University of Amsterdam

University of Amsterdam

BSc
Business Analytics

The impact of economic indicators on active military personnel in NATO countries

Author: Joosep Roots
Student ID: 14611201

Supervisor: Elias Dubbeldam

Company: NATO
Company Contact: Emir Karadag

Date: June 2025

Statement of originality

This document is written by Student Joosep Roots, who declares to take full responsibility for the contents of this document. I declare that the text and the work presented in this document are original and that no sources other than those mentioned in the text and its references have been used in creating it. This thesis contains only work done by the author, and no sections are based predominantly on group work. The Faculty of Economics and Business is responsible solely for the supervision of completion of the work, not for the contents.

Abstract

This thesis examines how key socioeconomic indicators — unemployment rate, GDP, defence spending and educational attainment — influence the number of active military personnel in NATO countries. It is motivated by recruitment and retention challenges faced by NATO countries amid increasing geopolitical tensions and shifting defence priorities. The study compiles a panel data set covering the years 2015-2023 across NATO countries and analyses it using a correlation analysis and a fixed-effects regression model. The research controls for country-specific heterogeneity to assess the individual effects of these socioeconomic factors on active military labour supply. The results indicate that a higher share of defence spending in a country's GDP and a country's GDP per capita are associated with increased military personnel, while a rising secondary education rate is related to decreases in military labour supply. This study contributes to a better understanding of military labour dynamics, provides evidence-based insights for NATO policymakers seeking to improve defence planning, and highlights the need for further cross-country, multivariate research in this field.

1. Introduction

Recruitment and retention of military personnel have become an increasingly important issue for many NATO countries due to demographic changes, socioeconomic conditions and geopolitical tensions (NATO, 2022; NATO Research and Technology Organization, 2007). Threats like Russia's invasion of Ukraine emphasise the need for sufficient manpower in NATO members' armed forces. At the same time, according to NATO Research and Technology Organization (2007), reports of large proportions of recruits leaving the military during their first term have not been unusual. Understanding the factors that influence military labour supply is crucial for providing advice to NATO policy makers to improve defence planning.

This thesis addresses the following research question: *How do unemployment rate, GDP, defense spending, and educational attainment affect the number of active military personnel in NATO countries?* It aims to provide empirical results on how broader economic conditions may affect military labour supply and thus support defense policy formulation.

Previous studies have shown that higher unemployment rates are often linked to increased enlistment, for example, in Sweden and the United States, while findings from the Czech Republic have suggested more complex relationships (Asch et al., 2010; Bäckström, 2019; Holcner et al., 2021). GDP and defense spending have been studied as signals of a country's civilian and military economy, which influences the choice of military employment, though results remain mixed (Holcner et al., 2021; Warner & Asch, 1995). Furthermore, educational attainment is seen as both a requirement for modern armed forces seeking highly skilled personnel and a challenge for personnel retention, as better-educated individuals may perceive better opportunities in the civilian sector (CNA's Resources and Force Readiness division, n.d.; Hof et al., 2023).

Crucially, existing research often focuses only on individual countries. These findings are context-specific and not easily applicable across NATO as a whole. Additionally, previous studies tend to analyse factors in isolation rather than as part of a multi-variable system. This approach does not account for the potential combined effects of the variables. As a result, the current knowledge lacks a comprehensive understanding of how multiple variables jointly influence military size. Current literature also provides conflicting findings, leading to uncertainty about the exact effects of various socioeconomic factors. This thesis aims to fill those gaps in the literature by taking a broader approach and empirically analysing the effect of multiple variables across all NATO countries and over several years. It hopes to find generalizable results, which NATO policymakers could use across all NATO countries for enhancing the defence planning strategy.

Following this introduction, Chapter 2 reviews the existing literature for links between socioeconomic conditions and military labour supply, and methods used to study those links. Chapter 3 describes the context of this study, connecting it with NATO's strategy and challenges. Chapter 4 outlines the methodology used for collecting the data and analysing it, while Chapter 5 presents the results of the analysis. Finally, Chapter 6 concludes the findings and presents future research opportunities.

2. Literature review

This literature review explores the research on the influence of socioeconomic factors on military labour supply. Recruitment and retention in the military have been studied using a wide range of different factors. According to a report by the NATO Research and Technology Organization (2007), the recruitment and retention issues in NATO countries are caused by factors such as low unemployment rates, higher civilian salaries, external competition for the labour supply, recruit quality and the shrinking pool of 18-24 year old individuals, among other challenges. This review aims to investigate how GDP, unemployment, education and defence spending impact the military's attractiveness as an employer. It focuses on quantitative studies in NATO countries to identify current knowledge on recruitment and retention, and identify gaps that this thesis will address. First, the effect of unemployment rates will be discussed, then GDP and defence spending and finally educational attainment.

The assertion that low unemployment rates cause recruitment and retention issues is supported by several authors. For example, Bäckström (2019), using a linear fixed-effects regression with panel data on Swedish counties in 2011-2015, found a positive and statistically significant correlation between unemployment and military application rates. Asch et al. (2010) also estimated a linear fixed-effects regression with panel data on U.S states across quarters, controlling for both time and state fixed effects. They similarly found the unemployment rate to be positively and significantly related to high-quality enlistment contracts. These findings align with research by Balcaen and Du Bois (2025) who used a linear mixed-effects regression, which included both fixed and random effects, to account for population and subgroup variation in data across Belgian provinces. This study found that a one percentage point increase in unemployment rate results in a 0.0137 percentage point increase in military application rates. However, some evidence also suggests the opposite, for example Holcner et al. (2021) found an inverse relationship between unemployment and military recruitment in the Czech Armed Forces.

