INSTITUTE
OF ECONOMIC STUDIES
Faculty of Social Sciences
Charles University

MILITARY EXPENDITURE AND ECONOMIC GROWTH: A META-ANALYSIS

Josef Simpartl

IES Working Paper 8/2024

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Bibliographic information:
Simpartl J. (2024): " Military Expenditure and Economic Growth: A Meta-Analysis" IES Working Papers 8/2024. IES FSV. Charles University.

This paper can be downloaded at: http://ies.fsv.cuni.cz

Military Expenditure and Economic Growth: A Meta-Analysis

Josef Simpartl[1]

[1]Institute of Economic Studies, Charles University, Prague, Czech Republic
62565549@fsv.cuni.cz

February 2024

Abstract:
In the face of wars and a geopolitically challenging environment, military expenditures have once again become political focal points in developed countries. However, the scientific literature remains inconclusive regarding their impact on economic growth. This paper conducts a meticulous meta-analysis, examining 405 estimates from 67 studies and incorporating over 30 variables to account for variations in their characteristics. The meta-analysis reveals a consistently negative average effect of military expenditures on economic growth, coupled with an absence or mild presence of publication bias. Both Bayesian and Frequentist model averaging highlight the diversity among individual estimates, attributing this variation to the data characteristics of individual studies. Notably, factors such as the panel structure, number of observations, number of countries, and time span emerge as crucial contributors to this diversity. The pivotal influence of data originating from the 1990s suggests the significance of de-escalation periods and hints at potential non-linearities within the observed effects. This paper makes notable contributions to prior meta-analyses by adopting an updated dataset, a more robust approach to publication bias analysis, and providing a more refined solution to addressing model uncertainty in the heterogeneity analysis.

JEL: C83, O11, H50
Keywords: meta-analysis, publication bias, model averaging, military expenditure, economic growth

Data and code are available in an online appendix at https://shorturl.at/cDERY

1 Introduction

Over the past five decades, there has been extensive scrutiny of the economic implications of military expenditures, seeking to discern whether such expenditures deplete resources from the economy or yield economic benefits. Alptekin and Levine (2012) undertook a comprehensive review of the resulting empirical literature examining the correlation between military expenditures and economic growth; however, viewed through a contemporary lens, their study reveals several shortcomings. Since its publication, the scientific discourse on this subject has experienced substantial growth, incorporating not only more expansive datasets in terms of both the number of countries and longer time series but also a broader array of control and interaction variables that might influence the relationship between military expenditures and economic growth. Furthermore, the published papers now employ a more diverse range of theoretical models and estimation techniques.

In addition to an updated dataset, the methodology of this meta-analysis constitutes a significant contribution to the scientific literature. It utilizes various methods for publication bias and heterogeneity analysis, reflecting advancements in this field. In addressing publication bias, this paper employs the FAT-PET-PEESE test with various linear specifications (weighted least squares, fixed-effects, and between-effects regressions) and non-linear methods, including the Stem-based test by Furukawa (2019), the Weighted Average of Adequately Powered (Ioannidis et al., 2017), the Selection model (Andrews and Kasy, 2019), and the Endogenous Kink by Bom and Rachinger (2019). Furthermore, the paper incorporates methods to control for endogeneity, such as FAT-PET-PEESE with an instrumental variable, p-uniform* by van Aert and van Assen (2018), and the Caliper test by Gerber and Malhotra (2008).

The analysis reveals that the average effect of military expenditures on economic growth is negative, with estimated effects ranging from -0.107 to -0.052 when controlling for publication bias. These estimates are highly statistically significant, except for the Stem-based test. The only economically significant estimate is derived from the study-weighted least squares, surpassing the threshold of -0.104 specified by Doucouliagos (2011). Publication bias is either absent or mild across all methods.

The heterogeneity analysis employs Bayesian model averaging following Zeugner and Feldkircher (2015), Frequentist model averaging with the Mallows criterion (Hansen, 2007), and a hybrid approach involving ordinary least squares using only significant variables from the Bayesian model (Gechert et al., 2022). The analysis indicates that data characteristics of the primary studies predominantly impact the economic effect of military expenditures. The panel structure of data and the number of observations in the primary-study datasets have a negative effect, while the number of included countries and the time span have a positive effect. Negative effects are also observed in studies analyzing data from the 1990s. The last discovery underscores the significant importance of the decade, widely regarded as a crucial stride toward international de-escalation post-Cold War. This revelation suggests a consequential implication: embracing diplomacy proves economically advantageous, enabling states to reduce military expenditures and alleviate the associated economic burden.

The paper is organized as follows: Section 2 provides a brief overview of the channels through which military expenditures may affect economic growth. Section 3 details the data collection procedure, data transformation, and independent variables and provides a brief descriptive analysis of the data. Section 4 introduces the methodology employed. Section 5 presents the results of the publication bias and heterogeneity analyses. Finally, Section 6 concludes the paper with a summary, acknowledgment of shortcomings, and implications.

2 Literature Review

The literature on peace economics offers varying evidence and conclusions regarding the impact of military spending. At its most basic level, military spending contributes to a secure environment, a factor that, while not strictly economic, carries significant economic externalities. A secure environment fosters economic growth by instilling confidence in property rights adherence, thereby incentivizing economic agents to invest and accumulate capital (Nugroho and Purwanti, 2021). Lipow and Antinori (1995) even argue that this effect intensifies in the presence of a security threat.

Military expenditures play a role in boosting aggregate demand, making use of previously unused production capacities (Khidmat et al., 2018). Nations with substantial military forces can stimulate aggregate demand through a positive employment effect (Heo, 2010). In arms-producing countries, increased military spending attracts high-skilled workers, elevates their wages, and enhances the profits of arms-producing companies (Rahman and Siddiqui, 2019).

The positive impact of military spending also extends to technology, infrastructure, and human capital development, all of which can be harnessed in the civilian economy (Chan, 1988). The military not only provides a disciplined and habituated labor force but also imparts technical and managerial skills (Faini et al., 1984). However, Dunne and Watson (2005) suggest a "spin-in" effect, contending that the military sector benefits from the development of civilian technologies, rather than the other way around. This may hint at a conjecture that the technology spillover effect found in the literature may be overestimated.

Conversely, the scientific literature outlines negative effects resulting from the impact of military expenditure on productivity, investment, savings, and fiscal conditions. The productivity argument suggests a trade-off between military and civilian economic sectors, with resource shifts towards the military sector potentially leading to an overall decrease in productivity (Ward and Davis, 1992). DeGrasse (1983) and Mintz (1989) observe a similar pattern concerning employment.

Military expenditures can also adversely affect economic growth through their impact on investments, particularly in social welfare, such as education, health care, and infrastructure (Ram, 2019). Faini et al. (1984) and Landau (1993) note a negative impact on private investments, specifically in the relationship between the military burden (military expenditures relative to GDP) and the investment ratio. Deger and Smith (1983) and Deger (1986) find a corresponding effect on the savings ratio. Adverse fiscal effects from heavier tax burdens or public budget deficits, are also attributed to military expenditure (Chan, 1988).

The delineation between positive and negative effects remains nuanced, suggesting a non-linear relationship between military expenditures and economic growth. Some argue that the actual amount of resources devoted to the military determines whether military expenditures promote or hinder economic growth, proposing a quadratic formalization of this effect. Landau (1996) first identifies this quadratic effect in OECD countries, a result echoed by Stroup and Heckelman (2001) in a dataset spanning 44 African and Latin American countries. Other studies suggest that non-linearities arise from interactions with other factors; for instance, Compton and Paterson (2016) propose that sound legal and political environments may mitigate the negative effects of defense expenditures.

While Alptekin and Levine (2012) initially review empirical papers, asserting a significantly positive economic impact of military expenditures, the wealth of contemporary evidence and refined meta-analysis methodologies necessitate a reconsideration of aggregating evidence on the economic effects of military expenditure.

3 Data

To create a dataset, I collect data from papers used in Alptekin and Levine (2012) and conduct a Google Scholar search, using search queries "*military/defense spending/expenditure*" or "*military burden*" and "*economic growth*". The search yields 600 papers, out of which 161 are duplicates and 48 papers are not available in the form of full-text or reported estimates.

After the initial search, 423 studies are available. Since this paper aims to extend the previous meta-analysis with more recent literature, I exclude 139 papers published before 2000. 57 papers contain unrelated topics, such as the relationship between military expenditure and some other variables than growth. For the sake of quality, I do not account for 4 bachelor and master theses and doctoral dissertations.

Next criteria apply to full-texts. I exclude 143 papers using the Granger causality test and other atheoretical empirical models, such as vector autoregression, error correction models, and various cointegration techniques, because these papers are generally uninformative of size and direction of the underlying effect (Dunne and Smith, 2010). Moreover, I remove 21 papers that use the Feder-Ram model, because of its complicated nature (Dunne et al, 2005) and 24 studies that do not contain any econometric outputs. The selection procedure yields 67 studies with 405 estimates and is summarized in the PRISMA diagram depicted in Figure 1.

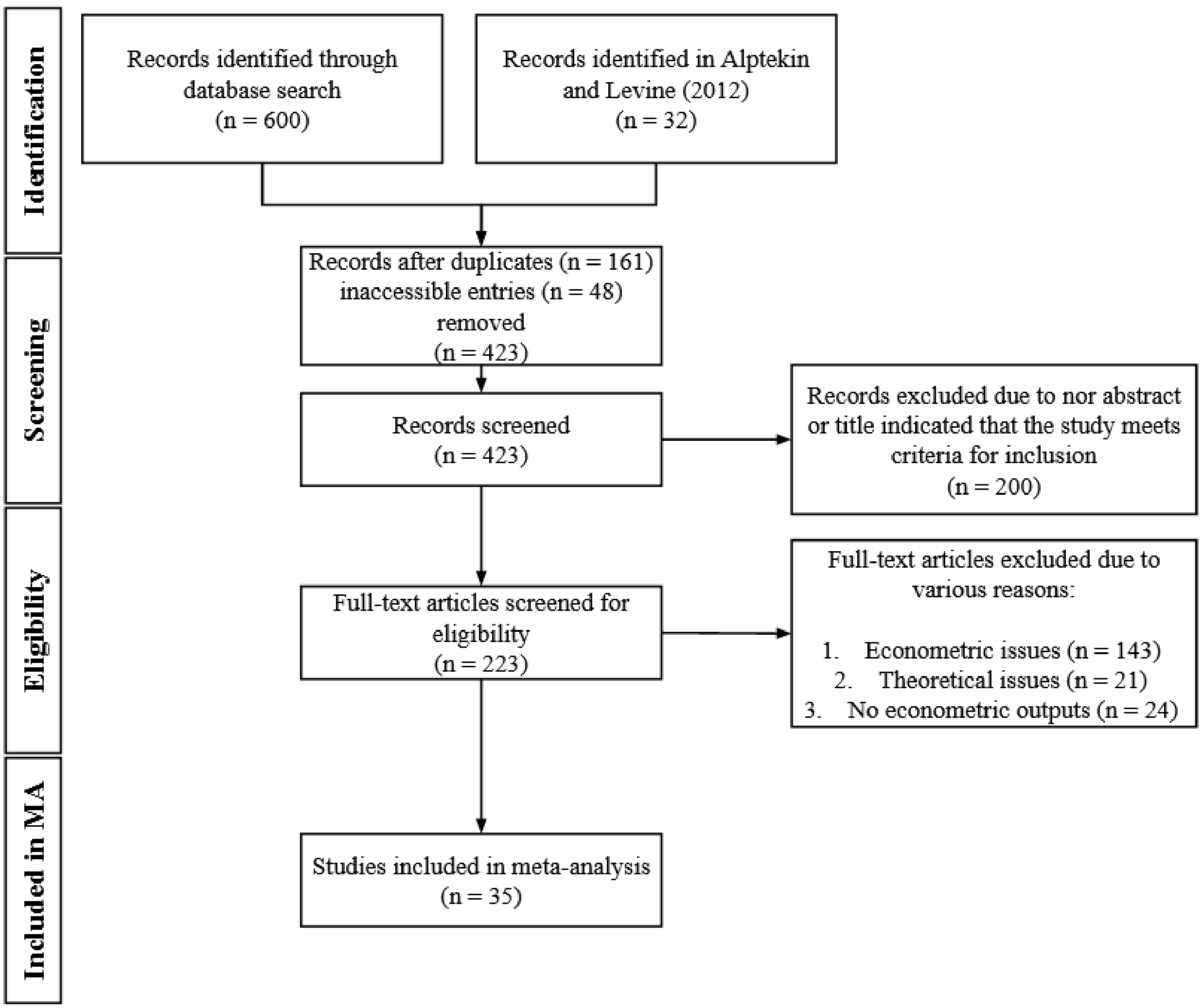

Figure 1: PRISMA diagram capturing the procedure of primary study selection

Since some of the estimates include squared and interaction terms, the estimates have to be transformed into marginal effects. The procedure follows Havranek and Irsova (2011), adjusting both estimated effects and standard errors. Equations (1) and (2) show the transformation of interaction effects and their standard errors respectively, while equations (3) and (4) show adjustment for squared terms:

