terms of compositional variation *within* defense budgets.

For example, "The percentage of defense budgets allocated to O&M expenditures correlates significantly and positively with several key indicators of allies' willingness to conduct operations outside their own territory, such as the ratio of land forces deployed to national wealth, the ratio of soldiers deployed in Afghanistan to total population, and the ratio of deployed personnel to total personnel (Becker & Malesky, 2017).” International relations and security studies scholars working on demand for defense spending may consider using data disaggregated at the level of defense budgets themselves to capture countries’ emphasis on particular aspects of their own defense enterprises.

Conclusions and Future Research

Our combined dataset including all available years of disaggregated defense spending for NATO and EU states is extensive and systematic in terms of years and country coverage. We have shown that substituting disaggregated data for topline spending data is associated with striking differences in results of canonical and newer research in the broad field of public choice economics, and in the specific subfields of defense economics, political economy of security and transatlantic studies. We have offered examples that can be further developed in ways that were not possible with the most widely used data on defense spending (SIPRI, 2018; World Bank, 2021). Much remains to be done – the analyses here as well as the works we have referenced that have already made use of disaggregated data remain limited. The overview above suggests that disaggregation is a very promising avenue toward better understanding of new and old research questions, as well as new ideas for research on national and collective economic, security and strategic resource allocation choices.

The dataset remains by its nature a work in progress, and as greater time series date is collected it will become even more useful. Extending the dataset to include non-Western states would be particularly valuable (Marksteiner, Tian & Kaplan, 2022). We will continue to update and extend both the disaggregated defense spending and the covariates analyzed above; perhaps more importantly, our replication files provide code enabling scholars to replicate the scraping of the data themselves in the future – the dataset can be updated as soon as the relevant international organizations publish new information. We expect that other researchers will benefit from the data and replication files associated with this article to extend upon the analysis above and the research we have referenced that has already made use of manually gathered disaggregated spending data.

References

Becker, Jordan (2017) The correlates of transatlantic burden sharing: revising the agenda for theoretical and policy analysis. *Defense & Security Analysis* 33(2): 131–157.

Becker, Jordan (2019a) Accidental rivals? EU fiscal rules, NATO, and transatlantic burden-sharing. *Journal of Peace Research* 56(5): 697–713.

Becker, Jordan (2019b) Arms Without Influence? Spatial Distribution of Defense Industrial Activity, Transatlantic Burden Sharing, and Strategy. (February) (https://papers.ssrn.com/abstract=3343493).

Becker, Jordan (2021) Rusty guns and buttery soldiers: unemployment and the domestic origins of defense spending. *European Political Science Review* 13(3): 307–330.

Becker, Jordan, Michael Duda & Douglas Lute (2022) From context to concept: history and strategic environment for NATO's 2022 strategic concept. *Defence Studies* 22(3): 489–496.

Becker, Jordan & Paul Dunne (2021) Military Spending Composition and Economic Growth. *Defence and Peace Economics* 0(0): 1–13.

Becker, Jordan & Edmund Malesky (2017) The Continent or the "Grand Large"? Strategic Culture and Operational Burden-Sharing in NATO. *International Studies Quarterly* 61(1): 163–180.

Bove, Vincenzo & Elisa Cavatorta (2012) From conscription to volunteers: budget shares in NATO defence spending. *Defence and Peace Economics* 23(3): 273–288.

Cooper, Scott & Kendall W Stiles (2021) Who commits the most to NATO? It depends on how we measure commitment. *Journal of Peace Research* 58(6): 1194–1206.

Cornes, Richard & Todd Sandler (1984) Easy Riders, Joint Production, and Public Goods. *The Economic Journal* 94(375): 580.

Danish Foreign Ministry (2015) Input-Output-Metrics-2014-2015.pdf (https://www.fmn.dk/temaer/nato/Documents/Input-Output-Metrics-2014-2015.PDF).

Dunne, Paul & Ron Smith (2020) Military Expenditure, Investment and Growth. *Defence and Peace Economics* 31(6): 601–614.

EDA (2018) Defence Data Portal (https://www.eda.europa.eu/info-hub/defence-data-portal).

European Council (2016) European Council Meeting (15 December 2016) - Conclusions (https://www.consilium.europa.eu/media/21929/15-euco-conclusions-final.pdf).

George, Justin & Todd Sandler (2018) Demand for military spending in NATO, 1968–2015: A spatial panel approach. *European Journal of Political Economy* 53(July): 222–236.

Hartley, Keith & Todd Sandler (1999) NATO Burden-Sharing: Past and Future. *Journal of Peace Research* 36(6): 665–680.

Kunertova, Dominika (2017) One measure cannot trump it all: lessons from NATO's early burden-sharing debates. *European Security* 26(4): 552–574.

Kuokštytė, Ringailė, Vytautas Kuokštis & Ieva Miklaševskaja (2021) External and domestic political determinants of defence spending: a time-series cross-section analysis of EU member states. *European Security* 30(2): 197–217.

Marksteiner, Alexandra, Nan Tian & Alexander Kaplan (2022) Improving transparency in military spending through disaggregated information | SIPRI (https://sipri.org/commentary/topical-backgrounder/2022/improving-transparency-military-spending-through-disaggregated-information).

McGerty, Fenella, Dominika Kunertova, Madison Sargeant & Andrew Webster (2022) NATO burden-sharing: past, present, future. *Defence Studies* 22(3): 533–540.

McInnis, Kathleen & Daniel Fata (2022) 2 Percent Defense Spending Is a Bad Target for NATO. *Foreign Policy* (https://foreignpolicy.com/2022/06/10/2-percent-defense-spending-nato-budget-bad-target/).

Missiroli, Antonio (2013) Enabling the future – European military capabilities 2013-2025: challenges and avenues. *European Union Institute for Security Studies* (https://www.iss.europa.eu/content/enabling-future-%E2%80%93-european-military-capabilities-2013-2025-challenges-and-avenues).

NATO (2014) Wales Summit Declaration issued by the Heads of State and Government participating in the meeting of the North Atlantic Council in Wales. *NATO* (http://www.nato.int/cps/en/natohq/official_texts_112964.htm).

NATO (2021) Information on defence expenditures. *NATO* (https://www.nato.int/cps/en/natohq/topics_49198.htm).

NATO (2022) Funding NATO. *NATO* (https://www.nato.int/cps/en/natohq/topics_67655.htm).

Olson, Mancur & Richard Zeckhauser (1966) An Economic Theory of Alliances. *The Review of Economics and Statistics* 48(3): 266–279.

Ringsmose, Jens (2009) Paying for protection: Denmark's military expenditure during the Cold War. *Cooperation and Conflict* 44(1): 73–97.

Sandler, Todd & Keith Hartley (1999) *The Political Economy of NATO: Past, Present and into the 21st Century*. Cambridge University Press.

SIPRI (2018) SIPRI Military Expenditure Database (https://www.sipri.org/databases/milex).

Stoltenberg, Jens (2019) Press conference by NATO Secretary General Jens Stoltenberg following the meetings of NATO Defence Ministers. *NATO* (http://www.nato.int/cps/en/natohq/opinions_163394.htm).

Wivel, Anders & Matthew Crandall (2019) Punching above their weight, but why? Explaining Denmark and Estonia in the transatlantic relationship. *Journal of Transatlantic Studies* 17(3): 392–419.

World Bank (2021) World Development Indicators (WDI) Data Catalog (https://datacatalog.worldbank.org/dataset/world-development-indicators).