Warner and Asch (1995) describe the decision to enlist and to remain in the military as a choice between employment in the military or the civilian sector. This study mentioned that a perfectly rational individual joins the military sector if the pay differential between the sectors exceeds the preference for civilian life. Warner and Asch (1995) also note that the USA and its allies spend a significant amount of their defence budgets on military personnel. This means that GDP and defence spending, along with their dynamics, could reflect employment opportunities in the civilian and military sectors and, in turn, be correlated with choices to enlist and stay in the military. Although Bäckström (2019) researched mostly unemployment rates, he highlights the fact that a stronger civilian economy increases the difficulty of recruiting new military personnel. On the other hand, Holcner et al. (2021) used a multiple linear regression model to find that the annual increase in GDP, indicating a growing economy, had a positive impact on the recruitment to the Czech Armed Forces. The annual increase in defence expenditure was also found to be correlated with a higher number of military recruits in that study.

The U.S Department of Defense has a benchmark of at least 90% of new military recruits having secondary education or higher (CNA's Resources and Force Readiness division, n.d.). This emphasises the fact that the military is looking for educated recruits. Asoni and Sanandaji (2013) also argue that the transition to a smaller and technologically advanced military has made the recruitment process more selective and less likely for an individual to be allowed in the military without a high school degree. Elster and Flyer (1982) found that the four-year retention rates were higher amongst U.S military recruits with a high-school degree, compared to those with lower educational attainment. Additionally, this study used grouped data in regression analysis to study the attrition rates from the military and found that high school graduates had lower attrition rates than those with lower educational attainment. However, this study was conducted with data from the 1970s, which may not reflect the current socioeconomic or military environment. In contrast, a more recent study by Hof et al. (2023) found that in the Dutch Armed Forces recruits training to become officers, who have a higher level of secondary education before enlistment, show higher intentions to quit basic training, than those training to become noncommissioned officers, who have a lower level of secondary education. This study hypothesised that recruits with a better educational background could believe that they have better opportunities in the civilian labour market. Asch et al. (2010) included a variable in their regression model for the percentage of high school graduates who are enrolled in college, but found no statistically significant relationship with the number of high-quality enlistment contracts, further reinforcing the mixed results on the impact of educational attainment on military personnel.

The reviewed literature predominantly relies on quantitative methods, commonly using panel regression models to estimate the impact of socioeconomic variables on military recruitment and retention. Fixed-effects models are often utilised to account for heterogeneity across geographical units and time, while random-effects models are also employed in some studies. These methods are appropriate for this research, as they can control for temporal and regional variation, allowing researchers to isolate the effects of different factors.

Overall, the literature suggests complex and sometimes even contradictory relationships between socioeconomic factors and military recruitment and retention. Many studies find unemployment rates to be positively correlated with enlistment, implying that individuals are more likely to join the military when civilian employment opportunities are scarce, while others have also reported the opposite. Similarly, although economic theory suggests that stronger civilian economies reduce the appeal of military employment while defence spending increases it, the empirical evidence remains inconsistent. This could indicate the need to rethink these assumptions or consider additional factors. Education also exhibits mixed results, with higher educational attainment improving eligibility for the military, while also potentially increasing attrition, as better-educated recruits may have better civilian opportunities. Socioeconomic conditions and military structures differ between nations and over time, which may lead to these inconsistencies. The mixed findings, along with most studies focusing on single countries or isolated variables or outdated data, underscore the need for comparative research across NATO countries to provide generalisable and consistent insights into military labour supply dynamics.

3. Case background

Established in 1949, NATO (North Atlantic Treaty Organization) is a security alliance, currently comprising 32 member countries from North America and Europe (NATO, n.d. U.S. Mission to NATO, n.d.). Its primary mission is to ensure the collective freedom and security of its members through political and military means (U.S. Mission to NATO, n.d.). Among others, NATO currently faces threats such as Russia's invasion of Ukraine, China's growing ambitions and conflicts in the Middle East and Africa, highlighting the need to further strengthen its deterrence and defence capabilities (NATO, 2022).

The events of September 11, 2001, led to a pivotal shift in global security. They altered the nature of perceived threats and military engagements. In response, most European countries redefined their military structures, abolishing large conscript armies in favour of all-voluntary active forces. This change led to the professionalisation of the military in many countries (Herranen, 2004). With the end of conscription in numerous countries, the size of the active military force became a critical measure of defence capacity, which is why the size of active armed forces is the key focus of this thesis.

It has become increasingly difficult in many NATO countries to recruit new and retain existing qualified military personnel. It is not uncommon for more than 30% of military recruits to leave on their first term of service, especially in specialist roles that are hard and costly to recruit and train. This can be attributed to a variety of factors, including socioeconomic conditions (NATO Research and Technology Organization, 2007). In light of that, this thesis explores how a selection of economic indicators influences the number of active military personnel in NATO countries.

The indicators investigated in this thesis were the unemployment rate, GDP, defence spending dynamics, and educational attainment. The data was collected from the World Bank (n.d.) and the International Institute for Strategic Studies (n.d.) for years 2015-2023 across all NATO countries. Russia was also included in the dataset to gain insight into its recruitment success during the invasion of Ukraine.

This study is relevant for NATO defence policymakers and strategic planners as it identifies how these socioeconomic indicators influence active military size. The findings of this thesis can support data-driven decision-making for recruitment and retention policies. As NATO continues to adapt to new geopolitical circumstances, understanding the socioeconomic drivers of military personnel supply is valuable for maintaining operational readiness.

4. Methodology

4.1 Data preparation

The data preparation process involved several steps to create a clean and structured dataset that could be used for analysis. The study combines military and socioeconomic data for NATO countries across multiple years. Data was collected from various sources, transformed into a consistent format and merged into a single dataset.

First, the military personnel and defence spending data were acquired from different issues of The Military Balance by the International Institute for Strategic Studies (n.d.) as each issue contained data for a specific year. Data for active armed forces numbers, defence spending per capita and defence spending's share of GDP were selected for NATO member countries from each table. To ensure that the data is comparable, the effect of inflation was removed from defence spending per capita data by adjusting the values to 2015 USD using the Consumer Price Index (CPI) of different years. The CPI data was sourced from Federal Reserve Bank of Minneapolis (n.d.). The tables from different years were then merged into a single long-format table, where each row represented observations for a specific country in a specific year, as this format is suitable for panel data analysis.