$$ME_{milex} = ES_{milex} + ES_{IT} * \overline{VAR_{IT}} \tag{1}$$

$$SE_{ME_{milex}} = \sqrt{SE_{milex}^2 + SE_{IT}^2 * \overline{VAR_{IT}}^2} \tag{2}$$

$$ME_{milex} = ES_{milex} + 2 * ES_{milex^2} * \overline{milex} \tag{3}$$

$$SE_{ME_{milex}} = \sqrt{SE_{milex}^2 + 4 * SE_{milex^2}^2 * \overline{milex}^2}, \tag{4}$$

where ME denotes the marginal effect, ES denotes the original estimated linear effect, $\overline{VAR_{IT}}$ denotes the mean value of an interaction variable, SE denotes standard errors, and $\overline{milex}$ denotes the mean value of the military burden in the given study. These marginal effects need to be further adjusted to be directly comparable. Therefore, following Cazachevici et al (2020), I transform them into partial correlation coefficients (PCCs), as described in equations (5) for the PCC and (6) for its standard error:

$$PCC_{is} = \frac{t_{is}}{\sqrt{t_{is}^2 + df_{is}}} \tag{5}$$

$$SE_{PCC} = \frac{PCC_{is}}{t_{is}}, \tag{6}$$

where i denotes the i-th estimate within the study s, t denotes t-value of the corresponding estimate, and df specifies degrees of freedom.

The next step is the selection of explanatory variables for the heterogeneity analysis. There are two conditions for the inclusion of a collected variable in the heterogeneity analysis. First, its variance inflation factor (VIF) must be below 10. If this is not the case, the variable is included in the analysis only if sufficiently justified. Second, dummy variables are included if their mean value presented in summary statistics ranges between 0.03 and 0.97 to ensure sufficient variation. Variables that do not satisfy this condition are excluded or merged with other variables.

Next, I rescale the numerical variables using min-max normalization. The dataset consists of both numerical and dummy variables and there is a considerable risk of a distortion of the heterogeneity analysis results. For example, the time span of the primary studies, a numerical variable, has a much bigger range (between 1 and 72), than dummy variables.

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{7}$$

The variables are grouped into five predefined variable categories – data characteristics, estimation characteristics, publication characteristics, and structural variation.

Data characteristics capture the type of dataset (cross-section, time series, panel). However, the time-series variable is perfectly correlated with a variable capturing single-country studies. For the sake of a clearer interpretation, I choose the latter variable. Cross-sectional datasets serve as the base type of dataset and are not included among the regressors to avoid a dummy variable trap. The analysis also includes a number of observations within the primary studies, the time spans of their datasets, and the number of countries included within their datasets. Another aspect of the data is also whether the authors of primary studies used multi-year averages of the economic growth (mostly 5- or 7-year averages), which may indicate a long-term impact of the military expenditures.

Estimation characteristics account for differences in econometric techniques and theoretical models used in primary studies. The analysis includes variables accounting for studies using non-linear functional forms (squares and interactions with other variables) and distinguishes between the Augmented Solow model (ASM) and the Barro-type growth model (BM). It also includes variables capturing estimations using ordinary least squares (OLS), generalized method of moments (GMM), fixed-effects model (FEM), and generalized least squares (GLS).

Publication characteristics include a count of citations standardized by the number of years since the primary study was published, which serve as a proxy for its recognized quality. I also considered including a variable capturing years since the primary paper was published, but given its VIF of 12.1, I rejected it.

Structural variation is motivated by possible implicit differences in analytical outputs of primary studies caused by the environment. For this reason, this paper includes dummies for decades captured by datasets of primary studies. These consist of a period from the 1950s to the 2000s. The base decade is set to the 1980s, which marked the peak of the Cold War arms race and presents an interesting comparison with respect 1990s. The next variables are motivated by differences among geographic regions based on socio-economic background. The only sufficiently represented regions are Africa and Asia. The socio-economic aspect also allows to distinguish between less-developed countries (LDCs) and developed countries.

Moreover, the primary studies controlled for factors that may affect the impact of military expenditures on economic growth. The sufficiently represented variables are population growth, education, natural resources, conflicts, foreign direct investments, neighboring countries' military expenditure, democracy, and corruption.

The last aspect worth analyzing is the data source. Following a note from Lebovic and Ishaq (1987), the individual data sources may differ in distinguishing between civilian and military spending, the inclusion of internal security forces, the use of US dollars or national currency, and the use of constant or nominal prices. Therefore, this paper includes dummy variables capturing estimates using datasets from the Stockholm International Peace Research Institute (SIPRI), the World Bank, and the U.S. Arms Control and Disarmament Agency (ACDA).

The paper presents the list of independent variables in Table 6 within a the Appendix A, including a basic description and summary statistics.

3.1 Descriptive analysis

The statistical summary shows the average effect of military expenditure on economic growth to be negative. The numerical summary of chosen primary studies is in Table 7 in Appendix A, which contains the number of estimates produced by individual primary studies, a plain average of their PCCs, and their fixed effects mean, calculated with the following formula:

$$mean_{fixed}(PCC) = \frac{\sum w_{is} PCC_{is}}{\sum w_{is}}, \tag{8}$$

where the weight w_{is} is calculated as the squared inverse of PCC standard error of the estimate i in the study s.

The plain mean of PCCs takes a value of -0.057, while the fixed-effects mean is equal to -0.072 and statistically significant. This is an interesting turn in comparison to the results of Alptekin and Levine (2012), which finds a fixed-effects mean of 0.056. My result is therefore similar in magnitude but has an opposite sign. Following Doucouliagos (2011), the negative effect of military expenditures is economically non-significant, since it is smaller than 0.104 in absolute value. The value of 0.104 is specified for the subset of economic literature concerning economic growth. Should I use a threshold of 0.07 specified for economic literature in general, the fixed-effect mean would be economically significant. The turn in the effect direction may be explained by the newly included primary studies, which show a fixed-effects mean of -0.147, which is both statistically and economically significant, yet small as per Doucouliagos (2011). This paper presents more detailed results concerning the newly included studies in Appendix B. However, these estimates are valid only if no publication bias is present.

Histograms depicted in Figures 2 and 3 further demonstrate the effect sizes. Figure 3 shows that the blue histogram representing the newly collected PCCs is shifted to the left and are on average more negative than the PCCs from Alptekin and Levine (2012), colored in red.

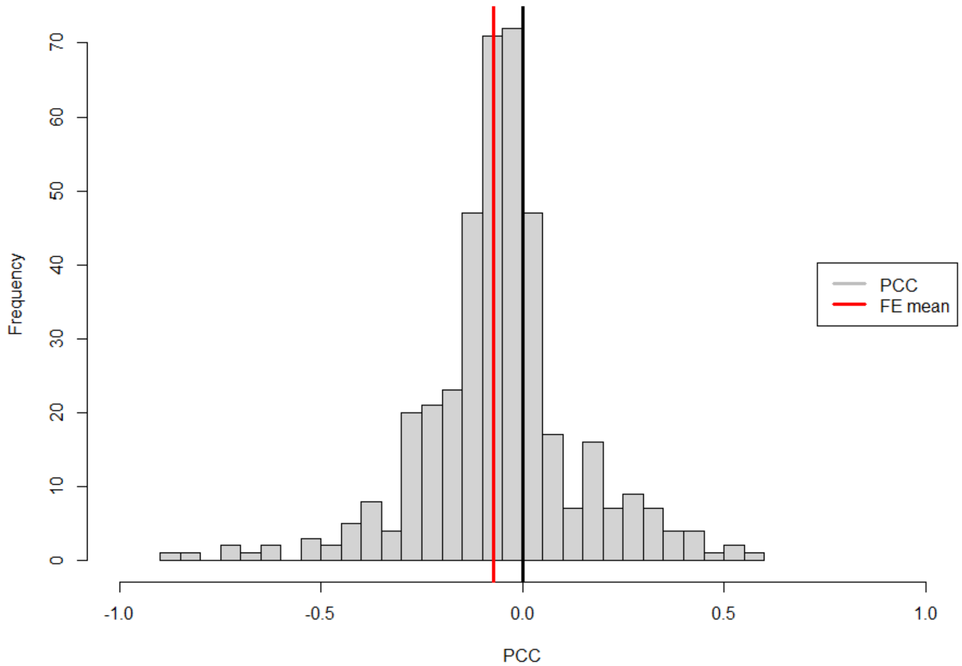

Figure 2: Histogram of all partial correlation coefficients from selected primary studies

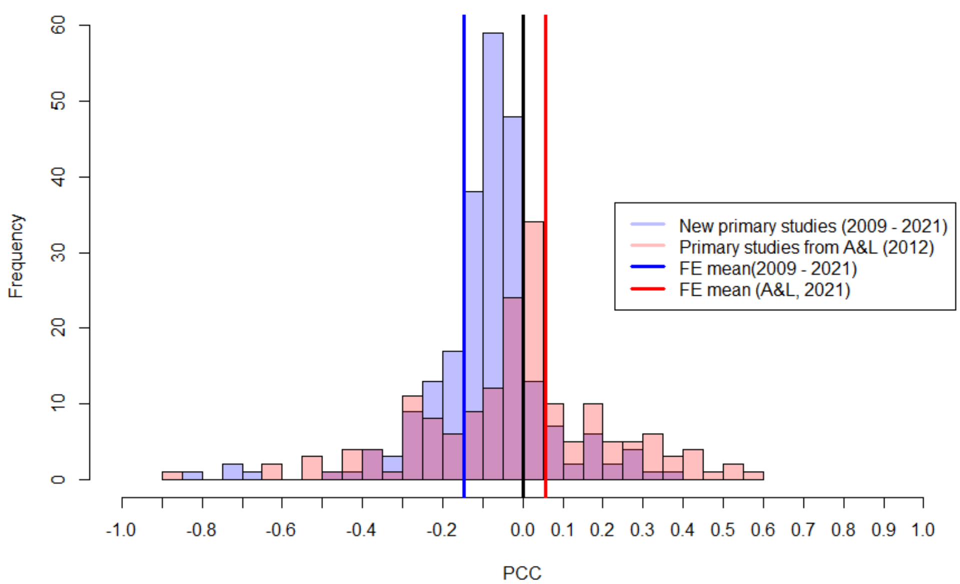

Figure 3: Comparison of partial correlation coefficients from Alptekin & Levine (2012) and partial correlation coefficients from new primary studies (2009 - 2021)

It is also worth inspecting, whether there are outliers that may affect further analysis, and therefore, whether I need to winsorize the data. Figure 4 depicts boxplots, which show some values outside the range of whiskers both for the estimates and the standard errors. The dataset needs to be winsorized. This paper performs winsorization as defined in Miller (1993) and adjusts 2.5 percent of both the maximum and minimum of the collected data points.