Next, the unemployment rate, GDP per capita in constant 2015 USD, population and secondary educational attainment data were collected from the open data provided by the World Bank (n.d.). Data was again filtered to include only NATO countries from 2015 to 2023. The educational attainment data had missing values that were handled using linear interpolation to estimate missing values based on neighbouring data, and backward/forward filling for data points at the start or end of the series, which could not be interpolated. The tables were then also transformed into a long format, so that they could be merged with the military personnel and defence spending data.

The military and economic data were then merged into a complete dataset. The active armed forces per capita column was created by dividing the active armed forces by population numbers. Additional columns for annual changes in GDP per capita, defence spending per capita and defence spending's share of GDP were calculated by finding their percentage changes from the previous year's value. Table 1 summarises the variables of the prepared dataset.

Table 1 *Variable Descriptions*

Variable	**Description**	**Source**	**Unit**	**Transformation**
Active armed forces per capita	Number of active military personnel per capita	Military Balance, World Bank	per capita	log()
Unemployment rate	National unemployment rate	World Bank	% (0-100)	none
Secondary education rate	Proportion of population (25+) with secondary education	World Bank	% (0-100)	none
GDP per capita	Gross Domestic Product per capita	World Bank	2015 USD	log()
Defence spending per capita	A country's defence expenditure divided by population	Military Balance	2015 USD	log()
Defence spending % of GDP	Defence spending's share of GDP	Military Balance	%	none
GDP per capita % change	Annual percentage change in GDP per capita	World Bank (own calculation)	%	none
Defence spending per capita % change	Annual percentage change in defence spending per capita	Military Balance (own calculation)	%	none
Defence spending % GDP % change	Annual percentage change in defence spending's share of GDP	Military Balance (own calculation)	%	none

Finally, a logarithmic transformation was applied to variables active armed forces per capita, GDP per capita and defence spending per capita. These variables can have skewed distributions due to disproportionate influence from larger countries. Logging can mitigate this skewness, stabilise variance and allow for interpretation of the regression coefficients in terms of elasticities. Other variables were not logarithmically transformed as they were expressed as percentages; therefore, logging would make interpretation difficult and distort the scales.

4.2 Data understanding

Table 2 reports the descriptive statistics for the variables used. The logged dependent variable, armed forces per capita, showed relatively low variation with a standard deviation of 0.44 and a range of values between -6.55 and -4.30. This could limit the explanatory power of the regression; however, the model might still provide insights into smaller changes in active military size per capita.

Table 2 *Descriptive statistics of variables*

	Armed Forces per cap.	GDP per cap.	Def. spend. per cap.	Unemployment rate	Secondary education rate	Def. spend. % GDP	GDP per cap. % change	Def. spend. per cap. % change	Def. spend. % GDP % change
mean	-5.71	7.86	74.50	10.03	5.71	1.61	2.18	4.82	4.57
std	0.44	4.58	14.64	0.76	0.81	0.66	3.67	18.73	16.39
min	-6.55	2.02	33.60	8.28	3.44	0.35	-15.21	-41.71	-41.03
25%	-5.96	4.83	69.07	9.52	5.26	1.14	0.76	-5.61	-3.59
50%	-5.79	6.54	77.78	9.93	5.81	1.44	2.27	1.89	1.48
75%	-5.46	9.41	86.00	10.70	6.29	1.98	4.37	10.16	10.67
max	-4.30	26.40	95.29	11.59	7.60	3.82	13.65	100.83	109.23

Defence spending per capita, its annual change and the annual change in defence spending's share of GDP, however, showed higher volatility, with standard deviations of 14.64, 18.73, and 16.39, respectively. These higher standard deviations could suggest heteroskedasticity. In addition to that, in panel data settings, standard errors are often correlated within entities (in this case, countries), which means assuming independent standard errors could inflate t-statistics and potentially overstate statistical significance in regressions. To account for this, a fixed-effects regression could be estimated using standard errors clustered at the country level.

The use of imputed educational attainment values was also investigated. There were a total of 23 interpolated or filled values across a total of 285 observations. No country had more than 2 imputed values in the timeframe. Given the small scale of imputed values, it was unlikely that a strong bias would be introduced. Nevertheless, a sensitivity analysis should be conducted to analyse the effect of imputation more in depth.

Outliers were identified using both Z-score and Interquartile Range (IQR) methods, which are commonly used in literature for outlier detection. The Z-score method calculates the distance of a data point from the mean in terms of standard deviations, using the formula:

$$Z = \frac{X - \mu}{\sigma}$$

where X is the data point, μ is the sample mean and σ is the sample standard deviation. A common threshold of ± 3 was used to identify values in approximately the outer 0.3% of normally distributed data. The IQR method is non-parametric and less dependent on distributional assumptions. The outliers were defined as values, which fall below $Q1 - 1.5 \cdot IQR$ or above $Q3 + 1.5 \cdot IQR$, where $Q1$ and $Q3$ are the first and third quartiles, reported in Table 2 as 25% and 75%, and $IQR = Q3 - Q1$. This approach helps identify skewed or non-normally distributed outliers. Comparing outliers to the complete dataset revealed no unrealistic values. Since the outliers appeared to be the result of natural variability, not data quality issues, they were not excluded from the dataset. (Barnett & Lewis, 1994)

4.3 Correlation analysis

This section describes the approach used to evaluate correlations and multicollinearity among the socioeconomic variables before estimating the regression model. Specifically, it outlines how Pearson correlation coefficients and Variance Inflation Factors (VIFs) were used to assess the relationships and potential collinearity among the independent variables and how the data were transformed to be consistent with the fixed-effects framework.

The correlation and multicollinearity analysis were performed on demeaned data, consistent with the fixed-effects regression. The fixed-effects model relies on within-country variation over time by subtracting country-specific means, so the data were transformed similarly. For each observation, the mean of the variable across all years of the corresponding country was subtracted from the original value. This demeaning process makes the data comparable across countries as the variables only reflect deviations from each country's average.

The correlation analysis involved calculating the Pearson correlation coefficient for each of the variables, resulting in a correlation matrix, which can be used to assess the direction and strength of bivariate relationships. The Pearson coefficient, denoted by r, is defined as:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

where x_i and y_i are observations of variables X and Y, and $\bar{x}$, $\bar{y}$ are their sample means. The coefficient values range from -1 to 1, where absolute values close to 1 indicate strong relationships and values near 0 suggest weak or no correlations (Cohen et al., 2003).

Variance Inflation Factors (VIF) were calculated for the independent variables to assess the potential presence of multicollinearity. VIFs indicate how much the variance of a regression coefficient is inflated due to relationships with other predictors. Multicollinearity can make estimates unstable and cause interpretability problems. The VIF for a variable X_i is calculated as:

$$VIF_i = \frac{1}{1 - R_i^2}$$

where R_i^2 is the coefficient of determination obrained by regressing X_i on all other independent variables. It is typically considered that VIF values from 1 to 5 are moderate and mostly acceptable, while values over 5, or more conservatively 10, can be potentially problematic (O'Brien, 2007).

4.4 Modelling

This section describes the core regression strategy used to estimate the relationships between socioeconomic variables and the number of active military personnel. It begins by motivating the use of a fixed-effects regression and outlines the base model specification. It then explains how country and year fixed effects were incorporated to control for unobserved heterogeneity and finally it presents how the robustness and sensitivity checks performed.

Studies such as Asch et al. (2010), Bäckström (2019) and Balcaen and Du Bois (2025) have demonstrated the suitability of a fixed-effects regression model for analysing panel data in the context of military recruitment and retention, which informed the methodological choice of this thesis. The following equation was used to estimate the base model:

$$
\begin{aligned}
log(ArmedForces_{it}) = {} & \beta_1 \cdot Unemployment_{it} + \beta_2 \cdot Education_{it} \\
& + \beta_3 \cdot log(GDPPerCap_{it}) + \beta_4 \cdot log(DefSpendingPerCap_{it}) \\
& + \beta_5 \cdot DefSpending\%GDP_{it} + \beta_6 \cdot \Delta GDPPerCap_{it} \\
& + \beta_7 \cdot \Delta DefSpendingPerCap_{it} + \beta_8 \cdot \Delta DefSpending\%GDP_{it} \\
& + \alpha_i + \gamma_t + \varepsilon_{it}
\end{aligned}
$$

where i denotes the country, t the time, Δ the annual change, a_i is the country-specific fixed effect and γ_t is the year fixed effect.

The regression analysis used a fixed-effects panel regression model to control for time-invariant heterogeneity across countries. This meant that it was not necessary to use the demeaned data anymore, because the model already took differences between entities into account. The regression model used country fixed effects to isolate the within-country variation over time, enabling more accurate estimation of the conditional effects of economic factors on military personnel levels. In other words, by introducing country fixed effects, the model was estimated on changes within countries, rather than between countries (Bäckström, 2019).

Time fixed effects were also included in the model to control for unobserved heterogeneity over time. Given global shocks, such as the COVID-19 pandemic and Russia's invasion of Ukraine, which could have affected all countries similarly, year fixed effects were necessary to avoid omitted variable bias and eliminate aggregate trends in the data (Bäckström, 2019). For comparison, a model with only country fixed effects and no time fixed effects is reported in Table 9 and Table 10 in the *Appendix*. It can be used to help assess the impact of including time fixed effects.

The regression models used clustered country-level standard errors to account for potential within-country correlation of error terms and heteroskedasticity across countries. In panel data, observations from the same entity are often not independent, which violates classical OLS assumptions. Clustering allows for arbitrary correlation within countries, ensuring a more robust and conservative model. While common trends may affect all countries in a given year, clustering at the time-level was not used, as the data has a limited number of time periods, which could lead to unstable clustering or an overly conservative model.

A robustness check and a sensitivity analysis were conducted to further specify the model. The robustness check included estimating two models, one with full variables and one with reduced variables, based on the results of the correlation analysis. The models were compared to assess the impact of removing multicollinear variables. The sensitivity analysis was used to determine whether to include interpolated and filled educational attainment data in the regression model. Three regression models were created: Model A with complete data, including observations with interpolated educational attainment values, Model B that excluded rows with interpolated and filled values, and Model C with complete data and an additional dummy variable indicating whether the education data was interpolated or not. The models were again compared to choose the most appropriate specification.