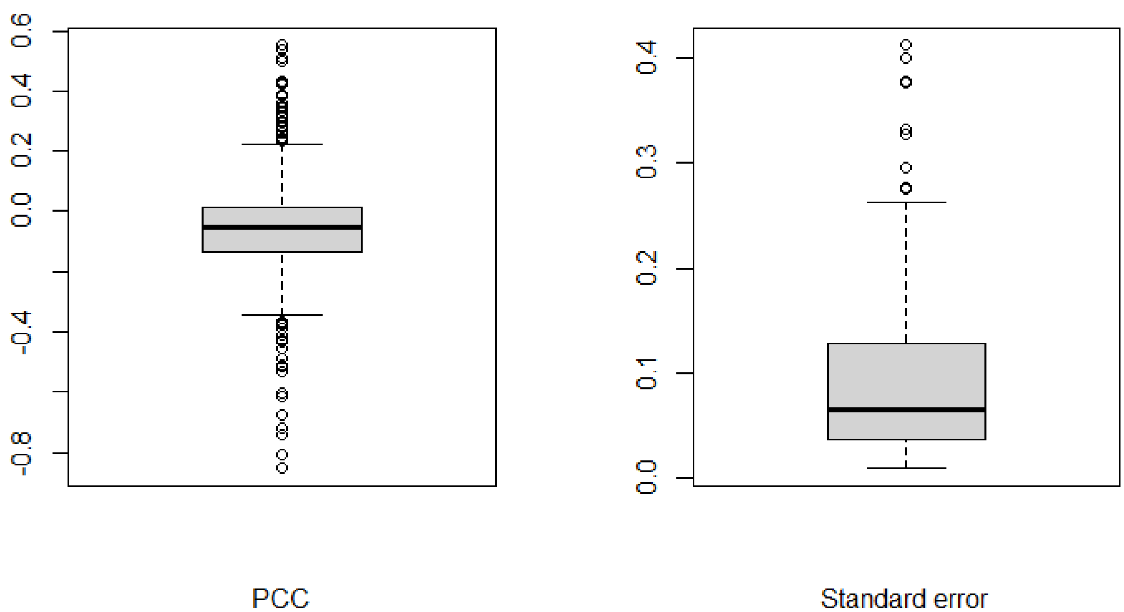

Figure 4: Boxplot of estimates and related standard errors

4 Methodology

The analysis consists of publication bias analysis and heterogeneity analysis. The publication bias analysis is divided into three parts – linear techniques, non-linear techniques, and techniques that allow for endogeneity in standard errors.

First, this paper uses a funnel plot for visual analysis (Egger et al., 1997). The funnel plot visualizes estimates of the effects against an inverse of the standard errors, which serve as precision proxies. However, such visual analysis is hardly sufficient and is accompanied by a FAT-PET-PEESE test (Stanley and Doucouliagos, 2012). The techniques used to estimate the linear tests are precision-weighted least squares, study-weighted least squares, fixed-effects model, and between-effects model.

There is no guarantee that the publication bias is linearly related to the standard error. For this reason, this paper employs non-linear techniques of publication bias analysis. These consist of four additional tests – the Weighted Average of Adequately Powered (WAAP) by Ioannidis et al. (2017), Stem-based method by Furukawa (2019), Selection model by Andrews and Kasy (2019), and Endogenous Kink by Bom and Rachinger (2019).

Techniques allowing for endogeneity fix a shortcoming of the prior techniques, which do not account for a relationship between estimates and standard errors and assume that any correlation between them arises due to publication bias. Such techniques are two-stage least squares, p-uniform* developed by van Aert and van Assen (2018), the Caliper test (Gerber and Malhotra, 2008), and the instrumental variable technique performed with linear techniques.

Heterogeneity analysis aims to describe which explanatory variables influence the estimates found in primary studies and how. In practice, it means using the following regression equation:

$$PCC_{is} = \beta_0 + \sum_{k}^{38} (\beta_k X_{k,is}) + \gamma SE(PCC_{is}) + \epsilon_{is}, \tag{9}$$

where X_k denotes the vector of independent variables and β_k denotes their respective coefficients. The equation also includes the standard error of the PCCs, which represents the publication bias. The intercept represents the PCC corrected for publication bias. However, in the case of the heterogeneity analysis, it should not be interpreted separately since it is conditional on other variables.

There are too many independent variables to use simple estimation techniques. Including all variables in the model would inflate the variance of parameters and make them useless and choosing a single suitable model is impossible since 38 independent variables collected yield 2^{38} model combinations. Since there is such a high degree of model uncertainty, the best approach is model averaging, specifically Bayesian model averaging (BMA). Furthermore, I use Frequentist model averaging (FMA), and hybrid frequentist-Bayesian model as robustness checks as in Gechert et al. (2022).

The BMA needs to be assigned with initial parameters. First, I assign the prior probabilities using a dilution prior to address possible collinearity problems (George, 2010). Second, I use unit information g-prior as suggested by Eicher et al. (2011). Since this paper uses a Bayesian model selection algorithm (Zeugner and Feldkircher, 2015), there are four more parameters. First, I choose the "birth-death" Markov Chain Monte Carlo (MCMC) algorithm. Second, I choose 300 000 model iterations. Third, I choose 100 000 burn-ins to let the MCMC algorithm form at least some distribution before converging to the equilibrium distribution (Johansen et al., 2010). Fourth, there are 50 000 "best" models to form the final model.

The FMA does not require any specification of prior probabilities and derives estimates only from provided data. Still, FMA needs a specification of weights since different weights yield different model properties (Wang, et al., 2009). The choice of weights is determined by Mallow's model averaging (MMA; Hansen, 2007). A one-by-one estimation of every model would be computationally very demanding, therefore I reduce the model space by orthogonalization of the covariates (Amini and Parmeter, 2012). The hybrid model averaging employs elements of both BMA and FMA. I choose the appropriate variables for the model using posterior inclusion probabilities (PIP) from BMA with respect to specified thresholds. In this specific case, the threshold is 0.5. There are 2 reasons for this threshold level. The basic logic is that a PIP higher than 0.5 indicates that the chance, that the variable should be part of the model, is better than random. The second reason is a classification of the significance of the variables. Generally, there are four categories of significant explanatory variables (Jeffreys, 1998). First, the explanatory variables are weakly significant if they are between the PIP value of 0.5 and 0.75. Second, the explanatory variable is substantially significant if the PIP is between 0.75 and 0.95. Third, if the PIP is larger than 0.95 and smaller than 0.99, the variable is strongly significant. Last, all variables with PIP larger than 0.99 are decisively significant. I estimate the hybrid model using OLS (Gechert, et al., 2022).

Given the heterogeneous nature of the analysis, I execute it using multiple means. Most tests and regressions are executed using RStudio. The Selection model and p-uniform* are performed in a web applications created by their authors. The R code is not available for the Endogenous Kink test, and therefore I conduct it in Stata using the code devised by Bom and Rachinger.

5 Results

5.1 Publication bias and average effect

Same as the descriptive analysis, the publication bias analysis shows the overall effect of military expenditure on economic growth to be negative and statistically significant. The first part of the publication bias analysis is visual using the funnel plot, which is displayed in Figure 5 and uses the data before winsorization. From the funnel plot, it is difficult to assess the publication bias, even though the values spread further from the mean value suggest some asymmetry, since there seem to be slightly more estimates to the left side of the plot. Nonetheless, this simplistic analysis hardly provides sufficient evidence of publication bias.

Table 1 presents the results of the publication bias tests. According to the Panel A, the precision-weighted WLS estimates the coefficient of the standard error equal to 0.316 and statistically insignificant. Therefore, I cannot reject the absence of publication bias. The same does not hold for the regression weighted by study size, fixed-effects regression, and between-effects regression. Still, following Doucouliagos and Stanley (2013), I classify the

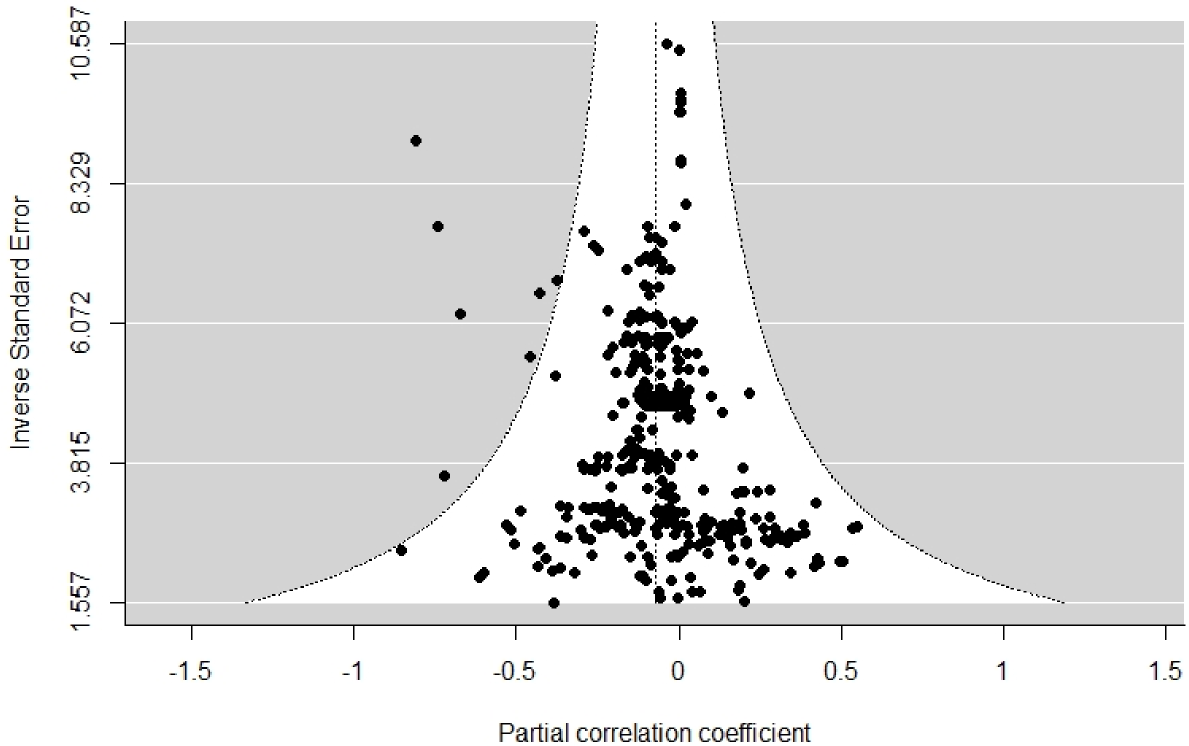

Figure 5: Funnel plot of sizes of partial correlation coefficients against precision measured as an inverse of estimate's standard error

publication bias as small, because the related coefficients do not exceed the absolute value of one. The mean beyond bias reveals coefficients ranging from -0.107 in the case of the model weighted by the study size and -0.059 for the fixed-effects model. Besides the study-weighted WLS, all other estimates are economically insignificant.