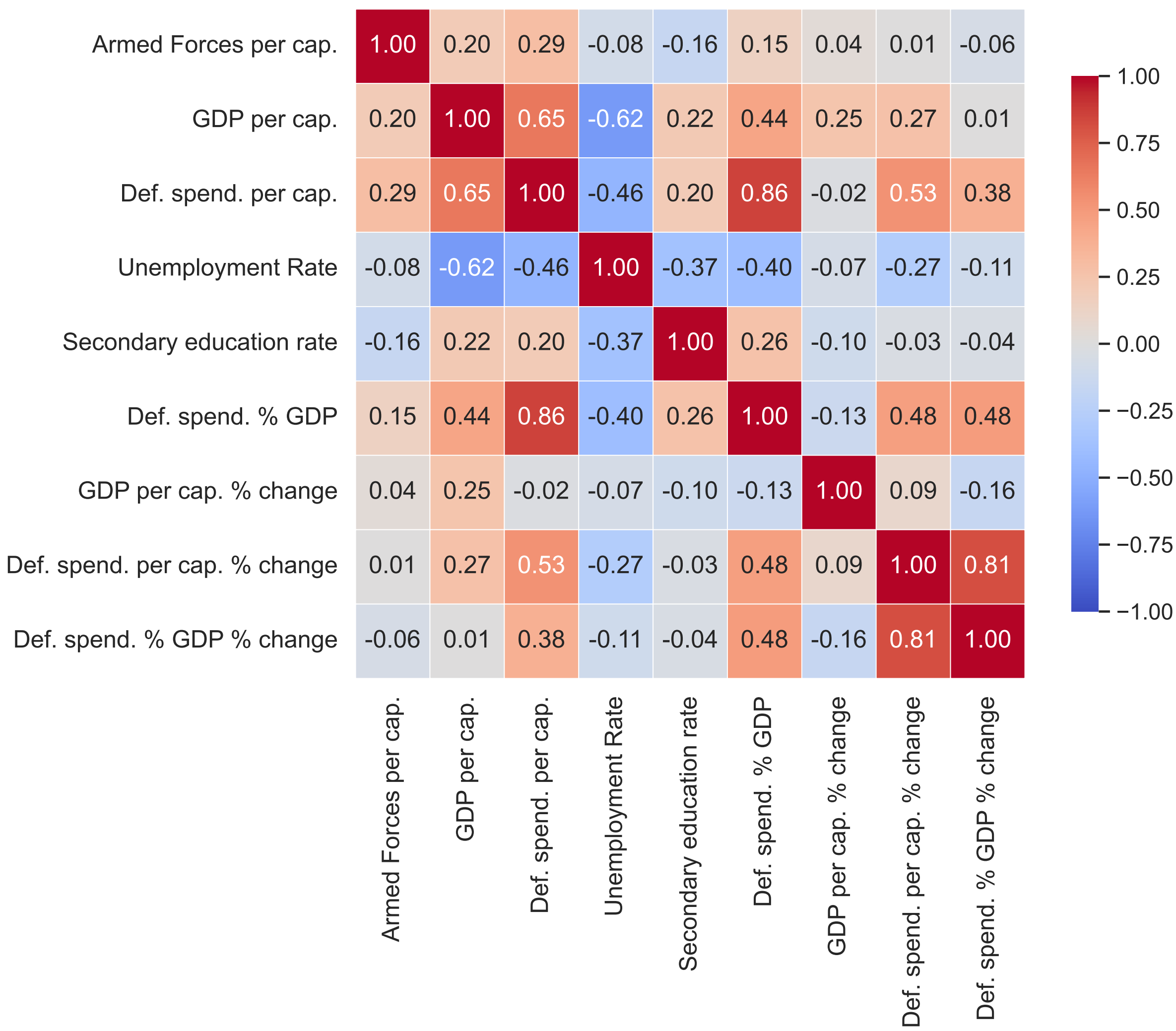
5. Results

5.1 Correlation analysis

5.1.1 Bivariate correlations

Figure 1

Correlation matrix



As shown in Figure 1, the correlation analysis revealed no single factor with a strong correlation to active military personnel numbers. Active armed forces per capita were found to have weak to moderate positive correlations with GDP per capita (0.17), defence spending proportion of GDP (0.14) and defence spending per capita (0.29). Weak negative correlations were observed with unemployment rate (-0.08), secondary education attainment rate (-0.12) and defence spending proportion of GDP annual change (-0.11). These findings suggest that military size may grow as a country's economy improves and defence spending rises, while higher unemployment and education attainment rates may be related to reductions in military personnel.

GDP per capita and unemployment rate exhibited a moderate negative correlation (-0.62), which could indicate that a stronger economy is associated with better labour market conditions. Even stronger correlations were found between defence expenditure per capita and defence expenditure's proportion of GDP (0.85), and between their annual changes (0.80). This was unsurprising given that these variables are derived from defence spending. However, such large correlations may indicate multicollinearity, which could bias regression estimates and inflate standard errors. This meant that multicollinearity should be assessed before running the regression.

While the correlation analysis provides useful preliminary insight into the relationships between the variables, it does not isolate the direct effect of each variable. Therefore, a fixed-effects regression is used to estimate the conditional effects of the economic variables on active armed forces per capita, controlling for unobserved time-invariant heterogeneity across countries.

5.1.2 Multicollinearity

The initial results, displayed in Table 3, showed moderate to high VIFs for defence spending per capita (6.00), defence spending as a percentage of GDP (4.93) and their respective annual percentage changes (4.28 and 4.43), indicating the presence of multicollinearity. This is expected as these variables were highly correlated and capture similar underlying dynamics. To mitigate this, the variables defence spending per capita and its annual change were removed, and VIF values were calculated again. It should be noted that a high VIF does not always justify the elimination of a variable; the exclusion should be theoretically motivated, for example, if two variables measure the same under-

Table 3 *Variance Inflation Factors (all variables)*

Variable	VIF
Unemployment rate	1.95
Secondary education rate	1.27
GDP per capita	2.73
Defence spending per capita	6.00
Defence spending % of GDP	4.93
GDP per capita % change	1.35
Defence spending per capita % change	4.28
Defence spending % GDP % change	4.43

lying factor (O'Brien, 2007). In this case, the exclusion makes sense, as the two defence spending variables and their annual changes are conceptually similar.

Table 4 reports the VIF values, excluding the possibly multicollinear variables. The VIF values reduced significantly, with defence spending's share of GDP now having 1.93 and its annual change having 1.50. The observed reduction in multicollinearity motivated the estimation of two regression models: a full model including all variables and a reduced model excluding the highly collinear variables. These models can then be used for a robustness check to ensure whether the findings of the models are stable or influenced by multicollinearity.

Table 4 *Variance Inflation Factors (reduced variables)*

Variable	VIF
Unemployment rate	1.89
Secondary education rate	1.23
GDP per capita	2.10
Defence spending % of GDP	1.93
GDP per capita % change	1.19
Defence spending % GDP % change	1.50

5.2 Regression modelling

5.2.1 Robustness check

Two fixed-effects regression models were estimated to assess the robustness of the model; Table 5 reports their summaries. The first was a base model, which included all variables, and the second was a reduced model with defence budget per capita and its annual change excluded, as these variables had exhibited multicollinearity with the defence budget's proportion of GDP and its annual change.

Table 5 *Robustness check*

Term	Base model	Reduced model
Coefficient estimates		
Unemployment rate	$-0.0108, p = 0.3263$	$-0.0097, p = 0.4007$
Secondary education attainment rate	$-0.0062, p = 0.0477$	$-0.0070, p = 0.0408$
GDP per capita	$0.3754, p = 0.1189$	$0.8725, p = 0.0033$
Def. spend. per capita	$0.3635, p = 0.0003$	-
Def. spend. % GDP	$-0.0984, p = 0.0898$	$0.0918, p = 0.0442$
GDP per capita % change	$-0.0032, p = 0.4352$	$-0.0073, p = 0.0989$
Def. spend. per capita % change	$0.0007, p = 0.5899$	-
Def. spend. % GDP % change	$-0.0019, p = 0.1055$	$-0.0011, p = 0.1599$
Model statistics		
R^2	0.2820	0.2038
F-statistic (robust)	$F(8, 237) = 4.35, p = 0.0001$	$F(6, 239) = 2.92, p = 0.0091$

The base model explained a larger proportion of the variance after accounting for country and year fixed effects ($R^2 = 0.2820$ vs. 0.2038) and included a statistically significant variable, defence spending per capita, which had a positive relationship with the dependent variable. However, the multicollinearity between defence spending per capita and defence spending's share of GDP could undermine the stability and interpretability of the model. This was evidenced by the change in the coefficient sign of defence spending's share of GDP from negative to positive. That variable was expected to exhibit effects in the same direction as defence spending per capita, as they are derived from the same underlying measure of defence budgets. The concerns of multicollinearity distorting the model were further supported by the change in significance of GDP per capita, its annual change, and defence spending's share of GDP. These inconsistencies supported the decision to remove the collinear variables and favour the reduced model, which could be more stable and interpretable.

5.2.2 Sensitivity analysis

Table 6 shows the three models that were estimated to analyse the effect of interpolated and filled educational attainment values. The secondary educational attainment rate coefficients stayed consistently negative across models, although in Model B, it became statistically insignificant. Additionally, other coefficients also remained similar in magnitude, indicating that the use of interpolated values did not meaningfully distort the relationship with the target variable. This was further supported by the fact that the dummy variable for interpolated values in Model C was not statistically significant, suggesting no systematic differences in observations, where the education rate was imputed. However, Model B showed shifts in significance for other variables, such as defence spending's share of GDP and its annual change, and the overall model lost statistical significance ($F(6, 216) = 2.0229$, $p = 0.0638$).

Based on these results, Model C with full data, including interpolated and filled values, and a dummy for flagging imputed values, was chosen as the preferred regression model. While imputed data could introduce noise or subtle bias into regression estimates, this did not appear to be the case this time, as the interpolation dummy variable was found to be statistically insignificant. Nevertheless, Model C was adopted as it retains the full sample size while explicitly controlling for data quality variation, thereby increasing the robustness of the model. Furthermore, Model B, which excluded imputed data, was overall statistically insignificant, again supporting the choice of the full sample and appropriate controls.

Table 6 *Sensitivity analysis models*

Term	**Model A (full)**	**Model B (excl. filled)**	**Model C (full + dummy)**
Coefficient estimates			
Unemployment rate	$-0.0097, p = 0.4007$	$-0.0111, p = 0.3499$	$-0.0097, p = 0.3975$
Secondary education attainment rate	$-0.0070, p = 0.0408$	$-0.0052, p = 0.1290$	$-0.0070, p = 0.0417$
GDP per capita	$0.8725, p = 0.0033$	$0.8047, p = 0.0066$	$0.8719, p = 0.0034$
Def. spend. % GDP	$0.0918, p = 0.0442$	$0.1058, p = 0.0521$	$0.0915, p = 0.0438$
GDP per capita % change	$-0.0073, p = 0.0989$	$-0.0072, p = 0.1098$	$-0.0072, p = 0.1044$
Def. spend. % GDP % change	$-0.0011, p = 0.1599$	$-0.0017, p = 0.0321$	$-0.0011, p = 0.1592$
Interpolation dummy	-	-	$0.0049, p = 0.8935$
Model statistics			
Number of observations	285	262	285
R^2	0.2038	0.2000	0.2039
F-statistic (robust)	$F(6, 239) = 2.9221, p = 0.0091$	$F(6, 216) = 2.0229, p = 0.0638$	$F(7, 238) = 3.7099, p = 0.0008$

5.2.3 Regression model analysis

The summary of the selected model specification is reported in Table 7. The selected fixed-effects regression model had a 0.2039 R^2, meaning the independent variables explained approximately 20.39% of the variation in the logarithm of active armed forces per capita after accounting for fixed effects. However, the R^2 before removing fixed effects was approximately 0.9555, meaning in total the model explained about 95.55% of the variation in the dependent variable. The robust F-statistic for the model was $F(7, 238) = 3.7099, p = 0.0008$, showing joint significance of the independent variables, while adjusting for potential heteroskedasticity. In other words, after accounting for country and year fixed effects, the included variables jointly have a significant effect on the dependent variable.

Table 7 *Regression Estimation Summary*

Dep. Variable	`Active Armed Forces per capita`
Estimator	PanelOLS (from Python library "linearmodels")
No. Observations	285
Entities	32
Time Periods	9
Cov. Estimator	Clustered
R-squared	0.2039
Log-likelihood	274.78
F-statistic (robust)	3.7099
P-value (F-stat)	0.0008
Distribution	F(7, 238)
F-test for Poolability	60.604
P-value	0.0000
Distribution	F(39, 238)
Included effects	Entity, Time

Table 8 reports the parameter estimates for the selected model specification. Unemployment rate proved to be statistically insignificant ($-0.0097, p = 0.3975$), indicating that it may not substantially affect military labour supply in the sampled country and year subset. Secondary education attainment rate exhibited a significant, modest negative relationship ($-0.0070, p = 0.0417$) with the dependent variable, meaning for every 1 percentage point increase in secondary education attainment rate, the active military size per capita decreased by approximately 0.70%. This negative relationship could be caused by multiple reasons. For example, more educated individuals may consider their opportunities better in the civilian sector, while it could also be that countries with higher educational attainment rates have more technologically advanced armed forces, which require less manpower.