In comparison to the linear tests, the non-linear methods presented in the Panel B resemble a lower magnitude of the estimated means beyond bias. The results are consistent with the previous ones since all of the non-linear models yield negative estimates of the mean beyond bias. The mean beyond bias revealed by the non-linear models yields values within an interval of (-0.072 ; -0.052). Stem-based method is the only one that yields a statistically non-significant estimate. Given its nature, I assume that this deviation in results arises because of the exclusion of too many estimates, making its result less robust. However, the rest of the non-linear methods reveal means beyond bias which are in line with previous analyses.

Table 1: Publication bias tests results

Panel A: Linear models				
	Precision	Study	FE	BE
Publication bias	0.316	0.616**	-0.186*	0.429**
	(0.171)	(0.225)	(0.074)	(0.141)
Mean beyond bias	-0.083***	-0.107***	-0.059***	-0.092***
	(0.013)	(0.027)	(0.003)	(0.011)
Panel B: Non-linear models				
	Stem-based	WAAP	Selection model	Endogenous kink
Mean beyond bias	-0.052	-0.057***	-0.072***	-0.070***
	(0.041)	(0.009)	(0.011)	(0.012)
Panel C: Models controlling for endogeneity				
	2SLS	p-uniform*		
Publication bias	0.376	0.189		
	(0.194)	(p = 0.091)		
Mean beyond bias	-0.088***	-0.065***		
	(0.014)	(0.015)		

The Panel C reports the results of the 2SLS and p-uniform*. The 2SLS estimates the publication bias as statistically insignificant, while the mean beyond bias is estimated at -0.088 and highly statistically significant, corresponding with the battery of linear tests. The p-uniform* finds a statistically significant effect size of -0.065 with a standard error of 0.015, which corresponds with the non-linear, rather than the linear tests, and provides another argument in favor of the economically non-significant effect of military expenditures on economic growth. The publication bias revealed by the test shows an estimate of 0.189 with a p-value of 0.091.

The final test, the caliper test, looks for jumps in t-statistics around a level of 1.96, which corresponds with a 5 percent statistical significance level. By a plain look at the distribution of the t-statistics depicted in Figure 6, there is a small increase in observations on the left side of the -1.96 threshold and a bigger one on the right side of the 1.96 threshold, which hints at a publication bias. The caliper test follows. The chosen caliper widths around the threshold for this test are 0.1, 0.2, and 0.3.

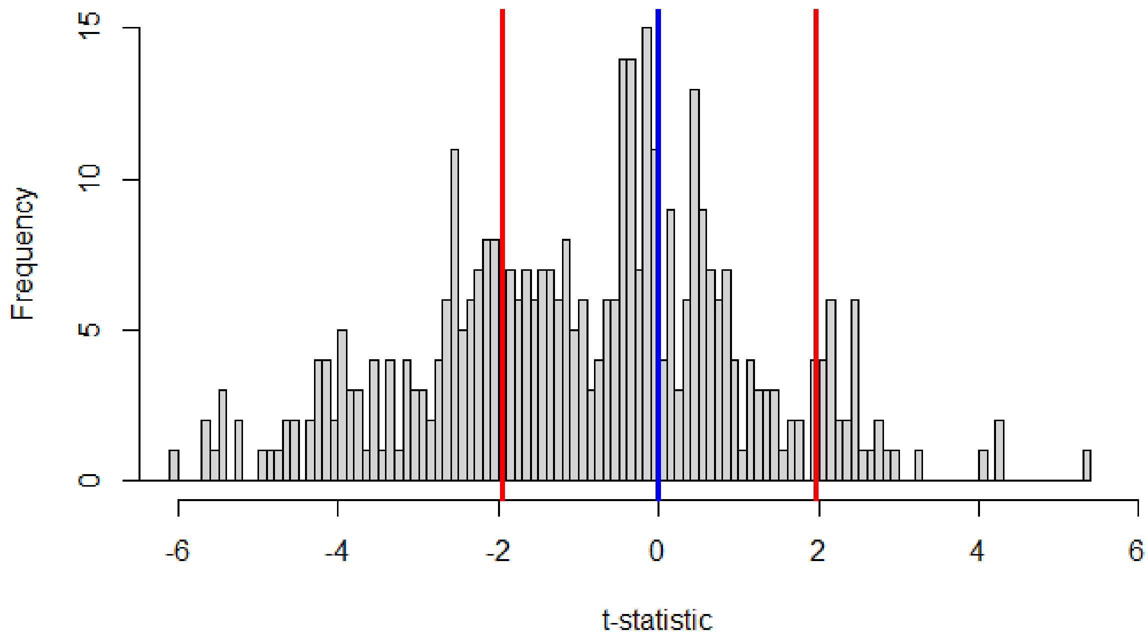

Figure 6: Histogram of t-statistics and vertical lines signifying -1.96, 0, and 1.96 values as visual analysis before the caliper test

The caliper tests in Table 2 show a significant difference between the number of estimates below and above the threshold. In the case of the interval between -2.06 and -1.86, 76 percent of estimates find themselves in the lower half of the interval. A similar percentage holds for the latter caliper widths, suggesting a presence of publication bias. The caliper test at the upper threshold of 1.96 reveals a significant disproportion in favor of estimates that find themselves on the right side of the threshold suggesting a publication bias towards statistically significant positive results.

Table 2: Caliper test results

	width = 0.1	width = 0.2	width = 0.3
threshold for t-statistic = -1.96	-0.260**	-0.275***	-0.292***
	(0.069)	(0.047)	(0.036)
Observations	14	27	42
threshold for t-statistic = 1.96	0.423**	0.422***	0.398***
	(0.080)	(0.047)	(0.041)
Observations	6	13	19

5.2 Heterogeneity analysis

Figure 7 depicts the first illustration of the heterogeneity analysis results. It shows the inclusion of individual explanatory variables in models drawn by the model-averaging process and sorts them by their PIP. The horizontal axis shows the posterior model probability (PMP). The colors of stripes associated with individual variables indicate whether their effect is rather positive (blue color) or negative (orange color). The variable capturing individual studies using data from the 1990s has the highest PIP of 0.714, marking it as weakly significant, since it is smaller than 0.75 (Jeffreys, 1998). A bundle of variables from the data characteristics category follows, all of which have PIPs between 0.54 and 0.60, and are therefore also weakly significant with respect to the PCCs.

Table 3 describes the results of the heterogeneity analysis. The first conclusion that can be drawn from it is that the publication bias is absent or very small since it has both a very low PIP and posterior mean. Therefore, the heterogeneity analysis, controlling for effects of other factors, supports the results of publication bias analysis and adds information that the effect of the standard error on the PCCs in previous parts may have been overestimated.

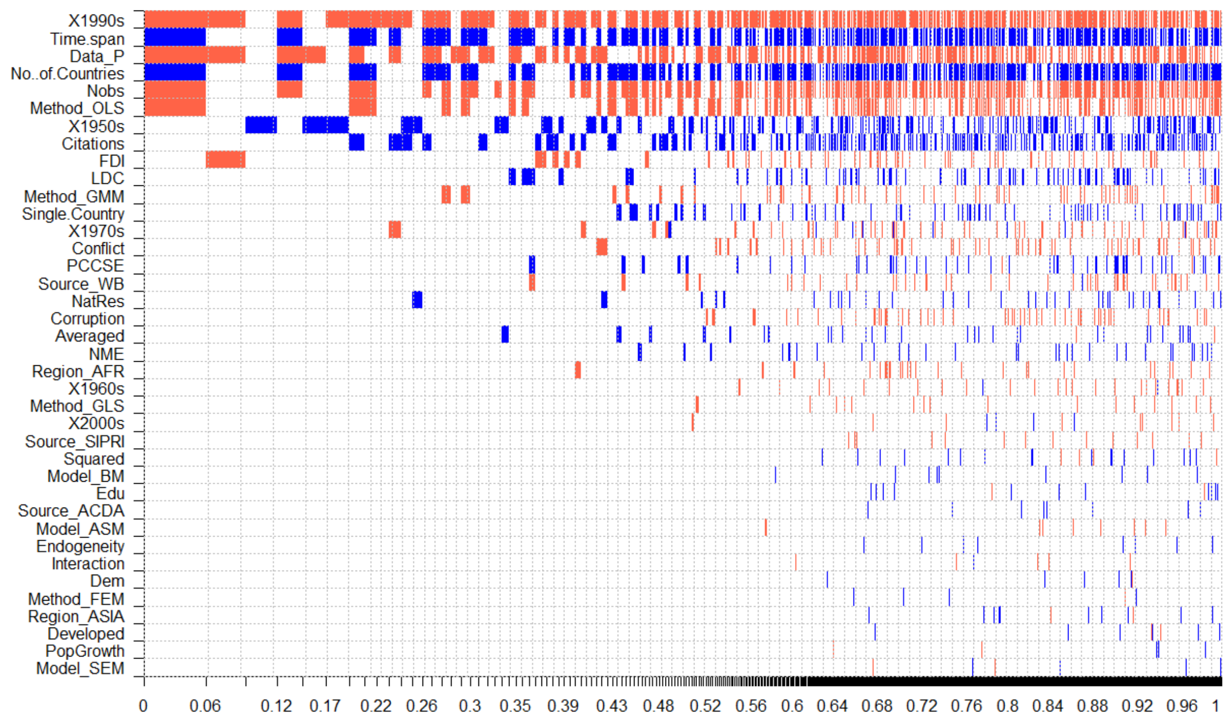

Figure 7: Model inclusion in Bayesian model averaging

Data characteristics The Table 3 shows, that most of the variables that significantly affect the PCC belong to the category of data characteristics. The PCC seems to be the most affected by the time span of individual studies. According to the posterior mean of the time span, a shift between the shortest and the longest time span of primary studies results in adding 0.165 to the PCC. This effect is even larger and statistically significant when using a hybrid frequentist robustness check. The explanation for such a result is provided by Alptekin and Levine (2012). Their reasoning claims that the longer the time series, the higher the probability that periods of high military expenditures are followed by periods of low military expenditures, and the estimated effect will converge to a neutral level. Since the average estimated effect is negative and the estimated effect of the time span is positive, this explanation holds for the case of this paper as well. This result also presents some indirect evidence in favor of the non-linearity hypothesis.

Table 3: Results of BMA and hybrid check

Response variable:	Bayesian model averaging			Frequentist check		
PCC	Post. mean	Post. SE	PIP	Coef.	SE	t-statistic
Intercept	-0.077		1.000	-0.075	0.026	-2.967
SE (publication bias)	0.038	0.150	0.081			
Data characteristics						
Panel	-0.040	0.041	0.545	-0.071	0.020	-3.469
Single-country study	0.010	0.035	0.102			
No. of observations	-0.128	0.128	0.552	-0.193	0.056	-3.444
No. of countries	0.136	0.133	0.568	0.209	0.046	4.550
Time span	0.165	0.150	0.595	0.276	0.054	5.147
Averaged data	0.002	0.013	0.047			
Structural variation						
SIPRI	-0.001	0.006	0.024			
WB	-0.005	0.021	0.067			
ACDA	0.001	0.009	0.017			
Africa	-0.003	0.015	0.043			
Asia	0.000	0.006	0.014			
Developed country	0.000	0.004	0.014			
LDC	0.008	0.022	0.139			
D1950	0.036	0.062	0.284			
D1960	-0.001	0.010	0.034			
D1970	-0.006	0.026	0.103			
D1990	-0.062	0.047	0.714	-0.078	0.019	-4.142
D2000	-0.000	0.007	0.021			
NME	0.004	0.020	0.051			
Education	0.000	0.005	0.020			
Natural resources	0.004	0.020	0.063			
Population growth	-0.000	0.002	0.012			
Corruption	-0.005	0.020	0.061			
Conflict	-0.007	0.025	0.094			
FDI	-0.014	0.037	0.137			
Democracy	0.001	0.008	0.018			

Response variable:	Bayesian model averaging			Frequentist check		
PCC	Post. mean	Post. SE	PIP	Coef.	SE	t-statistic
Estimation characteristics						
ASM	-0.000	0.003	0.014			
BM	0.000	0.005	0.019			
SEM	0.000	0.005	0.012			
OLS	-0.023	0.034	0.361			
GMM	-0.009	0.026	0.123			
GLS	-0.002	0.015	0.035			
FEM	0.000	0.004	0.015			
Endogeneity	0.000	0.004	0.017			
Squared	0.000	0.007	0.018			
Interaction	-0.000	0.004	0.015			
Publication characteristics						
Citations	0.052	0.086	0.310			
Studies	67			67		
Observations	405			405		

The number of countries within each study is significant and positive as well, but to a smaller degree, adding 0.136 to the magnitude of the PCC, resp. 0.209 as shown by the hybrid check. The number of observations included in the primary studies has a negative coefficient and a difference between the minimum and maximum values of this variable decreases the PCC by 0.128. According to the hybrid model, the decrease is even bigger.