The logarithm of GDP per capita had a positive and significant relationship ($0.8719, p = 0.0034$) with armed forces per capita, meaning a 1% increase in GDP per capita was associated with approximately a 0.87% increase in active military size per capita. This means that wealthier countries tend to have larger active militaries, which

Table 8 *Parameter Estimates*

Variable	Coef.	Std. Err.	t-stat	p-value	CI Lower	CI Upper
Unemployment rate	-0.0097	0.0115	-0.8476	0.3975	-0.0323	0.0129
Secondary education attainment rate	-0.0070	0.0034	-2.0480	0.0417	-0.0137	-0.0003
GDP per capita	0.8719	0.2950	2.9554	0.0034	0.2907	1.4532
Def. spend. % GDP	0.0915	0.0451	2.0268	0.0438	0.0026	0.1804
GDP per capita % change	-0.0072	0.0044	-1.6302	0.1044	-0.0160	0.0015
Def. spend. % GDP % change	-0.0011	0.0008	-1.4120	0.1592	-0.0027	0.0004
Education Dummy	0.0049	0.0368	0.1340	0.8935	-0.0675	0.0774

could be due to their ability to invest more in the defence sector. Wealthier countries may also want to maintain a larger military presence to project power and influence globally. The annual change in GDP per capita was found to be statistically insignificant ($-0.0072, p = 0.1044$), indicating that short-term economic changes might not influence active military personnel size.

Defence spending's share of GDP, however, exhibited a significant and positive relationship ($0.0915, p = 0.0438$), meaning that for each percentage point increase in defence spending's proportion of GDP, active military size per capita increased approximately 9.15%. The annual changes in defence spending's proportion of GDP proved to be an insignificant predictor of active armed forces numbers ($-0.0011, p = 0.1592$). These findings could indicate that a higher baseline proportion of defence spending in GDP may signal long-term commitments to maintaining a large military, while short-term changes in the proportion may not directly predict a change in active military personnel. It could also be that short-term investments in defence spending were used for assets other than manpower.

It is important to note for the interpretation of these coefficients that the dependent variable was active military size per capita, meaning the coefficients reflect relative changes in active military personnel with respect to population size. Assuming a stable population, the reported percentage changes in active military size per capita can also be interpreted as approximate percentage changes in total active military size. However, in cases where the population can vary notably, the per capita measure provides a more consistent comparison.

6. Conclusion

This thesis set out to investigate how socioeconomic factors like unemployment rate, educational attainment, GDP and defence spending influence active military personnel size in NATO member countries from 2015 to 2023. The analysis was conducted on a panel dataset, using a fixed-effects regression model. Several findings were discovered for academic research and defence policy.

One of the most important and statistically significant findings was that a higher share of defence spending in a country's GDP is associated with a larger active armed forces size per capita. This is particularly valuable as the share of defence spending in GDP is currently a relevant and widely discussed metric in the context of rising geopolitical tensions and shifting defence priorities. In 2014, NATO Heads of State and Government committed to meeting a threshold of 2% defence spending in GDP, and in 2025, all NATO countries are expected to meet or exceed this target (NATO, 2025). According to the findings, this increase could also result in larger active armed forces.

Secondary education attainment rates were significantly and negatively related to active military size per capita. This finding could mean that the military labour supply may decrease as populations become more educated. The results align with Hof et al. (2023), who found that better-educated military recruits have higher intentions to quit basic training. This negative relationship may reflect, for example, better civilian opportunities for more educated individuals or a shift toward more technology-oriented, rather than manpower-intensive armed forces.

Contrary to some prior studies, the unemployment rate did not exhibit statistically significant relationships with armed forces personnel. Military service is often regarded as a stable employment opportunity, especially in bad labour market conditions; however, this study found no evidence of labour market dynamics influencing military size. In contrast, GDP per capita, which can indicate better labour market conditions, as it showed a fairly strong negative correlation with unemployment rates, was found to be statistically significant and positively related to active military size. Other than labour market conditions, this finding may also reflect that wealthier countries can afford to maintain a larger military force, because they have more resources available for defence spending.

The annual changes in GDP per capita and defence spending's share in GDP were found to be insignificant predictors of active military size. Short-term changes in GDP per capita may be too volatile to have direct effects on active military personnel, as defence planning could follow more long-term frameworks. The insignificance of annual changes in defence spending's proportion of GDP, contrasted by the significance of its overall level, may reflect that a sustained commitment to higher defence spending is necessary to grow military manpower, while short-term fluctuations may fund investments in other assets and therefore have limited immediate impact on military size.

For defence policy recommendations, the current study suggests keeping priority on sustained investments in the defence sector, as long-term commitments to a larger proportion of defence spending in GDP had a positive impact on military size. Additionally, GDP per capita had a significant positive relationship with armed forces per capita, highlighting the role of broader economic development in supporting military capacity. This suggests that defence planning should account for rising national income in supporting the maintenance of larger militaries. It also recommends considering and addressing the reasons why educational attainment might negatively impact the military labour supply. New recruitment and retention strategies could be adapted to appeal to more educated populations.

A limitation of this study is that the independent variables themselves only explained about 20% of the variation in active armed forces per capita with the fixed effects removed, while including fixed effects explained around 96%. This study aimed to investigate the relationships among the selected variables on active military size, which it successfully did; however, the variables themselves, without the fixed effects, do not predict a large portion of the variation in active armed forces per capita. For future studies looking to create a better prediction model, more variables, and for example, the effect of defence planning policy or the technological evolution of militaries should be taken into account, as personnel size may not depend only on socioeconomic factors. Another limitation is the sample size. According to Harrell (2015), in order to estimate a reliable multivariate regression, the data should have at least 10-20 observations per estimated parameter, which this model did. However, to increase the statistical power of a fixed-effects regression, future research could collect more observations per country or include additional countries for more generalizable results. An additional future research opportunity might be investigating the reasons why exactly educational attainment has a negative correlation to active military size, as this study does not reveal the causal effects.

Bibliography

Asch, B. J., Heaton, P., Hosek, J., Martorell, F., Simon, C., & Warner, J. T. (Eds.). (2010). *Cash incentives and military enlistment, attrition, and reenlistment*. Rand Corp. https://www.rand.org/content/dam/rand/pubs/monographs/2010/RAND_MG950.pdf

Asoni, A., & Sanandaji, T. (2013). Rich man's war, poor man's fight? Socioeconomic representativeness in the modern military. *IFN Working Paper No. 965*. https://ssrn.com/abstract=2542143

Bäckström, P. (2019). Are economic upturns bad for military recruitment? A study on Swedish regional data 2011–2015. *Defence and Peace Economics*, *30*(7), 813–829. https://doi.org/10.1080/10242694.2018.1522572

Balcaen, P., & Du Bois, C. (2025). Unemployment and military labour supply: A study on Belgian data for the period 2005-2020. *Defence and Peace Economics*, *36*(1), 20–35. https://doi.org/10.1080/10242694.2023.2252653

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (3rd Edition). Wiley. https://www.wiley.com/en-us/Outliers+in+Statistical+Data%2C+3rd+Edition-p-9780471930945

CNA's Resources and Force Readiness division. (n.d.). *Fiscal year 2019 summary report*. CNA. Retrieved May 22, 2025, from https://www.cna.org/pop-rep/2019/summary/summary.pdf

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Bivariate correlation and regression. In *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Routledge. https://doi.org/10.4324/9780203774441

Elster, R. S., & Flyer, E. (1982). *A study of relationships between educational credentials and military performance criteria*. Naval Postgraduate School. https://hdl.handle.net/10945/30228

Federal Reserve Bank of Minneapolis. (n.d.). *Consumer price index*. Retrieved June 1, 2025, from https://www.minneapolisfed.org/about-us/monetary-policy/inflation-calculator/consumer-price-index-1913-

Harrell, F. E. (2015). Multivariable modeling strategies. In *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis* (pp. 63–102). Springer International Publishing. https://doi.org/10.1007/978-3-319-19425-7_4

Herranen, H. (2004). Professional and efficient in action but conscript oriented: The Finnish defence forces. In *Bulilding sustainable and effective military capabilities. a systemic comparison of professional and conscript forces* (pp. 97–108). IOS Press. https://www.iospress.com/node15242/books/building-sustainable-and-effective-military-capabilities

Hof, T., Zuidema, P. M., & Pennings, H. J. M. (2023). Quality of life, psychosocial characteristics, and study skills affecting recruits' intention to quit basic military training. *Military Psychology*, *35*(5), 467–479. https://doi.org/10.1080/08995605.2022.2124790

Holcner, V., Davidová, M., Neubauer, J., Kubínyi, Ľ., & Flachbart, A. (2021). Military recruitment and Czech labour market. *Prague Economic Papers*, *30*(4), 489–505. https://doi.org/10.18267/j.pep.778

International Institute for Strategic Studies. (n.d.). *The Military Balance*. Retrieved June 1, 2025, from https://www.tandfonline.com/journals/tmib20

NATO. (2022). *NATO 2022 strategic concept*. NATO. https://www.nato.int/strategic-concept/

NATO. (2025). *Defence expenditures and NATO's 2% guideline*. https://www.nato.int/cps/en/natohq/topics_49198.htm

NATO. (n.d.). *What is NATO?* Retrieved June 1, 2025, from https://www.nato.int/nato-welcome/index.html

NATO Research and Technology Organization. (2007). *Recruiting and retention of military personnel* (RTO-TR-HFM-107). NATO Research & Technology Organisation. https://www.nato.int/issues/women_nato/Recruiting%20&%20Retention%20of%20Mil%20Personnel.pdf

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

U.S. Mission to NATO. (n.d.). *About NATO*. Retrieved June 1, 2025, from https://nato.usmission.gov/about-nato/

Warner, J. T., & Asch, B. J. (1995). Chapter 13: The economics of military manpower. In *Handbook of defense economics* (pp. 347–398, Vol. 1). Elsevier. https://doi.org/10.1016/S1574-0013(05)80015-8

World Bank. (n.d.). World bank open data. Retrieved June 1, 2025, from https://data.worldbank.org

Appendix

Regression without time fixed effects

To verify the robustness of the main findings, a one-way fixed-effects model including only entity effects was estimated. It is reported in Table 9 and Table 10. The other main predictors remained statistically significant, while GDP per capita lost significance in this model. The R^2 decreased to 0.1247, meaning the model explains about 12.47% of the within-country variation in active armed forces per capita, which is less than the model with time fixed effects included.

Table 9 *PanelOLS Estimation Summary of One-way Model*

Dep. Variable	`Active Armed Forces per capita`
Estimator	PanelOLS
No. Observations	285
Entities	32
Time Periods	9
Cov. Estimator	Clustered
R-squared	0.1247
Log-likelihood	259.79
F-statistic (robust)	2.3521
P-value (F-stat)	0.0242
Distribution	F(7, 246)
F-test for Poolability	70.146
P-value	0.0000
Distribution	F(31, 246)
Included effects	Entity

Table 10 *Parameter Estimates of One-way Model*

Variable	Coef.	Std. Err.	t-stat	p-value	CI Lower	CI Upper
Unemployment rate	-0.0011	0.0088	-0.1208	0.9039	-0.0185	0.0163
Secondary education attainment rate	-0.0084	0.0041	-2.0404	0.0424	-0.0165	-0.0003
GDP per capita	0.2878	0.2506	1.1484	0.2519	-0.2058	0.7813
Def. spend. % GDP	0.0882	0.0379	2.3250	0.0209	0.0135	0.1629
GDP per capita % change	-0.0014	0.0016	-0.8991	0.3695	-0.0045	0.0017
Def. spend. % GDP % change	-0.0013	0.0007	-1.7632	0.0791	-0.0027	0.0001
Education Dummy	-0.0299	0.0223	-1.3412	0.1811	-0.0739	0.0140