The dummy variable capturing the use of panel datasets in primary studies has the weakest impact among the significant data characteristics. A higher posterior standard error signifies its relatively weaker impact. The hybrid robustness check confirms both the significance of the chosen variables and their signs. The only small deviation from the BMA is the statistically significant coefficient of the panel dummy variable.

Structural variation Among the variables from the structural variation category, the only variable with high PIP is the variable capturing estimates using data from the 1990s.

This is not surprising since the 1990s presented a major de-escalation period after the end of the Cold War and the end of an arms race in the 1980s. The important role of the 1990s is confirmed by the hybrid model.

Estimation characteristics The estimation characteristics yield no variable that significantly impacts the PCC. The only variable, that reads at least a little elevated PIP, is the one capturing the use of ordinary least squares with a PIP equal to 0.361. Still, it refutes the claim from d'Agostino et al. (2013) that estimation techniques not controlling for endogeneity tend to push estimated effects upward. The theoretical models provide no sizeable effect as well.

Publication characteristics The number of citations standardized by the number of years since the publication of the primary study does not have a significant effect on the PCC. This means that the subsequent research literature uses all types of results and does not favor any found direction of estimated effect. A possible explanation may be the political and economic controversy of military expenditures, which may motivate researchers to explain all mechanisms through which military expenditures may affect economic growth and provide extensive evidence. Alternatively, the result may also lead to a conclusion that the primary studies are of the same perceived quality, regardless of their results.

The results of MMA are in the Appendix A (Table 8) and present two important similarities with previous results. First, the time span of the primary study is positively associated with the PCC and is statistically significant. Second, the dummy variable capturing the period of the 1990s negatively affects the PCC and is statistically significant as well. However, the BMA and subsequent hybrid approach revealed a lower magnitude of the effect.

The economic effects may be deduced from an application of changes in the key variables and the subsequent effect on the PCC. Table 4 shows that even though panel datasets and 1990s exhibit statistical significance, their economic effect is insignificant as specified by Doucouliagos (2011). On the other hand the rest of the variables, resp. their maximum change has a "small" economic impact. A large difference in the effects when comparing one-standard-deviation change and maximum change shows how influential the numeric variables are.

Table 4: Economic significance of key variables

	One-std.-dev. change		Maximum change	
	Effect on PCC	% of FE mean	Effect on PCC	% of FE mean
Panel	-0.018	-25 %	-0,040	-56 %
No. of observations	-0.024	-33 %	-0.128	-178 %
No. of countries	0.030	41 %	0.136	189 %
Time span	0.030	42 %	0.165	229 %
D1990	-0.030	-42 %	-0.062	-86 %

The analysis of the whole dataset puts more emphasis on the general characteristics of the data, rather than other specifications. This may be a result of a connection of two datasets, one by Alptekin and Levine and the one created by the author of this paper, into one. The differences between the two datasets are embedded mostly in their general characteristics since in the last decade, panel datasets became more popular and better analyzable. This, together with more data at disposal may explain why the results favor data characteristics so heavily and why the model does not provide significant results with respect to other classes of variables.

Appendix provides more information about the new studies used in this analysis.

6 Conclusion

This paper aims to provide a comprehensive summary of the literature on peace economics and synthesize the findings from relevant studies. The meta-analysis, covering primary studies published from 1978 to 2021, reveals statistically significant but largely economically non-significant estimates of the impact of military expenditures on economic growth. These estimates range from -0.107 to -0.052. This paper updates both the dataset, now incorporating more primary studies and estimates along with additional explanatory variables, and the methodology, utilizing multiple tests for publication bias and diverse methods for heterogeneity analysis, including Bayesian model averaging, frequentist analysis, and hybrid robustness checks. It stands as the first extensive meta-analysis of the influence of military expenditures on economic growth, employing modern methods to ensure robust results.

In addition to identifying a statistically significant negative overall effect of military expenditures on economic growth, the heterogeneity analysis uncovers a statistically significant impact of data characteristics from primary studies on this effect. These include the panel structure, number of observations, number of countries, and time span of the primary studies' datasets. Notably, data from the 1990s also exhibit a statistically significant effect.

This reveals a potential limitation in this paper. The predominant focus on numerical data characteristics overshadowed other mainly categorical variables, limiting the specificity of conclusions drawn from the analysis. Beyond the challenges in the heterogeneity analysis, the publication bias analysis faced limitations, as the Stem-based test by Furukawa excluded a considerable amount of estimates to draw strong conclusions.

The practical implications of this paper arise from the overall negative effect of military spending on economic growth and the adverse impact of the 1990s. The former suggests a positive economic outcome from reducing military burden and fostering international dialogue between countries to resolve disputes, alleviating the pressure to maintain strong and costly armies. The latter result provides concrete evidence for this conclusion, given the overall reduction in international tensions during the 1990s, subsequently leading to a worldwide decrease in military expenditures.

References

Amini, S. M. & Parmeter, C. F. (2012), 'Comparison of model averaging techniques: Assessing growth determinants', *Journal of Applied Econometrics* **27**(5), 870–876.

Andrews, I. & Kasy, M. (2019), 'Identification of and correction for publication bias', *American Economic Review* **109**(8), 2766–94.

Bom, P. R. & Rachinger, H. (2019), 'A kinked meta-regression model for publication bias correction', *Research synthesis methods* **10**(4), 497–514.

Cazachevici, A., Havranek, T. & Horvath, R. (2020), 'Remittances and economic growth: A meta-analysis', *World Development* **134**, 105021.

Chan, S. (1988), 'Defense burden and economic growth: Unraveling the taiwanese enigma', *American Political Science Review* **82**(3), 913–920.

Compton, R. A. & Paterson, B. (2016), 'Military spending and growth: the role of institutions', *Defence and Peace Economics* **27**(3), 301–322.

d'Agostino, G., Dunne, J. P. & Pieroni, L. (2013), 'Military expenditure, endogeneity and economic growth'.

Deger, S. (1986), 'Economic development and defense expenditure', *Economic development and cultural change* **35**(1), 179–196.

Deger, S. & Sen, S. (1983), 'Military expenditure, spin-off and economic development', *Journal of development economics* **13**(1-2), 67–83.

DeGrasse, R. W. (2016), *Military expansion, economic decline: Impact of military spending on United States economic performance*, Routledge.

Doucouliagos, C. (2011), How large is large? preliminary and relative guidelines for interpreting partial correlations in economics, Technical report, Deakin University, Department of Economics.

Doucouliagos, C. & Stanley, T. D. (2013), 'Are all economic facts greatly exaggerated? theory competition and selectivity', *Journal of Economic surveys* **27**(2), 316–339.

Dunne, J., Watson, D. et al. (2005), Manufacturing growth, technological progress, and military expenditure, Technical report.

Egger, M., Smith, G. D., Schneider, M. & Minder, C. (1997), 'Bias in meta-analysis detected by a simple, graphical test', *Bmj* **315**(7109), 629–634.

Eicher, T. S., Papageorgiou, C. & Raftery, A. E. (2011), 'Default priors and predictive performance in bayesian model averaging, with application to growth determinants', *Journal of Applied Econometrics* **26**(1), 30–55.

Faini, R., Annez, P. & Taylor, L. (1984), 'Defense spending, economic structure, and growth: Evidence among countries and over time', *Economic development and cultural change* **32**(3), 487–498.

Furukawa, C. (2019), 'Publication bias under aggregation frictions: Theory, evidence, and a new correction method', *Evidence, and a New Correction Method (March 29, 2019)* .

Gechert, S., Havranek, T., Irsova, Z. & Kolcunova, D. (2022), 'Measuring capital-labor substitution: The importance of method choices and publication bias', *Review of Economic Dynamics* **45**, 55–82.

George, E. I. et al. (2010), 'Dilution priors: Compensating for model space redundancy', *Borrowing Strength: Theory Powering Applications–A Festschrift for Lawrence D. Brown* **6**, 158–165.

Gerber, A., Malhotra, N. et al. (2008), 'Do statistical reporting standards affect what is published? publication bias in two leading political science journals', *Quarterly Journal of Political Science* **3**(3), 313–326.

Hansen, B. E. (2007), 'Least squares model averaging', *Econometrica* **75**(4), 1175–1189.

Havranek, T. & Irsova, Z. (2011), 'Estimating vertical spillovers from fdi: Why results vary and what the true effect is', *Journal of International Economics* **85**(2), 234–244.

Heo, U. (2010), 'The relationship between defense spending and economic growth in the united states', *Political Research Quarterly* **63**(4), 760–770.

Ioannidis, J. P., Stanley, T. D. & Doucouliagos, H. (2017), 'The power of bias in economics research'.

Jeffreys, H. (1998), *The theory of probability*, OUP Oxford.

Johansen, A. M., Evers, L. & Whiteley, N. (2010), 'Monte carlo methods', *Lecture notes* **200**.

Khidmat, W. B., Wang, M. & Iqbal, F. (2018), 'Impact of defense spending on economic growth: Evidence from developing nations of asia', *Journal of Sustainable Development Studies* **11**(2).

Landau, D. (1996), 'Is one of the 'peace dividends' negative? military expenditure and economic growth in the wealthy oecd countries', *The Quarterly Review of Economics and Finance* **36**(2), 183–195.

Landau, D. L. (1993), *The economic impact of military expenditures*, Vol. 1138, World Bank Publications.

Lebovic, J. H. (1999), 'Using military spending data: The complexity of simple inference', *Journal of Peace Research* **36**(6), 681–697.

Lebovic, J. H. & Ishaq, A. (1987), 'Military burden, security needs, and economic growth in the middle east', *Journal of Conflict Resolution* **31**(1), 106–138.

Lipow, J. & Antinori, C. M. (1995), 'External security threats, defense expenditures, and the economic growth of less-developed countries', *Journal of policy modeling* **17**(6), 579–595.

Miller, J. N. (1993), 'Tutorial review—outliers in experimental data and their treatment', *Analyst* **118**(5), 455–461.

Mintz, A. (1989), 'Guns versus butter: A disaggregated analysis', *American Political Science Review* **83**(4), 1285–1293.

NATO (2022), 'Defence expenditures of nato countries (2014-2022)'.

Nugroho, D. A. & Purwanti, E. Y. (2021), 'Impact of military expenditure on economic growth encouraging or constraining?', *JEJAK: Jurnal Ekonomi dan Kebijakan* **14**(1), 9–20.

Rahman, T. & Siddiqui, D. A. (2019), 'The effect of military spending on economic growth in the presence of arms trade: A global analysis', *Available at SSRN 3401331* .

Ram, R. (2019), 'Conceptual linkages between defense spending and economic growth and development: A selective review', *Defense spending and economic growth* pp. 19–39.

Stanley, T. D. & Doucouliagos, H. (2012), *Meta-regression analysis in economics and business*, routledge.

Stroup, M. D. & Heckelman, J. C. (2001), 'Size of the military sector and economic growth: A panel data analysis of africa and latin america', *Journal of Applied Economics* **4**(2), 329–360.

van Aert, R. C. & van Assen, M. A. (2018), 'Correcting for publication bias in a meta-analysis with the p-uniform* method'.

Wang, H., Zhang, X. & Zou, G. (2009), 'Frequentist model averaging estimation: a review', *Journal of Systems Science and Complexity* **22**(4), 732–748.

Ward, M. D. & Davis, D. R. (1992), 'Sizing up the peace dividend: economic growth and military spending in the united states, 1948–1996', *American Political Science Review* **86**(3), 748–755.

Zeugner, S. & Feldkircher, M. (2015), 'Bayesian model averaging employing fixed and flexible priors: The bms package for r', *Journal of Statistical Software* **68**, 1–37.

Appendix A (for online publication)

A.1 Tables

Table 5: List of included primary studies

Study	Data type*	Model**	Time period
Benoit (1978)	CS	Other	1950 - 1965
Deger & Smith (1983)	CS	SEM	1965 - 1973
Lim (1983)	CS	Other	1950 - 1973
Deger & Sen (1983)	CS	SEM	1965 - 1973
Faini (1984)	TS	Other	1950 - 1972
Cappelen et al. (1984)	P	SEM	1960 - 1980
Landau (1986)	P	Other	1960 - 1980
Biswas & Ram (1986)	CS	Other	1960 - 1977
Deger (1986)	CS	Other	1965 - 1973
Lebovic & Ishaq (1987)	P	Other	1973 - 1982
Chan (1988)	TS	Other	1961 - 1985
Grobar and Porter (1989)	CS	Other	1950 - 1965
Gyimah-Brempong (1989)	P	SEM	1973 - 1983
Looney (1989)	CS	Other	1970 - 1982
Landau (1993)	CS/P	Other	1969 - 1989
Dunne & Mohammed (1995)	CS/P	SEM	1967 - 1985
Lipow & Antinori (1995)	CS	Other	1974 - 1989
Blomberg (1996)	P	BM	1967 - 1982
Knight et al. (1996)	CS/P	ASM	1971 - 1985
Landau (1996)	P	Other	1950 - 1990
Brumm (1997)	CS	BM	1974 - 1989
Antonakis (1997)	TS	SEM	1960 - 1990
Heo & DeRouen (1998)	TS	Other	1961 - 1990
DeRouen (2000)	TS	Other	1953 - 1992
Stroup & Heckelman (2001)	P	BM	1975 - 1989
Galvin (2003)	CS	SEM	1999
Aizenman & Glick (2006)	CS	BM	1989 - 1998
Bose et al. (2007)	P	BM	1970 - 1989
Kollias et al. (2007)	P	Other	1961 - 2000
Yakovlev (2007)	P	ASM/BM	1960 - 2000
Looney & McNab (2008)	P	Other	1999 - 2003
Cooray (2009)	CS	ASM	1996 - 2003
Pieroni (2009)	CS	BM	1989 - 1998

Study	Data type*	Model**	Time period
Pieroni & d'Agostino (2009)	CS	BM	1989 - 1998
Aikaeli & Mlamka (2010)	CS	Other	2001 - 2005
d'Agostino et al. (2010)	P	BM	2003 - 2007
Heo (2010)	TS	ASM	1954 - 2005
d'Agostino (2011)	TS	BM	1958 - 2005
Kalaitzidakis & Tzouvelekas (2011)	P	BM	1980 - 1995
Iftikar & Ali (2012)	P	Other	1984 - 2003
Hou & Chen (2013)	CS/P	ASM	1975 - 2009
Na & Bo (2013)	P	ASM	1990 - 2006
Hou & Chen (2014)	P	ASM	1960 - 2009
Hasnul (2015)	TS	Other	1970 - 2014
Mowlaei & Golkhandan (2015)	P	ASM	1988 - 2012
Biyase & Zwane (2016)	P	Other	1980 - 2005
Compton & Paterson (2016)	P	BM	1988 - 2012
d'Agostino et al. (2016)	P	BM	1996 - 2010
Heo & Ye (2016)	P	ASM	1990 - 2012
Serkan et al. (2016)	P	Other	1998 - 2012
Yildirim & Öcal (2016)	CS	ASM	2000 - 2010
Yolcu Karadam et al. (2016)	P	ASM	1988 - 2012
Augier et al. (2017)	TS/CS	ASM	1952 - 2012
Aziz & Asadullah (2017)	CS/P	BM	1990 - 2013
d'Agostino et al. (2017)	P	BM	1970 - 2014
Sheikh et al. (2017)	TS	ASM	1972 - 2016
Töngür & Elveren (2017)	P	ASM	1988 - 2008
d'Agostino (2018)	P	BM	1998 - 2012
Bayrak (2019)	TS	Other	1990 - 2017
Dunne & Smith (2020)	P	ASM	1960 - 2014
Rahman & Siddiqui (2019)	P	BM	1998 - 2017
d'Agostino et al. (2020)	P	BM	1984 - 2014
Riveros Gavilanes (2020)	P	ASM	1977 - 2016
Becker & Dunne (2021)	P	ASM	1970 - 2019
Nugroho & Purwanti (2021)	P	Other	2002 - 2018
Yolcu Karadam et al. (2021)	P	ASM	1988 - 2019

*CS = Cross-section dataset; TS = Time-series dataset; P = Panel dataset

**ASM = Augmented Solow model; BM = Barro-type endogenous growth model;

SEM = Simultaneous equation model; Other = unspecified theoretical model

Table 6: List of proposed explanatory variables

Variable	Description	Mean	SD
Data characteristics			
Panel (P)	=1 if the dataset used in the primary study is panel	0.724	0.448
Single-country study	=1 if the primary study examines only one country	0.069	0.254
Time span	= time span of the primary study (normalized)	0.314	0.183
No. of observations	= no. of observations within the primary study (normalized)	0.133	0.187
Countries	= no. of countries within the primary study (normalized)	0.285	0.217
Average	=1 if the dependent variable is averaged over a certain no. of years	0.41	0.492
Structural variation			
SIPRI	=1 if the data source of the primary study is Stockholm International Peace Research Institute	0.617	0.487
WB	=1 if the data source of the primary study is the World Bank	0.203	0.402
ACDA	=1 if the data source of the primary study is the Arms Control and Disarmament Agency	0.037	0.189
Africa	=1 if only African countries data is used	0.072	0.258
Asia	=1 if only South East Asian or South Asian countries data is used	0.057	0.232
MENAT	=1 if only Middle-Eastern (incl. Turkey) and North-African countries data is used	0.028	0.354
Developed	=1 if only developed countries data is used	0.215	0.411
LDC	=1 if only less-developed countries data us used	0.244	0.430
D1950	=1 if the primary study contains data from 1950s	0.067	0.250
D1960	=1 if the primary study contains data from 1960s	0.398	0.490
D1970	=1 if the primary study contains data from 1970s	0.489	0.500
D1990	=1 if the primary study contains data from 1990s	0.630	0.484
D2000	=1 if the primary study contains data from 2000s	0.539	0.49

Variable	Description	Mean	SD
NME	=1 if the primary study controls for milex of neighboring countries	0.072	0.258
Education	=1 if the primary study controls for education	0.279	0.449
Natural resources	=1 if the primary study controls for natural resources (oil, coal, etc)	0.124	0.329
Population growth	=1 if the primary study controls for population growth	0.425	0.495
Corruption	=1 if the primary study controls for corruption	0.064	0.245
Conflict	=1 if the primary study controls for external or internal conflicts	0.089	0.285
FDI	=1 if the primary study controls for foreign direct investments	0.086	0.281
Democracy	=1 if the primary study controls for the quality of democracy	0.035	0.183
Estimation characteristics			
ASM	=1 if primary study uses Augmented Solow model	0.269	0.444
BM	=1 if primary study uses Barro-type growth model	0.331	0.471
SEM	=1 if primary study uses Simultaneous equation model	0.059	0.236
OLS	=1 if primary study uses Ordinary least squares	0.361	0.481
GMM	=1 if primary study uses Generalized method of moments	0.136	0.343
GLS	=1 if the primary study uses Generalized least squares	0.044	0.206
FEM	=1 if primary study uses Fixed effects method estimation	0.136	0.343
Endogeneity	=1 if primary study takes endogeneity explicitly into account	0.136	0.343
Squared	=1 if primary study contains square form of milex	0.104	0.305
Interaction	=1 if primary study contains milex interacting with other variables	0.124	0.329
Publication characteristics			
Study age	=no. of years since the study was published (normalized)	0.336	0.293
Citations	=no. of times the primary study has been used as a reference standardized by the no. of years since it was published (normalized)	0.138	0.157

Table 7: Numerical description of primary studies

Author	Number of estimates	Simple mean	Weighted mean
Benoit (1978)	3	0.365	0.383
Deger and Smith (1983)	5	-0.013	0.124
Lim (1983)	12	-0.367	-0.389
Deger and Sen (1983)	1	0.146	0.146
Faini et al. (1984)	1	-0.121	-0.121
Cappelen et al. (1984)	8	0.031	0.027
Landau (1986)	12	-0.058	-0.003
Biswas and Ram (1986)	6	0.132	0.194
Deger (1986)	5	0.333	0.333
Lebovic and Ishaq (1987)	4	-0.151	-0.130
Chan (1988)	3	0.121	0.136
Grobar and Porter (1989)	5	0.216	0.253
Gyimah-Brempong (1989)	2	0.038	0.038
Looney (1993)	2	-0.051	-0.272
Landau (1993)	29	0.005	0.003
Dunne and Mohammed (1995)	3	-0.121	0.001
Lipow and Antinori (1995)	2	0.189	0.193
Blomberg (1996)	1	-0.042	-0.042
Knight et al (1996)	4	-0.035	-0.01
Landau (1996)	11	0.004	0.004
Brumm (1997)	2	0.305	0.316
Antonakis (1997)	2	-0.469	-0.471
Heo and DeRouen (1998)	5	-0.108	-0.111
DeRouen (2000)	1	0.207	0.207
Stroup and Heckelman (2001)	5	-0.007	-0.012
Galvin (2003)	9	-0.141	-0.144
Aizenman and Glick (2006)	7	-0.088	-0.108
Bose et al (2007)	4	0.354	0.409
Kollias et al (2007)	2	0.159	0.16
Yakovlev (2007)	10	-0.158	-0.16
Looney and McNab (2008)	3	-0.104	-0.128
Cooray (2009)	5	0.108	0.109
Pieroni (2009)	6	-0.136	-0.156
Pieroni and d'Agostino (2009)	1	0.068	0.068

Author	Number of estimates	Simple mean	Weighted mean
Aikaeli and Mlamka (2010)	1	-0.344	-0.344
D'Agostino et al (2010)	2	-0.143	-0.143
Heo (2010)	1	-0.347	-0.347
D'Agostino et al (2011)	1	-0.181	-0.181
Kalaitzidakis and Tzouvelekas (2011)	3	0.124	0.01
Iftikhar and Ali (2012)	6	-0.101	-0.107
Hou and Chen (2012)	6	-0.084	-0.096
Na and Bo (2013)	5	-0.368	-0.424
Hou and Chen (2014)	15	-0.067	-0.059
Yildirim and Öcal	4	0.203	0.203
Hasnul (2015)	2	0.046	0.045
Mowlaei, Golkhandan (2015)	3	-0.186	-0.172
Biyase and Zwane (2016)	10	-0.138	-0.141
Compton and Paterson (2016)	48	-0.043	-0.043
D'Agostino et al (2016)	4	-0.122	-0.121
Heo and Ye (2016)	1	-0.810	-0.810
Serkan et al (2016)	4	-0.278	-0.278
Daddi et al (2016)	6	0.191	0.168
Yolcu Karadam et al (2016)	4	-0.143	-0.145
Augier et al (2017)	4	0.236	0.242
Aziz and Asadullah (2017)	12	-0.106	-0.114
D'Agostino et al (2017)	12	-0.132	-0.113
Sheikh et al (2017)	1	0.279	0.279
Töngür and Elveren (2017)	12	-0.077	-0.078
D'Agostino et al (2018)	5	-0.069	-0.67
Bayrak (2019)	1	-0.085	-0.085
Smith and Dunne (2019)	24	-0.094	-0.089
Rahman and Siddiqui (2019)	3	-0.430	-0.515
D'Agostino et al (2020)	8	-0.273	-0.272
Riveros Gavilanes (2020)	3	0.026	0.026
Becker and Dunne (2021)	4	-0.116	-0.126
Nugroho and Purwanti (2021)	1	0.131	0.131
Yolcu Karadam et al (2021)	8	-0.102	-0.131
Total	405	-0.057	-0.072
Total excl. Alptekin & Levine (2012)	233	-0.086	-0.147

Table 8: Results of MMA for the full dataset

Response variable:	Frequentist model averaging		
PCC	Coefficient	SE	t-statistic
Intercept	-0.119	0.088	-1.352
SE (publication bias)	0.229	0.285	0.804
Data characteristics			
Panel	-0.031	0.039	-0.796
Single-country study	0.012	0.064	0.184
No. of observations	-0.213	0.091	-2.348
No. of countries	0.239	0.093	2.580
Time span	0.425	0.192	2.209
Averaged data	-0.004	0.028	-0.150
Structural variation			
SIPRI	-0.014	0.025	-0.571
WB	-0.038	0.048	-0.785
ACDA	0.012	0.041	0.294
Africa	-0.028	0.038	-0.747
Asia	0.032	0.046	0.684
Developed country	0.036	0.034	1.063
LDC	0.045	0.037	1.194
D1950	0.009	0.043	0.217
D1960	-0.060	0.044	-1.366
D1970	-0.064	0.056	-1.145
D1990	-0.092	0.058	-1.598
D2000	-0.032	0.041	-0.780
NME	0.102	0.103	0.994
Education	0.023	0.029	0.766
Natural resources	0.016	0.029	0.538
Population growth	0.017	0.021	0.814
Corruption	-0.044	0.046	-0.949
Conflict	-0.071	0.066	-1.066
FDI	0.042	0.051	0.832
Democracy	0.027	0.046	0.595

Response variable:	Frequentist model averaging		
PCC	Coeffiecient	SE	t-statistic
Estimation characteristics			
ASM	-0.009	0.022	-0.388
BM	0.018	0.030	0.627
SEM	0.008	0.032	0.240
OLS	-0.040	0.041	-0.980
GMM	-0.048	0.050	-0.964
GLS	-0.017	0.031	-0.549
FEM	0.008	0.022	0.371
Endogeneity	0.045	0.040	1.123
Squared	-0.037	0.045	-0.823
Interaction	0.001	0.017	0.050
Publication characteristics			
Citations	0.135	0.127	1.068
Studies	67		
Observations	405		

Appendix B (for online publication): Detailed analysis of the recent studies

B.1 Publication bias and average effect

Just as in the case of the main text, this appendix reports the results of the analysis of the most recent papers (2009 - 2021) in three parts - the numerical summary, publication bias analysis, and heterogeneity analysis.

The fixed-effects mean of the partial correlation coefficients found among the new studies is -0.147 with a 95 percent confidence interval of (-0.152 ; -0.143). Therefore, the effect size of relevant primary studies published between 2009 and 2021 is both statistically and economically significant and classified as small (Doucouliagos, 2011). The Figure 8 provides a visual illustration of the difference between the PCCs within the dataset used by Alptekin and Levine (2012) and the newly published studies.

The funnel plot created from the restricted dataset depicted in Figure 8 does not offer a clear conclusion to the publication bias analysis. Still, it should be commented on the four outliers in the upper left corner of the plot. These estimates resemble both a very negative partial correlation coefficient and a very low standard error. Even though the estimates do not belong to the same study, there is a combination of several factors, which may be responsible. These estimates stem from panel datasets, with a relatively short time span located in the 1990s and 2000s, and having SIPRI as a data source.

The results of linear publication-bias tests for the limited dataset depicted in Table 9, Panel A, show a significant shift of the coefficients in comparison to the whole dataset. The estimate of WLS using precision as weight and between-effects model still presents only a small publication bias. The WLS model using study size as weight and fixed-effects model show substantial publication bias. Given the significant heterogeneity in the study size (see number of estimates in Table 7), the results of the fixed-effects model should be treated with reservations. Therefore, the publication bias is small or moderate. The restricted dataset yields a significantly negative mean beyond bias ranging from -0.179 to -0.130, all of which can be classified as small.

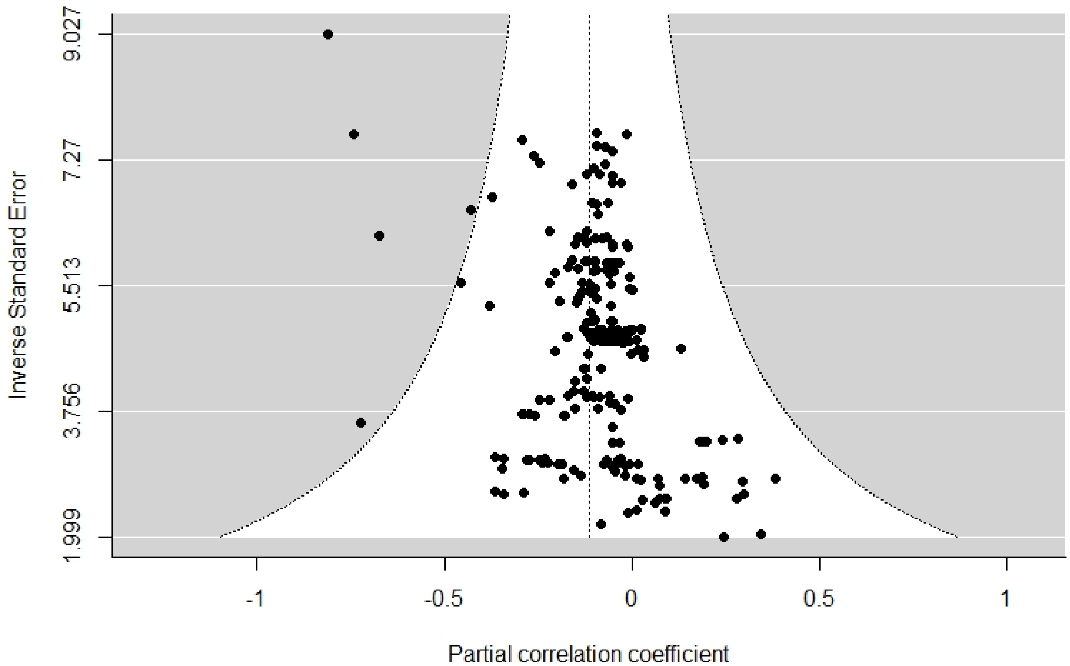

Figure 8: Funnel plot of sizes of estimates against precision measured as an inverse of estimate's standard error (limited dataset using only newly collected estimates)

The non-linear tests in Panel B also show an increase in the magnitude of the mean beyond bias. The estimates also show a larger dispersion ranging from -0.246 indicated by the Endogenous kink model to -0.081 given by the Weighted average of adequately powered. Three out of four estimates again show an economically significant small effect size, while the Stem-based estimate finds itself near the threshold of -0.226 for moderate economic effect as specified by Doucouliagos (2011) and the Endogenous kink surpasses it.

In Panel C, the 2SLS shows a statistically and economically significant mean effect of -0.134 and a small publication bias. The p-uniform* controlling for endogeneity estimates the mean effect size at -0.112, which is statistically significant given a standard error of 0.019.

It also indicates an estimate for the publication bias statistically significant at a 10 percent level with an estimated value of 0.503 and a p-value of 0.078.

Table 9: Publication bias tests results for the restricted dataset (2009 - 2021)

Panel A: Linear models				
	Precision	Study	FE	BE
Publication bias	0.970***	1.135*	-1.459***	0.793***
	(0.278)	(0.480)	(0.135)	(0.226)
Mean beyond bias	-0.141***	-0.179***	-0.161***	-0.130***
	(0.016)	(0.034)	(0.005)	(0.013)
Panel B: Non-linear models				
	Stem-based	WAAP	Selection model	Endogenous kink
Mean beyond bias	-0.212***	-0.081***	-0.109***	-0.246***
	(0.058)	(0.008)	(0.015)	(0.024)
Panel C: Models controlling for endogeneity				
	2SLS	p-uniform*		
Publication bias	0.852***	0.503		
	(0.280)	(p = 0.078)		
Mean beyond bias	-0.134***	-0.112***		
	(0.015)	(0.019)		

The caliper test follows. The histogram of t-statistics does not provide a clear conclusion at the 1.96 thresholds on either side of the histogram depicted in Figure 9. Except for the smallest width caliper, the caliper test (Table 10) indicates the presence of publication bias, placing roughly 77 percent of observed t-statistics beyond the threshold of -1.96 within the limits of the caliper. Still, I approach these results with caution since they come from an analysis using a very limited number of observations. Looking at the results of the caliper test around the threshold of 1.96, these results seem to be completely unusable due to a very low number of observations.

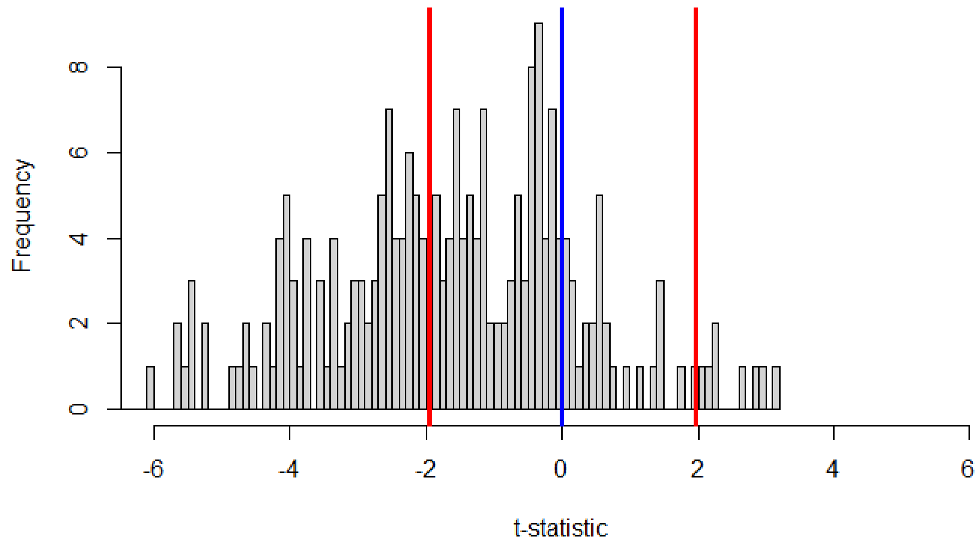

Figure 9: Histogram of t-statistics and vertical lines signifying -1.96, 0, and 1.96 values used as visual analysis before the caliper test (restricted dataset)

Table 10: Caliper test results at t = -1.96 for the restricted dataset (2009 - 2021)

	width = 0.1	width = 0.2	width = 0.3
threshold for t-statistic = -1.96	-0.134	-0.275***	-0.269***
	(0.083)	(0.047)	(0.046)
Observations	8	15	26
threshold for t-statistic = 1.96	0.501***	0.389**	0.411***
	(0.004)	(0.114)	(0.070)
Observations	2	4	6

B.2 Heterogeneity analysis

As exhibited by Figure 10, the list of variables with a posterior inclusion probability high enough now contains only one variable from the data characteristics - the dummy variable for panel datasets with a PIP of 0.975. Nonetheless, there are four variables from the structural variation - dummy variables for the SIPRI data source with a PIP of 0.623, African countries with a PIP of 0.854, the decade of the 1990s with a PIP of 0.941, and education with a PIP of 0.832. The number of years since the study was published, a publication characteristic, has a PIP equal to 0.922. This means that the panel dataset dummy variable is strongly significant. Dummy variables capturing years since publishing, African countries, the 1990s, and education are substantially significant. The only variable that is weakly significant is SIPRI.

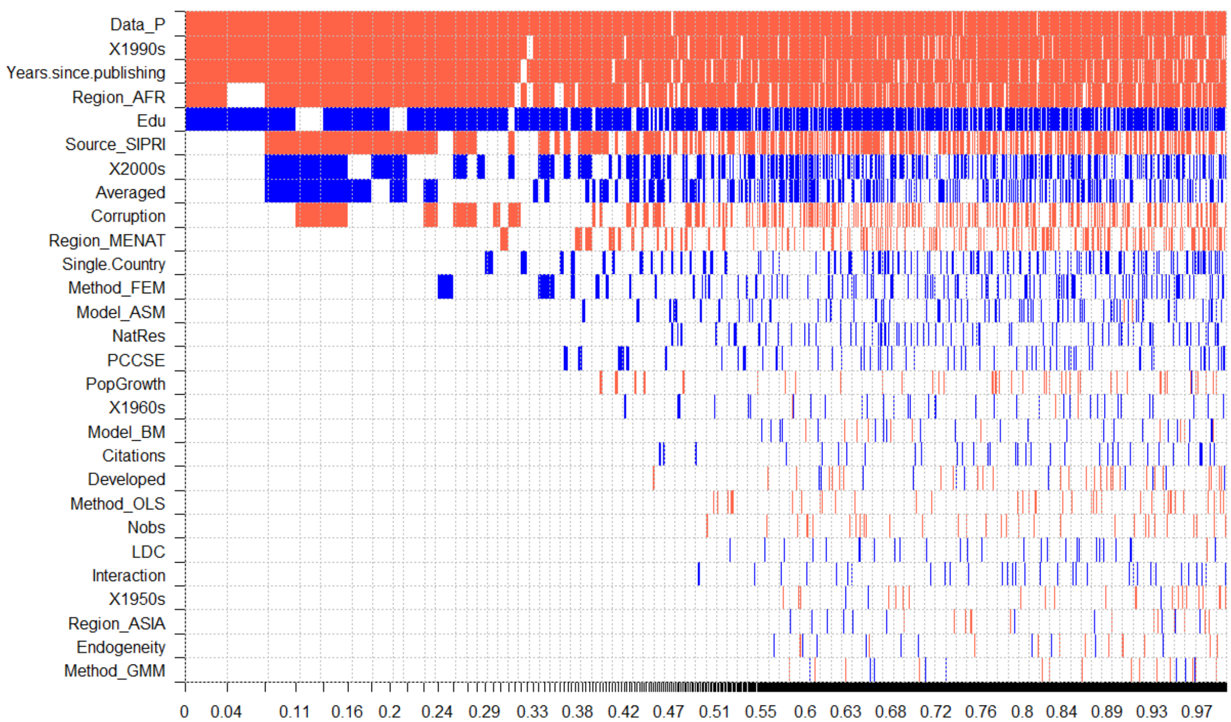

Figure 10: Model inclusion in Bayesian model averaging (restricted dataset)

Table 11: Results of BMA and hybrid check for the restricted dataset (2009 - 2021)

Response variable:	Bayesian model averaging			Frequentist check		
PCC	Post. mean	Post. SE	PIP	Coef.	SE	t-statistic
Intercept	0.193		1.000	0.209	0.039	5.251
SE (publication bias)	0.054	0.212	0.010			
Data characteristics						
Panel	-0.167	0.046	0.975	-0.186	0.023	-7.931
Single-country study	0.019	0.049	0.182			
No. of observations	-0.002	0.013	0.047			
Averaged data	0.026	0.035	0.405			
Structural variation						
SIPRI	-0.045	0.041	0.623	-0.046	0.018	-2.494
Africa	-0.094	0.050	0.854	-0.087	0.027	-3.263
Asia	-0.001	0.010	0.039			
MENAT	-0.013	0.032	0.173			
Developed country	-0.001	0.009	0.053			
LDC	0.001	0.008	0.037			
D1950	-0.001	0.013	0.034			
D1960	0.001	0.010	0.055			
D1990	-0.116	0.050	0.941	-0.111	0.025	-4.475
D2000	0.041	0.051	0.446			
Education	0.061	0.035	0.832	0.072	0.018	3.976
Population growth	-0.001	0.007	0.065			
Corruption	-0.019	0.030	0.334			
Estimation characteristics						
ASM	0.005	0.018	0.110			
BM	0.002	0.012	0.057			
OLS	-0.001	0.005	0.042			
GMM	-0.000	0.004	0.032			
FEM	0.005	0.015	0.142			
Endogeneity	-0.000	0.007	0.071			
Interaction	-0.000	0.005	0.032			
Publication characteristics						
Study age	-0.141	0.058	0.922	-0.151	0.035	-4.357
Citations	0.003	0.019	0.055			
Studies	35			35		
Observations	233			233		

According to Table 11, the publication bias is reduced significantly once other variables are controlled for.

Data characteristics The Table 11 shows a statistically significant and negative coefficient of the panel dataset with respect to the PCCs. Since the second most prevalent data type in the dataset is cross-section data, this result may be interpreted as evidence of the importance of the longitudinal dimension of data as proposed by Stroup and Heckelman (2001).

Structural variation The limited dataset provides many structural variation variables that have a sizable impact on the partial correlation coefficient. First, the choice of data source does play a significant role in the analysis. The choice of SIPRI as a source for the military expenditure data seems to decrease the effect size by 0.046 compared to other data sources. This favors the hypothesis of Lebovic (1999), even though the effect is not large. The Africa dummy variable pushes the PCC down as well. A possible interpretation for this may be that the African countries are devoting too many resources to their armies, because of occurring conflicts. Since the African states gained independence, there have been numerous inter-state conflicts, insurgencies, and coups d'etat, which have remained prevalent to this day. The effect of the decade of the 1990s is negative. The interpretation of this result remains unchanged. Education seems to be a factor, which brings positive effects. This points to the plausibility of a spillover hypothesis, that the education, expertise, and skills received during military service spread into the general economy.

Estimation characteristics Neither the theoretical nor empirical properties seem to play any role in the magnitude of partial correlation coefficients. None of the variables have a PIP higher than 0.150.

Publication characteristics From the publication characteristics, only the study age is significant. On average, the studies published in 2009 have an effect size lower by 0.141 compared to studies published in 2021. This may be because of the inclusion of more recent data. Many argue that many armies of developed countries have gradually received less financing over the last 20 years (NATO, 2022), making their detrimental impact on economic growth less severe.

The MMA results provided in Table 13 provide another frequentist robustness check, which is consistent with the results of the BMA and the hybrid approach. The dummy variables capturing the decade of the 1990s, African countries, and the study age are all statistically significant. The dummy variable capturing the use of the SIPRI data source lost its statistical significance, but it is still statistically significant at a 10 percent significance level. No other variable is statistically significant, which contrasts the previous analysis in case of the variables capturing panel datasets and education.

Table 12: Economic significance of key variables in the restricted dataset (2009 - 2021)

	One-std.-dev. change		Maximum change	
	Effect on PCC	% of FE mean	Effect on PCC	% of FE mean
Panel	-0.060	41 %	-0,167	114 %
SIPRI	-0.022	15 %	-0.045	31 %
Africa	-0.024	17 %	-0.094	64 %
D1990	-0.034	23 %	-0.116	79 %
Education	0.029	20 %	0.061	42 %
Study age	-0.035	24 %	-0.151	103 %

The economic significance of the variables selected from the Bayesian model (Table 12) shows small economic significance in the case of panel datasets, the 1990s and years since the primary study was published. Other variables are economically insignificant according to Doucouliagos (2011). Technically, should we relax the criteria and use the threshold of 0.07 indicated for general economic literature, the effect for African countries would be economically significant as well, and the education variable would approach this threshold as well. The result, however, seems to be informative, more so when it is combined with the coefficient of 0.072 when accounting for the hybrid robustness check.

B.3 Tables

Table 13: Results of MMA for the restricted dataset (2009 - 2021)

Response variable:	Frequentist model averaging		
PCC	Coefficient	SE	t-statistic
Intercept	0.068	0.118	0.576
SE (publication bias)	0.539	0.588	0.917
Data characteristics			
Panel	-0.068	0.057	-1.192
Single-country study	0.118	0.079	1.505
No. of observations	-0.027	0.055	-0.049
Averaged data	0.047	0.037	1.266
Structural variation			
SIPRI	-0.060	0.034	-1.752
Africa	-0.110	0.055	-2.018
Asia	-0.039	0.047	-0.817
MENAT	-0.033	0.042	-0.773
Developed country	-0.048	0.035	-1.354
LDC	0.010	0.034	0.298
D1950	-0.024	0.062	-0.396
D1960	0.004	0.030	0.135
D1990	-0.122	0.057	-2.156
D2000	0.043	0.042	1.018
Education	0.035	0.027	1.314
Population growth	-0.013	0.018	-0.731
Corruption	-0.042	0.031	-1.351
Estimation characteristics			
ASM	0.060	0.055	1.094
BM	0.037	0.041	0.901
OLS	-0.001	0.011	-0.142
GMM	-0.002	0.004	-0.443
FEM	0.023	0.026	0.878
Endogeneity	-0.018	0.026	-0.665
Interaction	0.015	0.023	0.658
Publication characteristics			
Study age	-0.152	0.073	-2.092
Citations	0.055	0.058	0.953
Studies	35		
Observations	233		

IES Working Paper Series

2024

1. Nino Buliskeria, Jaromir Baxa, Tomáš Šestořád: *Uncertain Trends in Economic Policy Uncertainty*
2. Martina Lušková: *The Effect of Face Masks on Covid Transmission: A Meta-Analysis*
3. Jaromir Baxa, Tomáš Šestořád: *How Different are the Alternative Economic Policy Uncertainty Indices? The Case of European Countries.*
4. Sophie Ghvanidze, Soo K. Kang, Milan Ščasný, Jon Henrich Hanf: *Profiling Cannabis Consumption Motivation and Situations as Casual Leisure*
5. Lorena Skufi, Meri Papavangjeli, Adam Gersl: *Migration, Remittances, and Wage-Inflation Spillovers: The Case of Albania*
6. Katarina Gomoryova: *Female Leadership and Financial Performance: A Meta-Analysis*
7. Fisnik Bajrami: *Macroprudential Policies and Dollarisation: Implications for the Financial System and a Cross-Exchange Rate Regime Analysis*
8. Josef Simpart: *Military Expenditure and Economic Growth: A Meta-Analysis*

All papers can be downloaded at: http://ies.fsv.cuni.cz

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz