Stockholm School of Economics
Department of Economics
BE551 Degree project in economics
Spring 2026

Shocking News: Macroeconomic Consequences of Anticipated Future Defense Spending in Europe

Jonathan Rasmusson (26040) and Carl Wulff Sahlén (25885)

Abstract

European governments have recently committed to substantial rearmament programs, yet the macroeconomic consequences of anticipated future defense spending remain almost entirely unstudied in a European setting. To address this gap, we introduce a novel data set of 35 European countries spanning seven decades. We estimate Bayesian SVARs with hierarchical pooling, identifying defense news shocks by finding innovations that maximize forecast error variance over a five-year horizon while staying independent of contemporaneous defense spending. The identified shocks explain approximately 36 percent of future defense spending variation over a ten-year horizon. Our impulse response functions show that defense news shocks generate an immediate and significant expansion in output, consumption, investment, and employment. The positive impact response of private consumption is inconsistent with standard neoclassical predictions and instead points to an aggregate demand channel. Fiscal multipliers are positive but imprecisely estimated, with wide credible intervals reflecting large heterogeneity across the broader European sample. Our findings suggest that policymakers should exercise caution in relying on crowding-in effects to offset the fiscal cost of rearmament.

Keywords: Bayesian SVAR, defense news shocks, European rearmament, fiscal foresight, fiscal multipliers, import dependence, impulse response functions

JEL: E62, E65, H30, H56, H87

Supervisor: Sampreet Goraya
Date submitted: 2026-05-11
Date examined: 2026-05-19
Discussants: Erik Nyström and Rasmus Samuelsson
Examiner: Johanna Wallenius

Acknowledgments

We would like to extend our sincere gratitude and appreciation to our supervisor, Sampreet Goraya, assistant professor at the Stockholm School of Economics Department of Economics, for his valuable guidance, recommendations, and support.

We also thank Nadav Ben Zeev, Evi Pappa, and Elena Scola Gagliardi for writing "The Offensive Power of Defense News in Europe" (2026). Our bachelor thesis was greatly inspired by their paper.

In Memory of Gustav Andersson

We would like to dedicate this thesis to our dear friend Gustav Andersson, who passed away during his studies at the Stockholm School of Economics. Gustav had a deep interest in defense and societal issues. He also had a rare ability to bring people together, including the authors of this thesis. May he rest in peace.

Contents

1 Introduction **1**

2 Economic Theory **3**

2.1 Fiscal Multipliers . . . 3
2.2 Causal Mechanisms . . . 3
2.3 Defense Spending Effects . . . 4
2.4 Anticipated Defense Spending Effects . . . 5
2.5 Import Dependence . . . 5

3 Literature Review **6**

3.1 Defense Spending Multipliers . . . 6
3.2 Fiscal Foresight . . . 7
3.3 European Heterogeneity . . . 7
3.4 Research Gap and Contribution . . . 8

4 Methodology **8**

4.1 Data . . . 8
4.2 Introduction to Structural Vector Autoregressions . . . 11
4.3 Overview of Empirical Strategy . . . 11
4.4 Reduced-Form VAR . . . 12
4.5 Mapping to Structural Shocks and Identification . . . 13
4.5.1 Exogeneity of Defense News Shocks . . . 15
4.6 Posterior Simulation . . . 16
4.6.1 Bayesian Estimation . . . 16
4.6.2 Hierarchical Pooling . . . 18
4.7 Fiscal Multipliers . . . 20
4.8 Import dependence . . . 21

5 Results **23**

5.1 Baseline SVAR . . . 23
5.2 Multipliers . . . 26
5.3 Import dependence . . . 28
5.4 Robustness checks . . . 29

6 Discussion **30**

6.1 Validity and Limitations . . . 30
6.1.1 Estimation and Inference Validity . . . 30
6.1.2 Internal Validity . . . 32
6.1.3 External Validity . . . 33
6.2 Comparison with Literature . . . 33
6.2.1 Multipliers . . . 33

6.2.2 Institutional and Regional Heterogeneity . . . 34
6.2.3 Import Dependence Heterogeneity . . . 35
6.3 Theoretical Implications . . . 36

7 Conclusion **37**
7.1 Summary of Findings . . . 37
7.2 Conclusions . . . 37
7.3 Future Research and Policy Recommendations . . . 38

References **39**

Appendices **44**

Appendix A: Supplementary Materials **44**
A.1 Robustness . . . 44
A.1.1 Truncation Horizon . . . 44
A.1.2 Oil-Price Control . . . 44
A.1.3 Terrorism Control . . . 45
A.1.4 Excluding the Post-COVID 19 Period . . . 46
A.1.5 Alternative Defense Spending Definition . . . 47
A.1.6 Geopolitical risk index VAR . . . 49
A.2 Import dependence and Regional IRFs . . . 51
A.3 Data . . . 53
A.3.1 Defense Spending . . . 53
A.3.2 Primary Sources . . . 54
A.3.3 Novel Data . . . 54
A.3.4 Extending Western European Series . . . 54
A.3.5 Serial Correlation . . . 55
A.4 Further Discussions on Limitations . . . 55
A.4.1 Currency Conversions . . . 55
A.4.2 Exchange Rate Dynamics . . . 56
A.4.3 Statistical Identification . . . 57
A.4.4 Linearity . . . 57

Appendix B: Artificial Intelligence Disclosure **59**
B.1 GenAI Tools Utilized and their Application . . . 59
B.2 Contribution to Thesis Quality . . . 59
B.3 Potential Risks and Mitigation Strategies . . . 59
B.4 Insights from using GenAI in the Thesis Process . . . 60

1 Introduction

Following Russia's full-scale invasion of Ukraine in 2022, European states have committed to significant rearmament objectives. This marks a substantial shift from the post-Cold War decline in military expenditure, during which defense spending in Europe fell to an average low of 1.27 percent of GDP in 2015 (see Figure 1). By contrast, NATO's 2025 Hague commitment requires allies to invest 5 percent of GDP annually in core defense requirements and related spending by 2035 (NATO, 2025). Further, the European Commission's Readiness 2030 plan aims to mobilize up to EUR 800 billion in additional defense investment across Europe (European Commission, 2025).

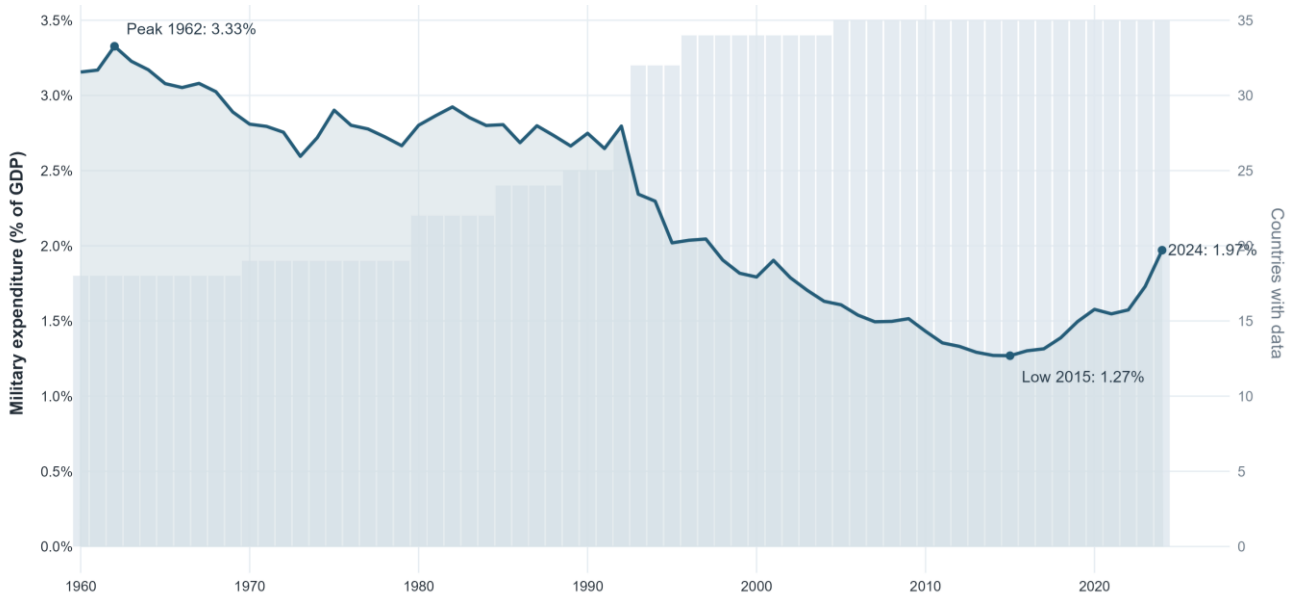


Figure 1: Average military expenditure in Europe-35 country sample, 1960-2024. Source: World Bank (2026), based on SIPRI military expenditure data.

These headline commitments raise an important but understudied question: how does news about future defense spending affect the economy? This issue is of direct policy relevance: the effects of announcements of government spending programs have both monetary and fiscal policy implications. For monetary policymakers, this matters because such announcements may alter macroeconomic conditions, potentially increasing inflationary pressures as forward-looking agents adjust their behavior (Bokan et al., 2025). For fiscal policymakers, understanding the true economic cost of rearmament requires investigating whether the policy crowds in private activity (Ben Zeev, Pappa, and Scola Gagliardi, 2026, pp. 1–2). An often neglected, but important, mechanism is that forward-looking agents adjust their economic behavior immediately to news about future spending. This anticipation effect is often neglected in the literature and can therefore lead to inaccurate estimates (Ramey, 2011, p. 1; Ben Zeev and Pappa, 2017, pp. 1–2).

Although the macroeconomic effects of government spending are widely studied, and defense spending has long been used to identify fiscal shocks, the role of defense spending anticipation remains less understood in the European context. To our knowledge, Ben Zeev et al. (2026) provide the only existing study of defense news shocks in Europe. We build on their contribution by extending the analysis to a significantly broader 35-country European sample. This expanded setting allows us to address the central research question of this thesis:

How do defense news shocks affect macroeconomic activity in Europe, and how do these effects vary?

To answer this question, we estimate average European responses and examine heterogeneity across regional, institutional, and import-dependence characteristics. Methodologically, the study relies on a Bayesian structural vector autoregression (SVAR) framework with hierarchical pooling and medium-run restrictions, identifying defense news shocks through maximum forecast error variance (MFEV) over a five-year horizon. We find that defense news shocks generate a gradual increase in defense spending and an immediate and significant increase in aggregate output. Consumption, investment, and employment rise following the news shock, suggesting that anticipated defense spending can increase private activity.

The associated fiscal multipliers are positive but imprecisely estimated, with wide credible intervals reflecting substantial cross-country heterogeneity. The heterogeneity analysis indicates that EU and NATO subsamples are broadly close to the baseline, while regional and import-dependence groupings point to more varied effects. Low-import-intensity economies appear to exhibit stronger multipliers than high-import-intensity economies, consistent with an import-leakage channel, although this evidence remains suggestive rather than conclusive. Overall, our results suggest that policymakers should be cautious when extrapolating results from one grouping to another, and they should not assume that private-sector crowding-in will reliably offset the fiscal costs of higher defense spending.

This paper proceeds as follows. Section 2 introduces foundational theory. Section 3 reviews relevant literature. Section 4 presents the data set and econometric methodology. Section 5 reports the results. Section 6 discusses and interprets the findings, limitations, and validity of the estimates. Section 7 concludes.

2 Economic Theory

This section introduces the foundational economic theory underpinning this study.

2.1 Fiscal Multipliers

The fiscal multiplier of government spending is generally defined as the ratio between a change in output and the change in spending that caused it. A multiplier equal to unity implies that output rises one-for-one with the increase in government spending. A multiplier above unity implies that output rises by more than the initial spending, as private sector activity is crowded in. Conversely, a multiplier below unity implies crowding-out effects (Ilzetzki, Mendoza, and Végh, 2013, p. 244; Morelli, 2025, pp. 1–2). Economic theory suggests that the size will depend on a multitude of variables. Ramey (2019, p.90) broadly identifies eight sources on which the relationship depends, such as the type of spending, monetary policy response, and whether it was anticipated.[1]

2.2 Causal Mechanisms

To interpret the estimated effects of defense news shocks, it is first necessary to clarify the mechanisms through which government spending can affect output. This varies across economic theories. Many introductory macroeconomics textbooks feature a simple Keynesian cross model. The model posits that output is, in the short run, demand-determined. An increase in government spending, therefore, directly raises output through aggregate demand, with the initial impulse reverberating through the economy as households adjust consumption in response to higher income. In its simplest form, the multiplier becomes a function of the marginal propensity to consume (MPC), equal to $\frac{1}{1-MPC}$, and as such exceeds unity when $MPC > 0$ (Ramey, 2019, p.91).

Among more advanced models, the key theoretical divide today is between the New Keynesian and neoclassical traditions (Morelli, 2025, p.2). According to Ramey (2019, p.92), neoclassical models where labor supply and capital stock are allowed to vary often imply positive multipliers through a supply-side effect. As government spending must be financed, the private sector anticipates a future negative wealth effect, and households, as such, supply more labor, pushing output upwards. However, as Woodford (2011, pp. 2–4) emphasizes, this same wealth effect also crowds out private consumption, so the positive output response does not necessarily imply an above-unity multiplier.

[1]The exact eight are: "1) the persistence of the change; 2) the type of spending or taxes that changed; 3) how the policy was financed; 4) whether it was anticipated; 5) how the policy was distributed across potentially heterogeneous agents; 6) how monetary policy reacted; 7) the state of the economy when the policy took effect; and 8) other features that characterize the economy such as level of development, exchange rate regime, and openness."

In contrast to neoclassical predictions, New Keynesian models can suggest crowding-in of consumption under some specifications, owing to an aggregate demand effect. Firstly, an increase in aggregate demand stemming from government spending can raise short-term consumption under sticky-price and imperfect-competition assumptions. Under these conditions, firms choose to increase their production rather than raise prices, instead reducing their markup. Increased production leads to more labor demand, in turn raising real wages and consumption (Morelli, 2025, pp. 2–3). Another mechanism arises if allowing for household heterogeneity. MPCs may then differ between households. This can be attributed to financial frictions that prevent consumption smoothing and force some households to consume all their disposable income. Hand-to-mouth consumers may also stem from low income and low or illiquid wealth. This yields a larger share of high-MPC households. If government spending is not immediately offset by higher taxes, these households raise consumption, leading to larger fiscal multipliers (Morelli, 2025, pp. 12–13; Auclert, Rognlie, and Straub, 2024, pp. 2–3).

Other New Keynesian models emphasize the role of monetary policy responses. New Keynesian dynamic stochastic general equilibrium models combine the Keynesian demand-side reasoning with the intertemporal effects found in neoclassical models (Woodford, 2011, p. 1; Ramey, 2019, p. 92). Baseline versions assuming representative households, sticky prices and no financial frictions typically lead to below-unity spending multipliers. The modeled fiscal expansion raises inflation, leading central banks to raise the real interest rate, in turn reducing private consumption and investment (Woodford, 2011, pp. 14–15). However, if sticky wages are assumed, workers may end up supplying more hours than they would freely choose at the prevailing wage, therefore dampening the rise in marginal costs and hence also in inflation, leading to less monetary policy tightening and, correspondingly, a larger multiplier (Woodford, 2011, pp. 6, 9).

2.3 Defense Spending Effects

A common approach to structuring the macroeconomic effects of defense spending is to distinguish between the security, supply, and demand channels (Dunne et al., 2005, pp. 450–451; Olejnik, 2023, p.6). The security channel treats defense spending as a public good, which is crucial for protecting life and ensuring the proper functioning of markets. It has been suggested that there is a certain level of security generated by military spending that maximizes economic growth (Shieh et al., 2002, p. 443).

The military spending supply channel has both positive and negative effects on growth. Increased growth may come from expenditure on national defense manufacturing that leads to profits that are reinvested in the industry to improve its production capabilities. Military R&D investments can also have civilian technological spillovers that drive long-term growth (Olejnik, 2023, p.6;

Ilzetzki, 2025, p. 19). On the negative side, large military spending is not governed by market processes, making it susceptible to inefficient resource allocation (Knight et al, 1996). Finally, institutions such as conscription may have adverse labor supply effects as conscripts are temporarily stunted in their human capital accumulation and may work below their potential productivity for the duration of their service (Olejnik, 2023, p.7).

The demand channel is the effect on the economy through aggregate demand. In a Keynesian setting, higher government spending raises output by increasing aggregate demand and current income. In neoclassical and standard New Keynesian DSGE settings, by contrast, the spending impulse is partly offset as households and firms anticipate a lower future disposable income. ECB modeling on increased defense spending illustrates these dynamics. Its semi-structural models omit Ricardian effects and thus feature a comparatively small impact of interest rates on consumption, generating higher multipliers as a result. Its DSGE models, by contrast, produce crowding-out of private demand as households and firms anticipate future taxation and lower disposable income. HANK models allow for varying MPCs across households: when expenditure is directed towards sectors that employ low-income, high-MPC workers, the income it generates is to a larger degree converted into consumption, and the resulting fiscal multiplier rises above ECB benchmark models (Bokan et al., 2025).

2.4 Anticipated Defense Spending Effects

Forward-looking economic agents react not only to contemporary conditions but also to expected future fiscal policy. This foresight gives rise to anticipation effects: announcements of future defense spending may affect economic agent behavior before the spending is realized (Ramey, 2016, p. 85). Ramey (2011, p. 3) shows that series from both professional forecasters and narratives over military-spending shocks in Granger-cause standard vector autoregression (VAR) shocks. This implies that normal VARs capture fiscal shocks later than the information became available to private agents and therefore miss the initial effects of news. It is useful to distinguish two types of innovations in a defense spending series. A defense news shock is information about future defense spending that arrives before the spending occurs, while a defense surprise shock is an innovation that raises spending on impact (Ben Zeev and Pappa, 2017, pp. 1568–1570).

2.5 Import Dependence

Standard economic theory would suggest that the multiplier is smaller in more open economies that import a large share of their goods and services. The reason is that in more open economies, part of the effects of increased government demand leaks abroad, attenuating the output response and reducing the multiplier (Barrell, Holland and Hurst, 2012, pp. 12–13).

3 Literature Review

This section explores previous empirical studies and situates our findings within the existing body of research. Our research relates to three strands of the fiscal multiplier literature: defense spending multipliers, fiscal foresight, and European multiplier heterogeneity. We discuss each strand in turn below.

3.1 Defense Spending Multipliers

A first strand of literature concerns the defense spending specific multiplier. Defense spending has long occupied a central space in the fiscal multiplier literature due to being a plausibly exogenous source of variation in government expenditure. Much of this literature has also focused on the United States, using military spending as an instrument for government spending in general (See Ramey and Shapiro, 1998; Ramey, 2011; Barro and Redlick, 2011; Nakamura and Steinsson, 2014; Antolin-Diaz and Surico, 2025). The broad finding from the U.S. evidence is that military spending tends to raise output, but with substantial variation across identification strategies: Ramey (2011) finds a multiplier of roughly one, Ramey and Zubairy (2018) find estimates closer to two-thirds, while Nakamura and Steinsson's (2014) local fiscal multiplier approach produces a larger estimate of around 1.5 (as reported by Ilzetzki, 2025, pp. 9-10).

Research on fiscal defense multipliers in a European context remains more limited. This reflects the U.S. orientation of the earlier literature and European data constraints. Defense spending data in Europe are generally lower-frequency, and long quarterly time series are unavailable (Furceri et al., 2026, p. 11; Ben Zeev et al., 2026, p. 2). Instead, European estimates mostly rely on annual panels with pooled cross-country methods. Ilzetzki (2025, p. 1) reports a consensus that the economy expands to accommodate at least some military spending increases. However, findings diverge on the magnitude and whether the private sector is crowded out. He calculates that a 2 to 3.5 percent rise in defense spending as a share of GDP would raise Europe-wide GDP by 0.9 to 1.5 percent, implying a multiplier of about 0.6 to 1 (Furceri et al., 2026, p. 10). García-Serrador, Sarasa-Flores, and Ulloa (2025, p. 1) estimate an EU-27 defense multiplier of 1.4 after one year and 1.6 after two years. Furceri et al. (2026) find an EU-27 defense spending cumulative multiplier of around 1.3 on impact and around 1.9 two years after the shock (Furceri et al., 2026, p. 14).

The methodology used in this thesis most closely resembles Ben Zeev et al. (2026). Their identified defense news shocks explain a large share of the forecast error variance in military spending and raise consumption, investment, employment, and output (Ben Zeev et al., 2026, p. 1). Their baseline European median output multiplier reaches 1.24 after one year, 1.06 after four

years, before declining to 0.63 after ten years. Surprise defense shocks are substantially weaker and turn negative at longer horizons (Ben Zeev et al., 2026, p. 11).

3.2 Fiscal Foresight

A second strand of literature addresses the empirical challenge arising from the fiscal foresight of economic agents. Ignoring this anticipation can bias estimates of the multiplier as the timing of the shock is misidentified. Most of the literature treats defense spending as unanticipated, even though fiscal policy is frequently announced before implementation (Ben Zeev et al., 2026, pp. 1–2). Ramey (2011, p. 21) addresses this by constructing a narrative series of shocks with historical defense news. Fisher and Peters (2010) address the same issue using cumulative excess returns on defense-contractor stocks as a proxy for anticipated changes in defense spending. Finally, Ben Zeev and Pappa (2017) identify defense news shocks as innovations that maximize the five-year forecast error variance of future defense spending while remaining orthogonal to current defense spending (Ben Zeev and Pappa, 2017, pp. 1–2). They find that these shocks are correlated with Ramey's news shocks, but explain a larger share of macroeconomic variance, while not requiring the manual construction of a narrative series (Ben Zeev and Pappa, 2017, p. 2). Ben Zeev et al. (2026) extend this strategy to 16 Western European economies in a Bayesian SVAR model with hierarchical pooling.

3.3 European Heterogeneity

A third strand of the literature investigates heterogeneity in the macroeconomic effects of military spending across European economies. While differences in output responses to government spending are well documented in the broader fiscal multiplier literature, the specific context of European defense spending has not been thoroughly investigated. For general fiscal multiplier heterogeneity, an established point of reference is Ilzetzki, Mendoza, and Végh (2013, pp. 239–240). They find evidence that a diverse set of conditions can affect the output response, such as the level of development, exchange-rate regime, trade openness, public indebtedness, and spending composition. In a European defense context, Ben Zeev et al. (2026, pp. 2, 19–20) find evidence for larger multipliers among NATO members and high-debt economies. García-Serrador et al. (2025, pp. 1, 38–39) document strong temporal variations that are associated with business-cycle fluctuations. They find that multipliers are larger in downturns and when fiscal space is large. Furceri et al. (2026, p. 18) similarly find larger EU defense multipliers under favorable borrowing conditions.

Import dependence is another specific source of heterogeneity. García-Serrador et al. (2025, pp. 31–32, 38–39) find that low reliance on imported military goods increases the defense spending

multiplier, while high import dependence can considerably weaken it. In their import dependence specification, the multiplier exceeds 1.5 under low military-import reliance, whereas high import dependence yields multipliers below unity in the first three years and negative by year six. Furceri et al. (2026, p. 18) report a similar pattern: the one-year output multiplier is around 3 in low-import-intensity environments, compared with about 1 in high-import-intensity environments. To our knowledge, no previous study has examined whether responses to defense news vary with the level of import dependence.

3.4 Research Gap and Contribution

Although several studies have assessed the macroeconomic impact of increased defense spending, and others have addressed anticipation effects by identifying news shocks, Ben Zeev et al. (2026) is the only previous paper to employ a comprehensive defense news shock identification strategy in a European context. However, their panel is limited to only 16 Western European countries. This leaves it ambiguous whether their results generalize to a broader European sample and to relevant institutional groupings such as the EU-27 and all European NATO members. We address this gap by constructing a novel dataset covering 35 European economies, expanding both the geographic coverage and the usable country-year observations from the original baseline sample. This allows us to assess the generalizability of the existing evidence on European defense news shocks. The larger sample also enables us to examine heterogeneity across Europe. A novel contribution of our analysis is to examine one such source of heterogeneity: whether responses to defense news vary along the import-dependence dimension.

4 Methodology

4.1 Data

We construct a novel unbalanced panel of 35 European countries, with annual country series that vary in length, the longest spanning 1961–2025. The panel includes all EU-27 member states and all European NATO members. Table 1 provides an overview of the variable definitions and transformations, while Table 2 details the panel of 35 European countries.

Table 1: Variable definitions and transformations.

Variable	Definition	Transformations
Defense spending	Real military expenditure in local currency calendar years, current prices	Rebased to 2020 prices with national GDP-deflator ratio; log-levels and per capita

Variable	Definition	Transformations
Output	Real gross domestic product in local currency, 2020 prices	Log-levels and per capita
Consumption	Real private final consumption expenditure in local currency, 2020 prices	Log-levels and per capita
Investment	Real gross fixed capital formation in local currency, 2020 prices	Log-levels and per capita
Employment	Total employment, persons	Log-levels and per capita
Real wage	Real product wage, 2020 = 100 (national currency)	Deflated with national GDP deflator; log-levels
Interest rate	Nominal short-term interest rates (percentage), defined as the three-month money market rates. For the euro area the rate is the Euribor	-
Inflation	Consumer price index (2020 = 100)	Annual percentage change

Defense spending represents our fundamental variable and is sourced from the Stockholm International Peace Research Institute (SIPRI). It comprises expenditure on personnel, operations and maintenance, procurement, military R&D, military infrastructure spending, and military aid. The information is based on open sources from national budget documents, international statistics, and specialist journals and newspapers (SIPRI, 2026). For a thorough explanation of our defense spending definition, see appendix A.3.1.

Macroeconomic data is primarily sourced from the annual macro-economic database of the European Commission's Directorate General for Economic and Financial Affairs (AMECO) (European Commission, 2026). AMECO data is itself primarily sourced from Eurostat (the Statistical Office of the European Commission). To extend our time series where possible and introduce more countries, we supplement our data set with OECD (2026) sources. These sources allow us to create a data set equivalent to the one described by Ben Zeev et al. (2026). Upon close inspection of this data, we observe that three variables pose a constraint to extending our series and introducing more countries: employment, real wage, and interest rate. To extend the panel into a novel data set, we conduct thorough data collection, drawing from several national statistical offices, central banks, and data aggregation platforms. In the case of four countries (Albania, North Macedonia, Serbia, and Switzerland), we recalculate the real wage series thanks to longer series from national statistical offices. We obtain interest rate data from four countries (Albania, Montenegro, North Macedonia, and Serbia) by implementing Python pipelines for web

scraping of central bank data and image processing of historical charts. For detailed descriptions on data collection methods, see appendix A.3.3.

Table 2: Country sample and available observation periods.

Country	Period	Obs.	Country	Period	Obs.	Country	Period	Obs.
Albania	2002-2025	24	Greece	1980-2025	46	Poland	1992-2025	34
Austria	1967-2025	59	Hungary	1995-2025	31	Portugal	1966-2025	60
Belgium	1961-2025	65	Ireland	1961-2025	65	Romania	1995-2025	31
Bulgaria	1998-2025	28	Italy	1961-2025	65	Serbia	2005-2025	21
Croatia	2002-2025	24	Latvia	1995-2025	31	Slovak Republic	1995-2025	31
Cyprus	1999-2024	26	Lithuania	1999-2025	27	Slovenia	1998-2025	28
Czechia	1993-2025	33	Luxembourg	1961-2025	65	Spain	1961-2025	65
Denmark	1965-2025	61	Malta	1995-2024	30	Sweden	1961-2025	65
Estonia	1996-2025	30	Montenegro	2005-2025	21	Switzerland	1974-2025	52
Finland	1962-2025	64	Netherlands	1961-2025	65	Türkiye	1990-2025	35
France	1961-2025	65	North Macedonia	2007-2025	19	United Kingdom	1961-2025	65
Germany	1991-2025	35	Norway	1970-2025	56			

Summary statistics: Longest series (65 years) include Belgium, France, Ireland, Italy, Luxembourg, the Netherlands, Spain, Sweden, and the UK. The shortest series (19 years) is in North Macedonia. Obs. denotes annual observations available for the country-level VAR after harmonizing the source series. The variable with the shortest country-year time series in country i determines the length of that country's sample. This amounts to 1522 country-year observations in our baseline panel.

4.2 Introduction to Structural Vector Autoregressions

A standard vector autoregression (VAR) is a multi-equation model that captures dynamic interactions among macroeconomic variables. In a reduced-form VAR, each variable is expressed as a linear function of the system's past values and an unobserved error term. While this framework efficiently summarizes historical data, these reduced-form residuals are typically correlated with each other, thus lacking direct economic interpretation (Stock and Watson, 2001, pp. 101-102; Kilian and Lütkepohl, 2017, p. 94). For a shock to become economically meaningful, researchers must impose identifying assumptions, thereby transforming the reduced-form model into a structural vector autoregression (SVAR).

The application of SVARs to macroeconomic modeling has evolved significantly over the past decades. The fundamental VAR framework traces its origins to Sims (1980) and was later introduced to fiscal policy research by Blanchard and Perotti (2002). However, these approaches fail to account for the implementation lag arising between the decision and the realization of government expenditure (Morelli, 2025, p.3). This leaves the identification vulnerable to anticipation effects, as shown by Ramey (2011). Ignoring this effect causes traditional SVARs to misidentify the timing of the shock, leading to biased estimates of fiscal multipliers. The historical limitation of failing to incorporate fiscal foresight into the SVAR methodology directly motivates our choice to employ a news-based identification strategy in a European setting.

4.3 Overview of Empirical Strategy

We follow the structural vector autoregression (SVAR) design with medium-run restrictions, as pioneered by Ben Zeev and Pappa (2017) in a United States setting, and extended with Bayesian estimation and hierarchical pooling for a European context by Ben Zeev et al. (2026). Medium-run restrictions are achieved by constructing a defense news shock, which is defined as the structural shock that maximally explains future defense spending variation over a five-year horizon, with the condition that defense spending does not respond to the defense news shock on impact. Formally, we can express this as the shock that maximizes the five-year forecast error variance of defense spending while staying orthogonal to contemporaneous defense spending. The identifying assumption for our SVAR is that defense spending is only driven by two shocks: one defense news shock that moves defense spending with a lag, and one defense surprise shock that moves defense spending on impact. To recover the defense news shock and trace its transmission mechanism through European economies, we construct a multivariate VAR with macroeconomic variables that are deemed endogenous. We implement the outlined methodology into an R script, which we use to estimate our baseline results and conduct robustness tests.

To decide the appropriate pooling strategy for our estimated country-level objects, we examine the cross-country movement of the defense news shocks in our sample of 35 European countries. The average pairwise correlation of the country-specific news shock is approximately seven percent, which signals high cross-country heterogeneity and motivates the estimation of SVARs at the country-level. To estimate a true Europe-wide average effect, the empirical framework must simultaneously account for cross-country heterogeneity and within-country estimation uncertainty. Instead of adopting a standard pooling framework of a panel SVAR by imposing homogeneity across countries, we apply the canonical Bayesian Normal–Normal hierarchical model (Gelman et al., 2013, Ch. 5), which is used by Ben Zeev et al. (2026). This hierarchical pooling approach is suitable given the limited time dimension and unbalanced structure of our Europe-35 data set. The Bayesian estimation treats objects of interest (impulse response functions, maximum forecast error variance, fiscal multipliers) as random variables instead of fixed values. Furthermore, Bayesian estimation allows for the creation of precise posterior distributions, even in cases of small samples (Sims and Zha, 1998, pp.949-950).

4.4 Reduced-Form VAR

We estimate a system of OLS regressions for each country, where each of the eight variables is estimated on the lags of all eight variables, plus a constant (intercept) and a residual.

$$y_{i,t} = \Phi_{i,1} y_{i,t-1} + \cdots + \Phi_{i,p} y_{i,t-p} + c_i + u_{i,t}, \quad u_{i,t} \sim N(0, \Sigma_i) \quad (1)$$

First, in equation 1, $y_{i,t}$ represents a $k \times 1$ vector of variables with defense spending ordered first ($k = 8$), containing the current values of all eight variables for country i at time t.

$$y_{i,t} = \begin{pmatrix} \text{defense spending}_t \\ \text{output}_t \\ \text{consumption}_t \\ \text{investment}_t \\ \text{employment}_t \\ \text{real wage}_t \\ \text{interest rate}_t \\ \text{inflation}_t \end{pmatrix}$$

Second, $\Phi_{i,1}$ is an 8×8 matrix of coefficients explaining how the lagged values of all variables predict this period's values. Following Ben Zeev et al. (2026), we set lag length $p = 1$ which produces sufficient serially uncorrelated defense news shock at an annual frequency, and we also test this empirically for each country (see appendix A.3.5). Defense news shocks must be serially uncorrelated to ensure that they cannot be predicted from their own past values. Since these shocks represent new information to the system, agents should not be able to anticipate the shock based on previous information. If agents were able to anticipate the shock with previous

information, it would undermine the economic interpretation of the shock as capturing news about future defense spending.

Third, $u_{i,t}$ is the *8 × 1* vector of residuals, i.e., the remaining variation of each variable's value this period that last period's variables could not predict, which we call reduced-form residuals or innovations. We assume the residuals, $u_{i,t}$, are drawn from a zero-mean multivariate normal distribution with an *8 × 8* positive-definite covariance matrices, Σ_i. This matrix captures both the unpredictable variation within each variable and its contemporaneous co-movement.

Assuming $u_{i,t}$ follows a multivariate normal distribution is critical because it enables both Bayesian estimation and our structural identification and has been established in the SVAR literature since its inception (Sims, 1980). This assumption is justified by the central limit theorem. Our variables represent macroeconomic aggregates, which are averages of many individual decisions by agents, thus making their unexpected movements (innovations) trend towards normality. This assumption is less likely to hold during times of severe economic crisis (such as the 2008 global financial crisis, the COVID-19 pandemic, and its post-inflation surge), which produce larger innovations than normality predicts. Economic crises also make many individual shocks correlated, which weakens the likelihood that events are independent from each other (International Monetary Fund, 2013, pp. 81-83). Furthermore, defense spending is generally driven by a small number of political decisions rather than many individual choices, making the normality assumption uncertain for our fundamental variable. These are theoretical limitations inherent to our methodology.

4.5 Mapping to Structural Shocks and Identification

The estimated innovations in Σ_i contain the key information that we use to identify the structural shocks $\varepsilon_{i,t}$, but they mix all economic causes simultaneously. We need to disentangle our shock of interest, the defense news shock, from all other possible shocks that can occur in an economy. To achieve this, we begin by rewriting the reduced-form VAR in an equivalent form known as the moving average representation (MA). It tells us that today's values can be expressed as the weighted sum of all past residuals, and we get one matrix for each lag length, telling us how past shocks affect today's economy.

The identification of structural shocks requires the assumption that they are independent, and the pattern of co-movement in Σ_i precisely equates to the pattern of co-movement in the structural shocks. This assumption on Σ_i admits infinitely many solutions, so we need both a starting point, and an identification strategy. Our starting point is to impose the Cholesky identification on Σ_i, which produces a unique lower-triangular matrix (see Figure 2). This is key for economic

interpretation because we ordered defense spending first in the VAR. This restriction ensures that on impact, only the first structural shock can move defense spending. This is the unanticipated defense (surprise) shock stated as one part of our identifying assumption. All other shocks (shocks 2-8), which includes the defense news shock, are restricted to having zero contemporaneous effect on defense spending. This aligns our methodology with the second part of our identifying assumption: a defense news shock represents novel information about future defense spending entering the system today. By construction, the news shock leaves current defense spending unchanged on impact.

Structural Impact Matrix (Cholesky Identification)

Lower Triangular Structure

Variable	Shock 1	Shock 2	Shock 3	Shock 4	Shock 5	Shock 6	Shock 7	Shock 8
Defense Spending	ε_{11}	0	0	0	0	0	0	0
Output	ε_{21}	ε_{22}	0	0	0	0	0	0
Consumption	ε_{31}	ε_{32}	ε_{33}	0	0	0	0	0
Investment	ε_{41}	ε_{42}	ε_{43}	ε_{44}	0	0	0	0
Employment	ε_{51}	ε_{52}	ε_{53}	ε_{54}	ε_{55}	0	0	0
Real Wage	ε_{61}	ε_{62}	ε_{63}	ε_{64}	ε_{65}	ε_{66}	0	0
Interest Rate	ε_{71}	ε_{72}	ε_{73}	ε_{74}	ε_{75}	ε_{76}	ε_{77}	0
Inflation	ε_{81}	ε_{82}	ε_{83}	ε_{84}	ε_{85}	ε_{86}	ε_{87}	ε_{88}

Figure 2: Structural Impact Matrix via Cholesky Identification

The Cholesky identification provides only one way of mapping reduced-form residuals into structural shocks. To get a complete set of valid structural impact matrices, we multiply the lower triangular matrix by an orthonormal matrix. By mapping all valid decompositions, we can identify the shock that corresponds to our definition of a defense news shock. We employ a maximum forecast error variance (MFEV) method to identify the orthonormal matrix that best captures news about future defense spending. Forecast error variance measures how much of the unexpected movement in a variable is driven by a particular shock at different horizons. The key idea is to search for the shock within the space of shocks 2-8 that maximally explains defense spending movement over a five-year horizon, while having zero effect on impact (shocks 2-8 are orthogonal to contemporaneous defense spending by construction as seen in Figure 2). This reflects the gradual materialization of defense news into defense spending.

Formally, we look for a vector inside the orthonormal matrix that solves this as a constrained optimization problem, which simplifies to an eigenvector problem. This is because we are searching for directions that maximize the contribution to future movements in defense spending, while keeping the vector size fixed. In such problems, the solution is the direction that has the

strongest effect on the outcome. This direction is called an eigenvector, and its strength is measured by its eigenvalue. The optimal solution is given by the eigenvector with the largest eigenvalue. The system of equations below encapsulates this idea. Equation (2) selects the vector γ_i that maximizes the contribution of the shock to future movements in defense spending. The remaining equations impose the identifying restrictions: (3) ensures that the news shock does not move defense spending on impact, (4) ensures that the news shock does not capture any defense surprise shocks, and (5) normalizes γ_i.

$$\gamma_i^* = \arg\max_{\gamma_i} \sum_{h=0}^{H} \Omega_{i,1,2}(h) \tag{2}$$

$$\text{subject to } A_i(1,s) = 0 \quad \forall s > 1 \tag{3}$$

$$\gamma_{i,1} = 0 \tag{4}$$

$$\gamma_i' \gamma_i = 1 \tag{5}$$

4.5.1 Exogeneity of Defense News Shocks

The validity of our empirical strategy rests entirely on the identifying assumption that defense spending is driven exclusively by two exogenous structural shocks: a surprise defense shock that hits on impact, and a defense news shock that affects spending with a lag. A potential violation of this identifying assumption would be the contamination of the defense news shock by general productivity news shock (Ben Zeev and Pappa, 2017, p.20). The literature on news-driven business cycles, pioneered by Beaudry and Portier (2006), posits that economic agents react today to news about future improvements in total factor productivity (TFP). A positive productivity news shock generates an immediate economic expansion through wealth effects, increasing output, consumption, and expected future tax revenues. The critique suggests that if a government operates under a procyclical fiscal framework or adheres to a constant defense-to-GDP ratio (such as the NATO two percent target), the anticipation of higher future GDP would automatically trigger an anticipated increase in future defense spending.[2] If this dynamic holds, the SVAR might misidentify a general productivity news shock as a defense news shock. In that case, the observed macroeconomic expansion is driven by expected technological progress rather than anticipated military expenditure.

However, we argue that our identifying assumption remains robust against this critique for two fundamental reasons, rooted in both institutional decision-making and our specific econometric

[2] Even after the 2014 Wales pledge, the 2 percent benchmark was not binding: only 3 Allies met it in 2014 and 10 in 2023, before a sharp increase to an estimated 23, in 2024 (NATO, 2024, p. 2).

design. First, defense spending is widely considered an exogenous component of government expenditure (Ben Zeev and Pappa, 2017, p.20). Unlike social transfers or general government investment, which are highly sensitive to domestic business cycles and tax-revenue fluctuations, military spending decisions are primarily driven by external causes. Geopolitical threats, foreign policy shifts, and international treaties are generally assumed to determine defense spending, rather than domestic economic considerations. Furceri et al. (2026, p.11), in their recent evaluation of European defense spending, test the exogeneity of defense spending shocks across EU countries. They conclude that national defense outlays and procurement contracts are independent from past macroeconomic variables and anticipated GDP changes.

Second, our reliance on the maximum forecast error variance identification strategy inherently protects the defense news shock from productivity shock contamination. The MFEV procedure specifically isolates the linear combination of reduced-form innovations that maximally contributes to the forecast error variance of future defense spending over a five-year horizon. While a productivity news shock might theoretically include a marginal, second order increase in future defense spending due to a larger tax base, its primary explanatory power lies in the variance of output and consumption. By mathematically forcing the algorithm to find the shock that explains the absolute maximum share of defense spending variance, the procedure recovers a shock fundamentally distinct from a general business cycle or productivity signal. Ben Zeev and Pappa (2017, p.23) validate this exact mechanism, demonstrating that MFEV-identified defense news shocks capture strict defense-specific foresight rather than general technological optimism.

4.6 Posterior Simulation

4.6.1 Bayesian Estimation

Classical frequentist inference treats population parameters as fixed, unknown constants, with estimation focused on the likelihood of observing the data given those parameters. In contrast, Bayesian inference treats parameters as random variables characterized by probability distributions. This framework allows for recursive updating, where initial beliefs are refined as new observations enter the system. Formally, instead of conditioning the data y on a fixed parameter θ, Bayesian logic conditions the parameter θ on observed data y. This is achieved by combining the likelihood function $p(y|\theta)$ with the prior distribution $p(\theta)$, which represents existing information or beliefs about the parameter (Gelman et al., 2013, Ch. 1.3). Expressed as Bayes' rule, the posterior distribution is equal to the product of the prior distribution and the likelihood function, divided by $p(y)$, which ensures the posterior integrates to one (see equation 6).

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)} \tag{6}$$

Furthermore, the posterior distribution can be considered proportional to the product of the prior and the likelihood, as $p(y)$ is independent of the parameters θ (see equation 7).

$$p(\theta|y) \propto p(\theta)p(y|\theta) \tag{7}$$

Following the Bayesian theoretical framework, we generate posterior draws of the structural objects of interest, i.e., impulse response functions, forecast error variance decompositions, and fiscal multipliers, which we denote as $\theta_i^{(m)}$ for each country *i* and draw *m*. This approach is particularly advantageous given our small sample size relative to the eight-variable reduced-form VAR, as the Bayesian framework is more robust to uncertainty than frequentist alternatives (Sims and Zha, 1998, pp.949-950).

To implement this estimation, we specify a prior for the reduced-form coefficients matrix Φ_i and the covariance matrix Σ_i. Following Ben Zeev et al. (2026), we employ a Jeffreys prior, which belongs to the normal-inverse Wishart family of priors (NIW). Jeffreys prior is an uninformative prior designed to be invariant under reparameterization, meaning that the prior makes no assumptions about the distribution of our data. Under this prior, the posterior degrees of freedom of Σ_i equal $T - q$, where T denotes the number of country-year observations for country *i* and q are the number of regressors per equation. In our reduced-form VAR model with eight variables k and an intercept ($q = k + 1 = 9$), this implies a minimum condition of $T > 2k$ observations for the posterior to be well-defined. This condition is satisfied for all countries in our sample, even for the smallest country-year sample size (North Macedonia, $T = 19$). The Jeffreys prior, which is called a flat prior, also makes the posterior distribution almost entirely driven by the uncovered patterns in the data, which in practice closely resembles a frequentist approach (Chao and Philips, 2002, p.252). However, we acknowledge the potential hidden bias of uniform priors as a "non-neutral" instrument on our data when moving from coefficients to individual impulse response functions, as highlighted by Arias, Rubio-Ramíres, and Waggoner (2025, p.699). Finally, it is worth noting that countries with shorter samples tend to exhibit more dispersed posterior draws, which can lead to a lower stability rate in the rejection sampling step. This pattern carries over to the hierarchical pooling stage, where countries with more dispersed posterior distributions receive lower precision weights, reflecting their higher estimation uncertainty.

The NIW (Jeffreys) prior has two parts matching the two parameters of the VAR. For each posterior draw, we draw a random $\Sigma_i^{(m)}$ from its inverse-Wishart posterior. This produces a plausible covariance matrix that is consistent with the data but differs slightly from the OLS estimate, reflecting estimation uncertainty. Next, we draw a random $\Phi_i^{(m)}$ from a multivariate

normal distribution conditional on the drawn $\Sigma_i^{(m)}$, centered at the OLS estimate. Likewise, this is a reasonable coefficient matrix that is consistent with the data while containing some estimation noise.

We then perform rejection sampling: if the drawn $\Phi_i^{(m)}$ implies an explosive VAR, i.e., one where the shocks grow without contraction over time, this draw is rejected, and a new draw is attempted. This ensures stationarity of the VAR, which is the primary statistical assumption. Rejecting unstable draws follows theory: an economy eventually returns to a steady state after a shock, which accepting an explosive draw would contradict. However, Ben Zeev et al. (2026) never explicitly state implementing rejection sampling of posterior draws; they only report that their procedure generates *M* sets of posterior objects of interest. This is a methodological assumption of our thesis, which we deem tenable, given that imposing rejection or stability is standard in the Bayesian SVAR literature (Uhlig, 2005, p.391; Rubio-Ramírez, Waggoner, and Zha, 2010, p.688).

Next, we identify the defense news shock. By using the stable drawn pair $(\Phi_i^{(m)}, \Sigma_i^{(m)})$ we perform the previously outlined steps: compute the MA representation, perform the Cholesky decomposition, and solve the constrained MFEV optimization problem to obtain the direction of the news shock $\gamma_i^{*(m)}$. Finally, we can use $\gamma_i^{*(m)}$ to compute the impulse responses, forecast error variance contributions, and fiscal multipliers for this specific draw and store the resulting objects as $\theta_i^{(m)}$. The posterior distribution of our computed objects has direct probabilistic interpretation: all values within the 2.5th and 97.5th percentiles represent the 95 percent credible interval of the specific object (IRF, FEV, or multiplier). Given our Bayesian SVAR model and our observations, there is 95 percent probability that the true value of the object lies within the credible interval. This is conceptually cleaner than a corresponding frequentist approach with 95 percent confidence interval, which would give the interpretation that 95 percent of repeatedly computed intervals will contain the true (fixed) object. This entire process is repeated separately for each country until we have $M = 1000$ sets of stable posterior draws per country in our sample.

4.6.2 Hierarchical Pooling

After obtaining a posterior distribution for each object and country θ_i, we pool our country-level posteriors by using a Bayesian Normal-Normal hierarchical model to compute a single Europe-wide average estimate (Gelman et al., 2013, Ch. 5). The purpose of hierarchical pooling is to strike a balance between accounting for cross-country heterogeneity and estimation noise. This yields a superior Europe-wide estimate compared to a naive pooling approach that computes the sample mean of all 35 country estimates. A naive approach would bias our estimate, since countries with

uncertain estimates would be assigned the same weight as those with more certain estimates. Formally, we begin by making the following assumptions:

$$\theta_i|\eta_i \sim N(\eta_i, \iota_i^2) \tag{8}$$

Each estimated country-object θ_i is assumed to be drawn from a normal distribution centered on the country's true value η_i with variance ι_i^2 (estimated with sample variance $\widehat{\iota_i^2}$ for each country i), capturing within-country estimation uncertainty (see equation 8).

$$\eta_i|\mu, \zeta^2 \sim N(\mu, \zeta^2) \tag{9}$$

Next, each country's true value η_i is assumed to itself be drawn from a common normal distribution centered on the true European-wide average μ with variance ζ^2, capturing cross-country heterogeneity (see equation 9). Similarly to our prior on the country-level draws, we follow Ben Zeev et al. (2026) and select a flat prior on μ. This places zero initial assumptions on the true Europe-wide average. The posterior distribution of μ is the precision-weighted true Europe-wide average credible interval. w_i is the precision-weight for each country i made up of the inverse variance of within-country estimation uncertainty and between-country heterogeneity.[3]

$$w_i = \frac{1}{(\iota_i^2 + \zeta^2)} \tag{10}$$

This means that countries with low estimation noise (small ι_i^2) or low heterogeneity compared to the rest of the samples (small ζ^2) receive more weight, and vice versa (see equation 10). In practice, hierarchical pooling is computed together with the country-level posterior draws $\theta_i^{(m)}$. This entire process of Bayesian estimation, coupled with hierarchical pooling, is called the posterior simulation.

We need to assume conditional independence on country-specific objects θ_i to distinguish ι_i^2 from ζ^2. This means that the European average μ and heterogeneity scale ζ^2 should capture the entire common European effect (we estimated the average pairwise correlation of country-specific news shocks at seven percent), and the remaining variation of the defense news shocks and their corresponding objects (IRFs, MFEV, multipliers) is country-specific. Any remaining heterogeneity after conditional independence is negligible, as shown by Ben Zeev et al. (2026, p.6), and this assumption is standard in Bayesian Normal-Normal Hierarchical models.

[3] Estimated country weights are as follows (listed in ranked order from largest to lowest weight reported in percent): Norway (4.3), Sweden (4.1), Denmark (4.1), Netherlands (4.1), Italy (4.0), Belgium (4.0), France (3.9), Austria (3.9), Poland (3.9), Spain (3.8), Portugal (3.7), Switzerland (3.7), United Kingdom (3.7), Luxembourg (3.6), Finland (3.6), Germany (3.3), Hungary (3.2), Czechia (2.9), Bulgaria (2.9), Greece (2.8), Slovenia (2.8), Ireland (2.7), Serbia (2.1), Estonia (2.1), Türkiye (2.1), Slovak Republic (2.0), Romania (2.0), Cyprus (1.9), Latvia (1.9), Malta (1.6), Albania (1.4), Lithuania (1.2), Croatia (1.2), North Macedonia (1.1), Montenegro (0.4).

The hierarchical pooling simulation follows the previous $M = 1000$ stable draws. At each draw m, we first estimate the between-country variance $\zeta^{2(m)}$, and together with the known within-country variance $\widehat{\iota_i^2}$, we compute the precision-weighted European average $\mu^{(m)}$. $\zeta^{2(m)}$ is calculated as the cross-sectional sample variance $s^{2(m)}$ subtracted by the average within-country variance $\bar{\iota}$. If this equation turns negative, we set $\zeta^{2(m)} = s^{2(m)}$ to ensure non-negative precision weights. Formally, this can be expressed as the methods-of-moments (MoM) estimate: $\zeta^{2(m)} = s^{2(m)} - \bar{\iota}$. Repeating this sequence until we have 1000 stable draws yields the posterior distribution of the true Europe-wide average. The median of this sequence is reported as the point estimate at each horizon, and the 2.5th and 97.5th percentiles form the 95 percent credible bands.

4.7 Fiscal Multipliers

To quantify the economic effects of defense news shocks, we follow the approach by Ben Zeev et al. (2026) in computing fiscal multipliers. The conventional approach in the literature computes this ratio at the horizon where defense spending reaches its highest point, comparing the total accumulated output response against the total accumulated defense spending response up to that point, scaled by the ratio of a country's GDP to its defense budget to convert the result into meaningful monetary units (Ramey and Zubairy, 2018, pp. 863-864). This scaling is necessary because the impulse responses from the reduced-form VAR measure percentage changes, not currency amounts. GDP is far larger than the defense budget, meaning a one percentage change in each represents widely different amounts expressed in monetary terms.

This standard approach fails for defense news shocks because of a fundamental timing mismatch. A news shock represents new information entering the system that signals future spending rather than spending today. Consequently, firms, households, and investors adjust their behavior immediately upon receiving the new information (Morelli, 2025, p.5). This means output, consumption, and investment all respond on impact, while actual defense expenditure only rises gradually over subsequent years. However, at the arrival of the defense news shock, defense spending is left unchanged. Dividing the output response by a zero-defense spending response on impact produces a multiplier that is mathematically undefined, making the standard formula inapplicable precisely at the impact horizon.

To resolve this, we adopt the modified method of Ben Zeev et al. (2026, p.10-11). This approach replaces the contemporaneous defense spending response in the denominator with the present value of the entire ten-year path of defense spending responses to the initial defense spending shock, discounted by the average real long-term interest rate of country *i*. This ensures defense spending several years in the future counts for less than spending arriving next year and is

grounded in rational economic behavior. Defense contractor firms, who receive information today that the government will increase government spending on defense in the future, will immediately begin adjusting production capacity based on the present value of that future revenue stream, not by the materialized government spending on defense.

Formally, for each stable posterior draw *m* and country *i*, the cumulative news multiplier *n* at horizon $h = \{1, 2, 4, 8, 10\}$ is computed as seen in equation 11.

$$M_{i,h,n} = \frac{\sum_{j=1}^{h} \frac{\partial \ln Y_{i,t+j-1}}{\partial \varepsilon_{2,i,t}}}{h \times \sum_{j=1}^{10} \frac{\partial \ln DF_{i,t+j-1}}{\partial \varepsilon_{2,i,t}} \times \frac{1}{(1+r_i)^{j-1}}} \times \frac{\bar{Y}_i}{\bar{DF}_i} \tag{11}$$

The numerator accumulates the log output responses to the defense news shock from the moment of impact to horizon *h*. The multiplier at horizon one represents the immediate output response, at horizon four the multiplier represents the sum of four years of output responses to the defense news shock, continuing up until $h = 10$ years. The denominator represents the present value of the full ten-year defense spending response, multiplied by a scaling factor of horizon *h*.

The discount factor $\frac{1}{(1+r_i)^{j-1}}$ ensures that anticipated defense spending further in the future is weighted lower than near-term spending. The *h* scaling in the denominator grows linearly with the horizon, which means *h* grows in proportion to how many periods of output response(s) have been accumulated in the numerator, keeping the ratio economically interpretable. Finally, the ratio $\frac{\bar{Y}_i}{\bar{DF}_i}$ represents the average nominal GDP divided by the average nominal defense spending for country *i*, which converts the ratio of logged responses into a fiscal multiplier expressed in currency-invariant units.

This formula is computed separately for each posterior draw, producing a posterior distribution of multipliers at each horizon from which we report the median and 95 percent credible bands. For surprise defense shocks, where actual defense spending moves on impact and no timing mismatch exists, we use the standard cumulative formula without discounting (Ramey and Zubairy, 2018). Both the numerator and denominator simply accumulate their respective responses from impact up to horizon h, and the logged response output-to-defense ratio is scaled by the same nominal-GDP-to-defense-spending ratio.

4.8 Import dependence

To examine whether defense spending multipliers vary with import dependence, we construct two country-level import-dependence measures (see equation 12 and 13) and then pool country-level Bayesian SVAR multiplier draws separately for low- and high-dependence groups. First, following García-Serrador et al. (2025, pp. 31-32), we measure military import reliance by using UN

Comtrade military-goods imports and SIPRI current-USD military expenditure over the years overlapping each country's SVAR sample.[4]

$$Military\ Import\ Reliance_i = \frac{Military\ Goods\ Imports_{i,t}}{Military\ Expenditure_{i,t}} \tag{12}$$

Second, following the aggregate import intensity variable used by Furceri et al. (2026, p.17) in their estimation of import-dependence effects for the output response of military spending, we define WDI imports of goods and services as percentage of GDP:

$$Aggregate\ Import\ Intensity_i = \frac{Imports\ Goods\ Services_{i,t}}{GDP_{i,t}} \tag{13}$$

For each country, we require at least ten overlapping observations between the import variable years, and the years used in our SVAR. This results in North Macedonia being excluded from the military import reliance comparison. To split the country sample into low- and high-import-dependence groups, we first calculate each country's average import dependence, then compare that country's average to the overall average import dependence observed across all included country-years. Countries above that benchmark are classified as "high"; countries below it as "low." After this, we apply the hierarchical pooling procedure to each group. This implementation strategy follows the logic of Ben Zeev et al. (2026, p. 19), who examine cross-country heterogeneity by separating countries into high- and low-debt groups. The country-level Bayesian SVAR design prevents us from employing a time-varying import-dependence variable method as is done by García-Serrador et al. (2025) and Furceri et al. (2026).[5]

[4] Military-goods imports are constructed from UN Comtrade annual Harmonized System (HS) import records, using import flow M, world aggregate partner, and commodity codes 93 ("Arms and ammunition; parts and accessories thereof"), 871000 ("Tanks and other armored fighting vehicles; motorized, whether or not fitted with weapons, and parts of such vehicles"), and 890610 ("Vessels; warships").

[5] The import dependence splits results in the following country compositions. Military import reliance split: low = Albania, Belgium, Bulgaria, Croatia, Denmark, France, Germany, Greece, Hungary, Ireland, Italy, Netherlands, Poland, Portugal, Romania, Serbia, Slovenia, Spain, Sweden, Türkiye and United Kingdom; high = Austria, Cyprus, Czechia, Estonia, Finland, Latvia, Lithuania, Luxembourg, Malta, Montenegro, Norway, Slovak Republic and Switzerland. North Macedonia is excluded from this split because it has only two overlapping military-import-reliance observations. Aggregate import intensity split: low = Albania, Austria, Croatia, Denmark, Finland, France, Germany, Greece, Italy, Norway, Poland, Portugal, Romania, Spain, Sweden, Switzerland, Türkiye and United Kingdom; high = Belgium, Bulgaria, Cyprus, Czechia, Estonia, Hungary, Ireland, Latvia, Lithuania, Luxembourg, Malta, Montenegro, Netherlands, North Macedonia, Serbia, Slovak Republic, and Slovenia.

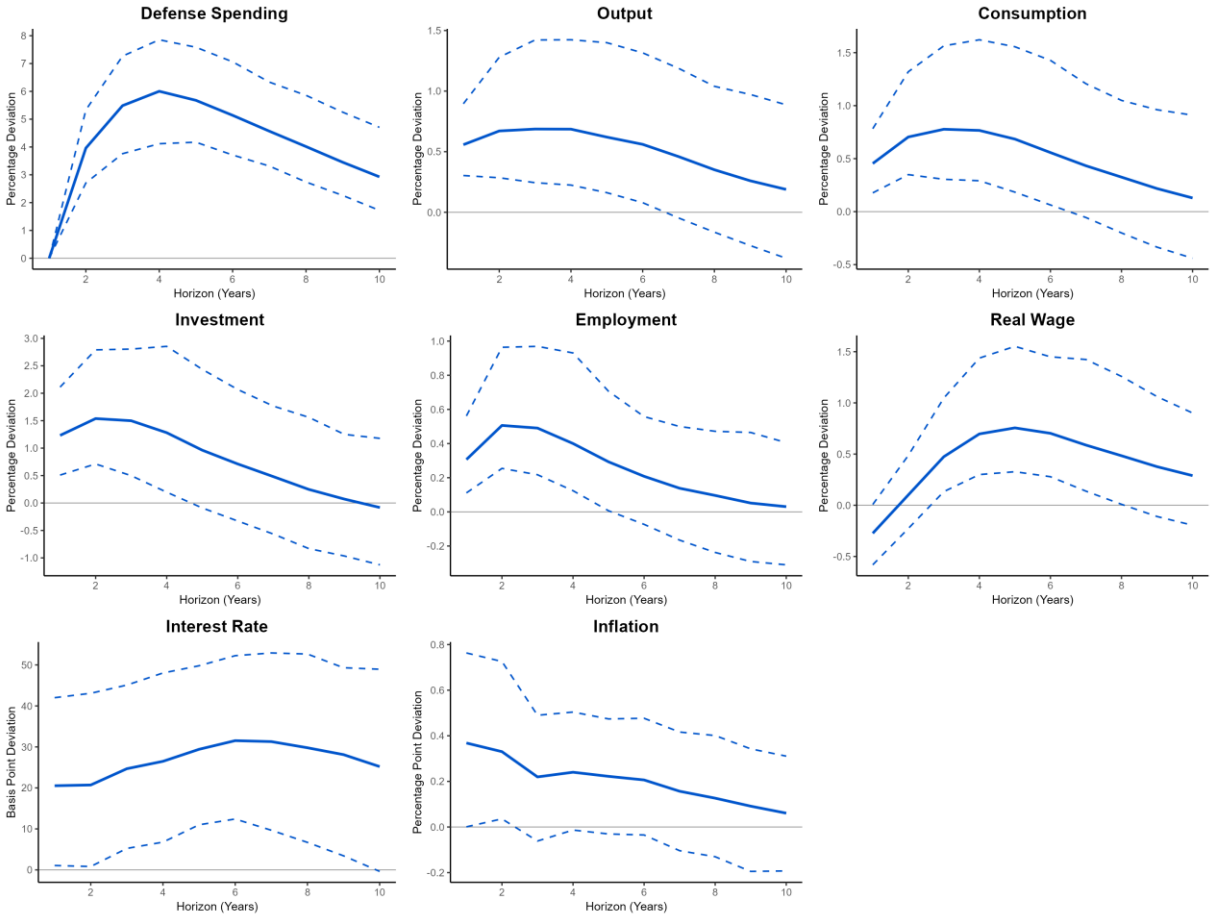
5 Results

5.1 Baseline SVAR

The baseline results from our hierarchically pooled SVAR for our Baseline European sample of 35 economies are reported below, primarily visualized through three graphs. Figures 3 and 5 report median pooled impulse response functions (IRFs) for each variable over 10 horizons. Horizon 1 corresponds to year 0, the impact year. The IRF of a variable traces its movement over time in response to an identified structural shock. The point estimates represent the medians of the hierarchically pooled IRFs. Following standard methodology, the IRFs reported below are the response to a one-standard-deviation defense news shock (Kilian and Lütkepohl, 2017, p. 112). It is measured as the percentage deviation (percentage point deviation for non-logged variables) from the counterfactual trajectory (estimated using the pre-shock information). As such, an output response of x at horizon y means that output is estimated to be x percent above its counterfactual path y years after the shock.

Figure 4 shows the median forecast error variance (FEV) contributions stemming from a defense news shock, for each of our variables. FEV contributions measure the share of a variable's forecast error variance that can be attributed to structural shock at horizon h (Kilian and Lütkepohl, 2017, pp. 113–116). In our methodology, Figure 4 therefore shows how much of each variable's forecast uncertainty is explained by the identified defense news shock. Thus, an FEV contribution of x for output at horizon y means that defense news shocks account for x percent of the model's y-year-ahead forecast uncertainty in output. A larger share implies that defense news shocks explain more of the unexpected movements in output at that horizon.

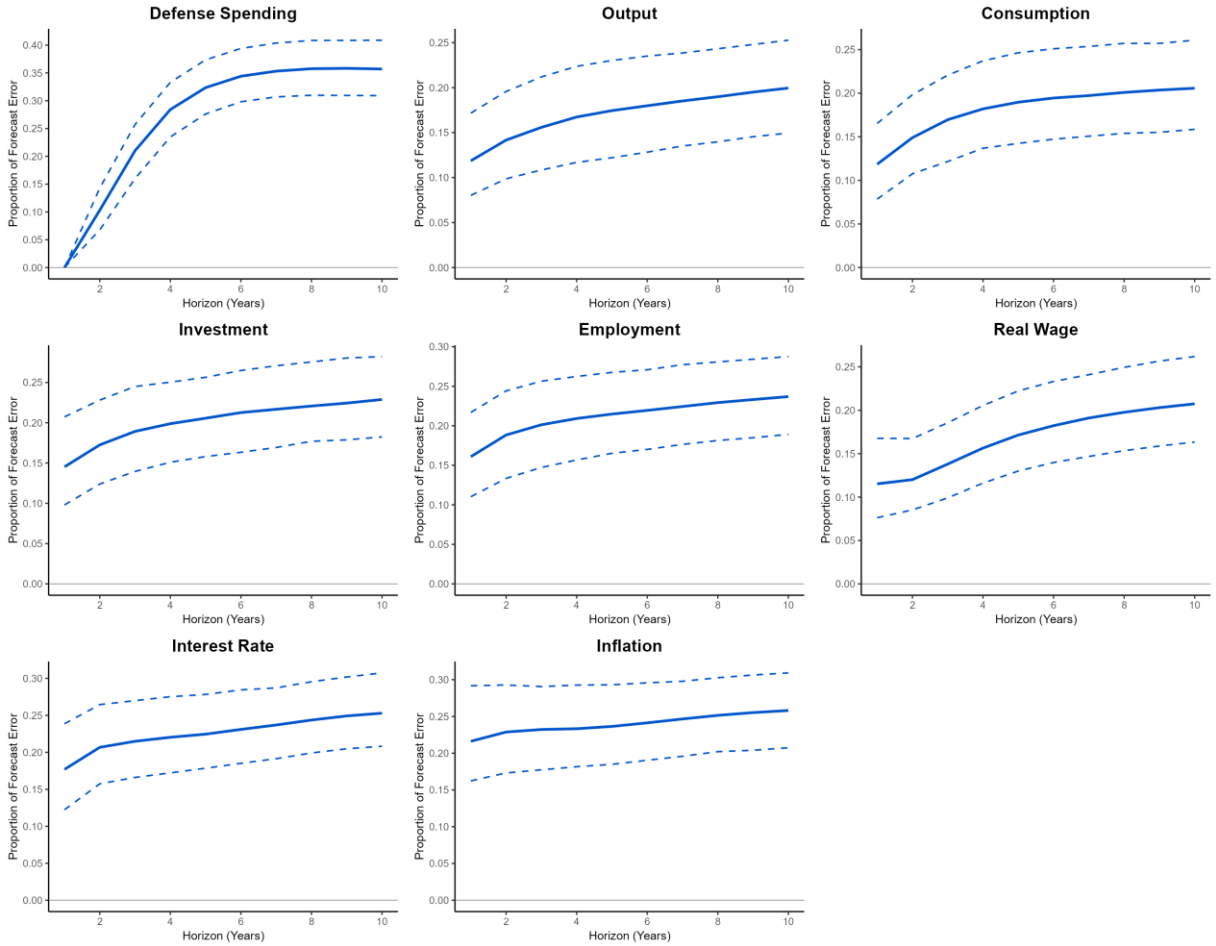
Figure 3 illustrates the impulse responses to a defense news shock in our baseline model. Our identified news shocks produce a significant and gradual rise in defense spending, peaking at around six percent three years after the initial shock. Aggregate output rises significantly on impact and reaches a peak of 0.69 percent around year two, where it remains essentially flat through year three. Both consumption and investment are also expanding. The labor market response is broadly positive: employment rises significantly on impact and peaks at 0.51 percent in year one. The real wage follows a hump shape, with an impact response that is statistically indistinguishable from zero before rising to a peak of 0.76 percent in year four. Inflation is observed to rise on impact, as indicated by a credible interval of [0.001, 0.76], which marginally excludes zero. The interest rate correspondingly rises by 21 basis points on impact and a peak of around 31 basis points in year five.



Note: Median impulse responses to a defense news shock, with 95% credible bands. Log-variable responses are reported as percentage deviations; interest rates are basis-point deviations and inflation is percentage-point deviation.

Figure 3: Baseline impulse responses to a defense news shock

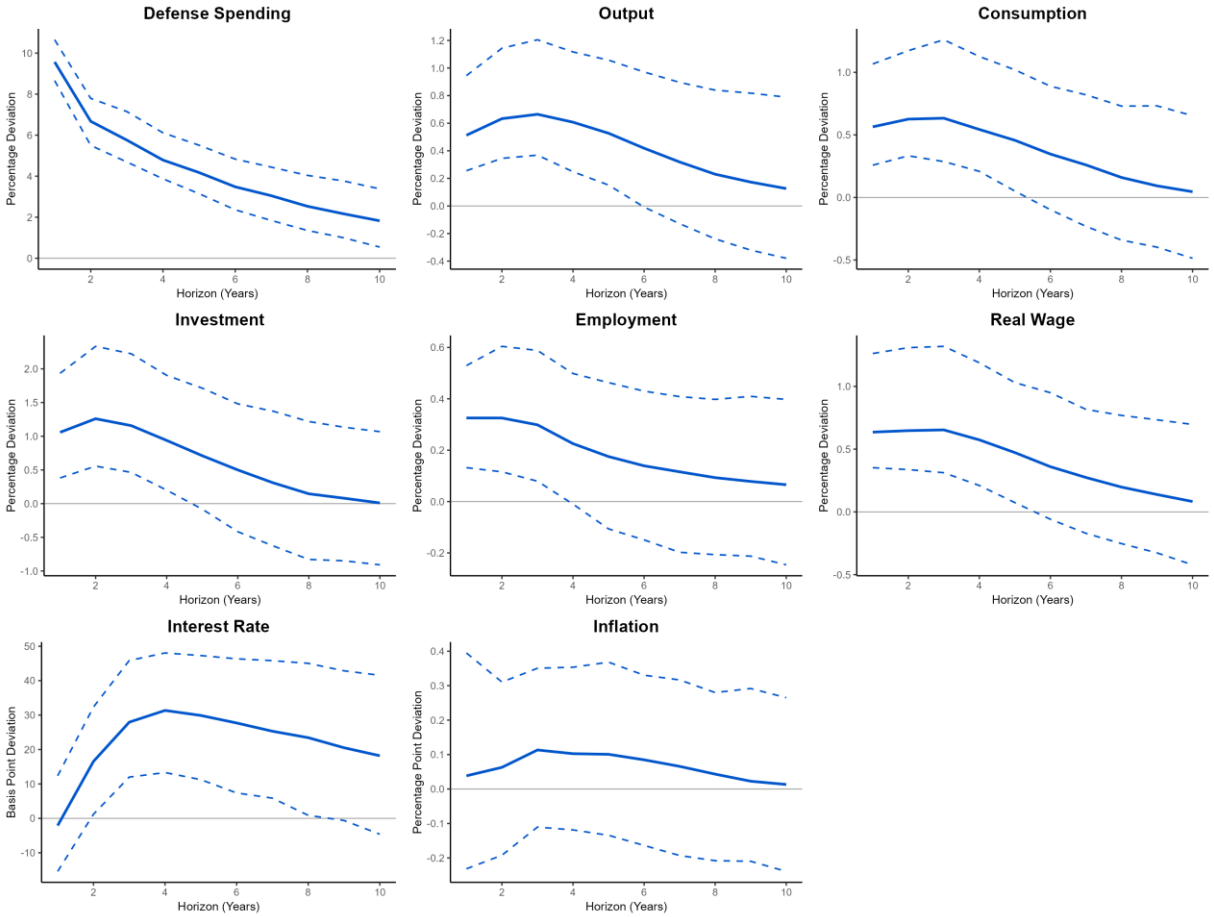
Figure 4 displays the forecast error variance decompositions for the defense news shock. The shock contributes increasingly to defense spending, rising from zero at impact (by construction) to approximately 36 percent at the ten-year horizon. News shocks also account for a moderate share of variance in other variables: output, consumption, investment, and employment each reach around 20 percent by year ten. For inflation, the initial share is approximately 21 percent, increasing to 27 percent by year nine. The interest rate contribution rises from about 13 percent to 26 percent over the same period, while the real wage contribution increases from roughly 12 percent to 21 percent. The credible bands' lower bounds remain consistently above zero for all variables. The news shock as such contributes meaningfully to macroeconomic variation across European economies.



Note: Median forecast error variance contributions of the defense news shock, with 95% credible bands.

Figure 4: Forecast error variance decomposition for the defense news shock.

Following Ben Zeev et al. (2026), we also estimate responses to defense surprise shocks, of which the IRFs are displayed in Figure 5. We find that aggregate output rises significantly on impact and peaks at 0.66 percent after one year. However, the response decays rapidly, falling to near zero by year four and becoming statistically insignificant by year 6. Consumption and investment both exhibit similar patterns. The labor market response is initially positive; employment rises by 0.33 percent, and the real wage by 0.64 percent, on impact. Nominal indicators exhibit weak responses. Inflation rises by only 0.04 percentage points on impact and remains insignificant across all horizons. The interest rate has a median of negative 2 basis points on impact, with credible bands containing zero. It then rises to a peak at around 31 basis points at horizon four. Overall, the surprise shock generates a short-lived expansion in real activity that becomes insignificant after four to six years.



Note: Median impulse responses to a defense surprise shock, with 95% credible bands. Log-variable responses are reported as percentage deviations; interest rates are basis-point deviations and inflation is percentage-point deviation.

Figure 5: Impulse responses to a defense surprise shock.

5.2 Multipliers

Table 3 presents the hierarchically pooled multipliers from our extended European sample, along with corresponding credibility bands. In contrast to the findings reported by Ben Zeev et al. (2026), who find European defense news shocks result in multipliers significantly above zero, the estimates here are substantially less precise. For our baseline sample of 35 economies, the impact multiplier has a median of 0.90, with the subsequent cumulative multiplier rising in horizon two, before declining and reaching 0.71 by H10. The 95 percent credible intervals are wide, including zero at every horizon. As such, these estimates cannot be interpreted as evidence of a positive output effect. Surprise shocks imply a statistically significant above unity response at impact and horizon 2, with medians rising from 3.45 to 4.87, respectively. The median multiplier continues to increase at longer horizons, peaking at 6.68 at horizon 8. However, inference at these horizons is weak due to very wide credible intervals that include zero.

Across subgroups, the estimates point to some heterogeneity, although the high posterior uncertainty limits the extent to which we can infer definitive differences. A notable contrast is

between the NATO and membership-timed non-NATO samples.[6] For NATO countries, the multiplier is close to the baseline Europe estimate and remains relatively stable across all horizons, declining from 0.90 at H1 to 0.75 at H10. By contrast, the non-NATO multiplier is substantially larger in the short run, at 1.94 at H1 and 1.68 at H2, but declines sharply thereafter, falling to 1.05 at H4, 0.19 at H8, and -0.14 at H10.

The EU-27 estimates are closer to the NATO and baseline results, with median multipliers of 0.87 at H1, 0.91 at H2, and 0.70 at H10. This is to be expected given that it constitutes a sample majority. The regional split also suggests some variation: Central Europe has relatively low short-run effects but higher medium-run estimates, rising from 0.27 at H1 to 1.69 at H4 and 1.63 at H8, whereas Western Europe sees smaller multipliers, declining from 1.04 at H1 to 0.48 at H10.[7]

Table 3: Multipliers of defense shocks across specifications and horizons

Specification	H1	H2	H4	H8	H10
Baseline Europe	0.90	0.99	0.98	0.81	0.71
	[-0.58, 4.62]	[-0.51, 3.73]	[-1.50, 3.40]	[-1.74, 3.19]	[-1.19, 2.67]
Baseline Europe (surprise)	3.45	4.87	5.40	6.68	5.83
	[1.93, 5.30]	[2.84, 6.87]	[-0.09, 10.47]	[-13.93, 65.00]	[-36.75, 78.07]
NATO	0.90	0.85	0.82	0.81	0.75
	[-0.87, 2.97]	[-0.63, 3.65]	[-0.35, 3.24]	[-0.77, 2.27]	[-1.00, 2.07]
Non-NATO	1.94	1.68	1.05	0.19	-0.14
	[-2.34, 8.88]	[-2.65, 10.56]	[-4.34, 9.35]	[-5.14, 5.69]	[-5.52, 6.32]
EU-27	0.87	0.91	0.85	0.78	0.70
	[-0.48, 3.65]	[-0.56, 3.28]	[-2.09, 2.89]	[-2.40, 2.30]	[-1.92, 1.97]

[6] To clarify, for the membership-timed non-NATO row: countries that are currently non-NATO members contribute data from their entire available sample period. Countries that joined NATO late are included only for the years prior to their accession, with data ending the year before they became NATO members. This approach ensures that the analysis for late joiners reflects solely their non-NATO status, allowing for a cleaner comparison between consistently non-NATO countries and those that transitioned to NATO membership. Due to insufficient time-series length, all countries that joined the alliance after the fifth accession round on March 29, 2004, are excluded from the NATO-group.

[7] Central Europe and Western Europe compositions are based on SIPRI definitions. The Central Europe sample consists of Albania, Bulgaria, Croatia, Czechia, Estonia, Hungary, Latvia, Lithuania, Montenegro, North Macedonia, Poland, Romania, Serbia, the Slovak Republic, and Slovenia. The Western Europe sample consists of Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Malta, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom. The regional split excludes Türkiye.

Specification	H1	H2	H4	H8	H10
Central Europe	0.27	0.98	1.69	1.63	1.44
	[-2.86, 7.07]	[-3.42, 6.38]	[-3.88, 5.14]	[-3.04, 4.30]	[-2.42, 3.90]
Western Europe	1.04	0.96	0.77	0.57	0.48
	[0.01, 2.54]	[-0.03, 2.59]	[-0.41, 2.51]	[-1.03, 1.81]	[-1.32, 1.78]

Notes: Median multiplier on first line; 95 percent credible intervals in brackets below. NATO and non-NATO rows use membership-timed samples. Results refer to news shocks for all multipliers except for the second row.

5.3 Import dependence

As seen in Figure 6, the aggregate import-intensity split indicates that low-import-intensity economies have defense-news multipliers significantly above zero up until H8: 1.05 at H1 [0.35, 2.09], 1.08 at H2 [0.36, 2.12], 1.00 at H4 [0.28, 2.03], and 0.89 at H8 [0.06, 1.80], before becoming statistically insignificant at H10. High-import-intensity economies have lower median multipliers at every horizon, ranging from 0.42 to 0.73, but the 95 percent intervals are wide and include zero across all horizons. The military-import-reliance split yields a similar descriptive pattern, with higher low-reliance medians at short horizons but very wide intervals for the high-reliance group. Thus, although low import dependence is associated with positive and economically meaningful defense news multipliers, the overlapping intervals prevent inference that the low- and high-import-dependence groups are statistically different.

An investigation into the temporal stability of country-group splits is necessary. These splits are determined by country-averages, while import dependence may fluctuate over time. If a country has frequently crossed between the high- or low-import-dependence threshold in the past, the average-based split is an unreliable proxy for import dependence. In our sample, this issue is less pronounced for aggregate import intensity, as only 10.1 percent of country-years are classified differently from the country's average-based group. In contrast, the problem is more significant for military import reliance, where 19.5 percent of country-years do not align with the static split. Further, 44.1 percent of countries in this group correspond to their static group in fewer than 80 percent of all its observed years. This indicates that the aggregate import-intensity split remains relatively stable, but the military-import-reliance split should be interpreted with great caution. Figure 13 in appendix A.2 shows a comparison of the IRFs between the different groupings.

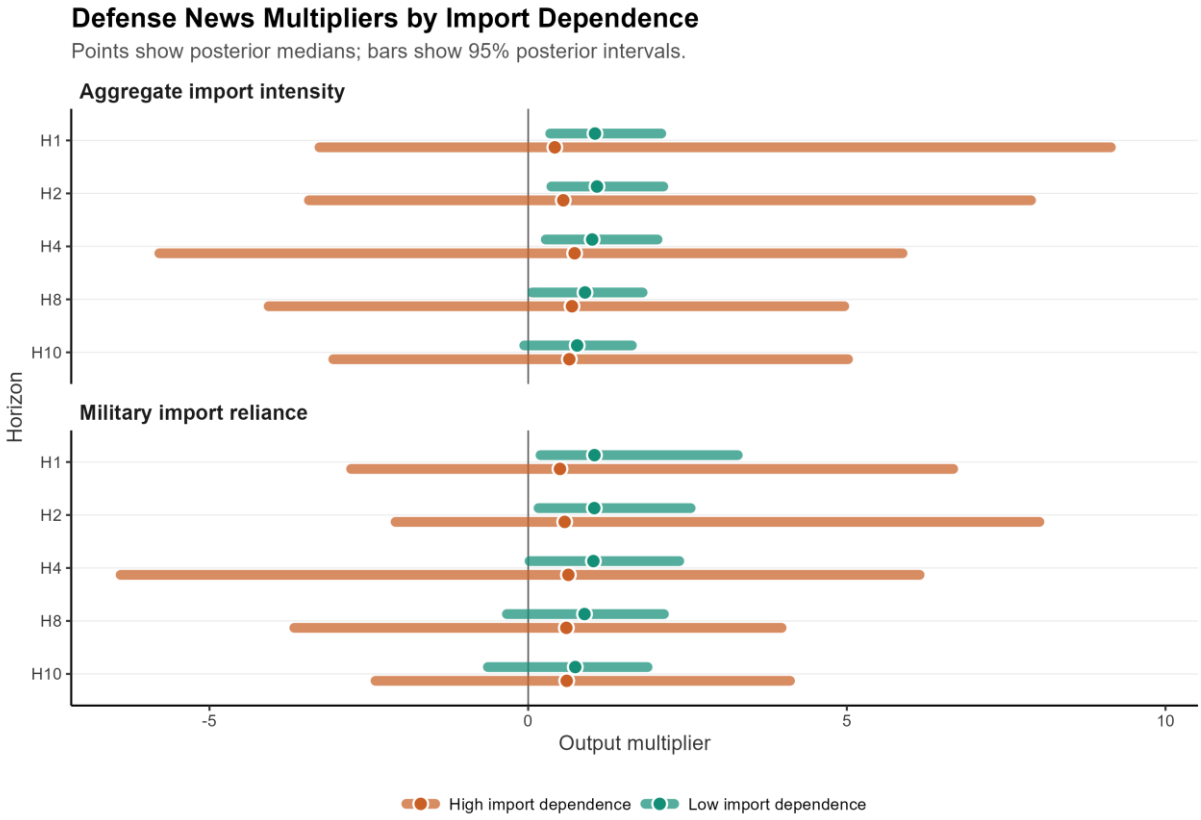


Figure 6: Defense News Multipliers by Import Dependence.

5.4 Robustness checks

Following Ben Zeev et al. (2026, p.13-16), we conduct multiple robustness checks of our baseline findings. These include changing the identification horizon, the sample period, the definition of defense spending, as well as accounting for confounding variables such as oil price, geopolitical risk and terrorist attacks. Overall, the robustness checks reported in appendix A.1 indicate that the baseline results are robust. The impulse responses and multiplier estimates are broadly stable when changing the MFEV truncation horizon, adding oil prices, controlling for terrorist fatalities, and adding geopolitical risk (GPR) to the SVAR. In the case of GPR, this does not imply that defense news is unrelated to geopolitical events, but rather that measured geopolitical risk does not appear to drive the estimated macroeconomic responses. Overall, these results suggest that defense news shocks are not primarily capturing confounding influences. Some differences do however arise. The shorter MFEV horizon and the pre-COVID-19 sample both yield lower median multipliers. The COFOG defense spending definition produces larger short-run multipliers than the SIPRI/NATO-style baseline. Taken together, the robustness checks support the main IRF patterns in our baseline. However, multiplier magnitudes are sensitive to the recent post-2019 period and to the definition of defense expenditure.

6 Discussion

This section discusses the preceding results and gives an economic interpretation of the empirical findings. First, we evaluate the methodological validity of our estimates and outline the main limitations of our analysis. Secondly, we compare our estimated multipliers to other findings within the recent defense spending literature. Finally, we relate the baseline results to fiscal multiplier theory and fiscal foresight effects.

6.1 Validity and Limitations

This subsection discusses the implementation of our method and the validity of our inference, before proceeding with internal and external validity. For discussions on additional limitations of this thesis, we refer to appendix A.4.

6.1.1 Estimation and Inference Validity

Before interpreting our results in the context of previous literature and economic theory, we must first establish the methodological validity of our model. This subsection therefore discusses whether our results are consistent with a correct implementation of the method. We therefore assess whether the estimates align with the patterns expected from a correctly implemented model. We also compare our findings to those presented in Ben Zeev et al. (2026) by running our model on a similar sample. The result of this comparison is presented in Table 4. The following assessment implies a correct implementation of our method, but wider credible intervals and larger surprise shocks constitute discrepancies.

Table 4: Defense news and surprise shocks; medians and 95 percent credible intervals by horizon

Shock	Horizon	Ben Zeev et al. median	Ben Zeev et al. 95 percent CI	Our 16-country median	Our 16-country 95 percent CI	Median diff.
News	**H1**	1.24	[0.81, 1.68]	1.17	[0.11, 2.17]	-0.07
News	**H2**	1.21	[0.83, 1.62]	1.02	[0.13, 2.01]	-0.19
News	**H4**	1.06	[0.66, 1.46]	0.86	[0.09, 1.94]	-0.20
News	**H8**	0.76	[0.33, 1.24]	0.80	[-0.03, 1.70]	0.04
News	**H10**	0.63	[0.15, 1.14]	0.87	[-0.08, 1.77]	0.24
Surprise	**H1**	0.49	[0.18, 0.80]	2.26	[0.14, 4.43]	1.77
Surprise	**H2**	0.54	[0.10, 1.00]	3.92	[1.62, 6.71]	3.38

Shock	Horizon	Ben Zeev et al. median	Ben Zeev et al. 95 percent CI	Our 16-country median	Our 16-country 95 percent CI	Median diff.
Surprise	**H4**	-0.15	[-1.16, 0.82]	5.09	[2.00, 8.69]	5.24
Surprise	**H8**	-2.39	[-4.92, 0.00]	1.93	[-3.35, 6.55]	4.32
Surprise	**H10**	-3.38	[-6.82, -0.18]	0.06	[-7.76, 7.19]	3.44

Notes: Ben Zeev et al. values are taken from Table 1 in Ben Zeev et al. (2026, p.11-12).

The IRFs and FEV graphs (Figures 3–5) in section 5.1 broadly validate a correct application of the methodology. Firstly, the IRFs of defense news and surprise news both display the expected pattern, most clearly visible in the respective shock's induced effect on the defense spending variable, with the former resulting in a delayed, gradual response, and the latter in an immediate effect on defense spending. Secondly, the forecast error variance (FEV) contributions confirm that defense news explains a substantial share of the variation in military spending (peaking at 36 percent). This result closely mirrors that of Ben Zeev et al. (2026, p. 8), where a corresponding figure of 35 percent was reported. Overall, the results reported in these figures follow similar patterns as those by Ben Zeev et al. (2026). In addition, the medians of our fiscal news multipliers are reasonably close to their paper, particularly if we estimate a model where our data and sample adhere closely to their 16-country panel.[8]

However, while our IRFs and FEV graphs look reasonable and consistent with the methodology, the credible intervals of our fiscal multipliers are very wide. These intervals, reported in Table 3, include zero across most horizons and specifications. Posterior uncertainty for objects of interest is caused by two metrics in our method, as explained in section 4.6.2 (see equation 10). These are the uncertainty in the estimated multiplier for each country, and the heterogeneity between countries. In the extended 35-country sample, several added countries have shorter and more volatile time series, which increases within-country uncertainty. Cross-country heterogeneity also increases with our more dispersed sample. This helps explain why credible bands are wider than in the original 16-country setting. Regarding the multipliers, the calculation includes a ratio. It rescales the cumulative output response by the cumulative defense spending response, including the $\frac{\bar{Y}_i}{\bar{DE}_i}$ ratio. As a result, uncertainty in the output and defense spending IRFs is amplified, so multiplier bands become much wider even though the IRFs have reasonable credible bands.

[8] Discrepancies from Ben Zeev et al. (2026) likely remains between their sample and our replication attempt. This stems from ambiguities in their data construction. As one example, Ben Zeev et al. describe their baseline unbalanced panel as having "a total of 859 observations, with an average number of 51.1 observations across our 16 countries," but 859 observations over 16 countries implies an average of 53.7 observations (unsurprising given that Ben Zeev et al. (2026) is still a working paper). This and similar ambiguities make it difficult to exactly and completely replicate their findings.

Beyond this, our 16-country specification still yields larger credible bands than those reported by Ben Zeev et al. (2026, p.11); ours are about 2.15 times wider on average. The remaining gap suggests that unreported details pertaining to the implementation of how the posterior simulation and handling of draws may affect the width of the credible interval. Ben Zeev et al. (2026) never comment on the handling of unstable posterior draws, but this approach is standard in SVAR literature and follows economic theory as explained in section 4.6.1. Similarly, our 16-country sample yields much larger credible intervals for the surprise multipliers. Compared to Ben Zeev et al. (2026, p.11), our intervals are about 4.04 times larger. For this category, our median multipliers are also much larger and more persistent, owing to an output response function that is stronger on impact and longer lasting relative to our corresponding defense spending response.[9] The reason that our identified surprise shocks result in stronger output responses remains unclear. The comparison with Ben Zeev et al. (2026) suggests that the defense surprise shocks may be less reliably identified, so we treat these results cautiously and limit the focus on surprise shocks.

Overall, this assessment suggests that the model is correctly implemented in its main features, particularly for the identification of defense news shocks. The IRFs, FEVs, and 16-country news multiplier medians are broadly consistent with Ben Zeev et al. (2026), which support the validity of the estimation procedure. However, the wide credible intervals and the larger surprise shock multipliers indicate reason for cautious interpretation. We therefore regard our findings as suitable for identifying the direction and timing of the defense news shock effects, particularly as conveyed in the impulse response functions, but less suited for drawing precise conclusions about the exact size of fiscal multipliers or statistically significant differences between them.

6.1.2 Internal Validity

First, AMECO data uses a fixed euro conversion for data collected before countries' Eurozone accession. Within-country time series are consistent, but the conversion limits the comparability of pre-euro absolute values across countries. This is unlikely to materially affect our country-level SVAR identification, but it prevents us from interpreting raw cross-country level differences. Second, our baseline model does not include real exchange rates meaning changes in the purchasing power of defense budgets and imported procurement costs are absorbed into the aggregate macroeconomic responses rather than separately identified. Third, the MFEV approach identifies defense news shocks statistically rather than narratively. Unlike event-based news measures such as Ramey's (2011), our shocks cannot be tied directly to the specific

[9] The pooled output response to a defense-surprise shock in our 16-country sample is 0.24 percent at H1, rises to 0.47-0.48 percent at H2-H3, and remains positive through H9 before reaching 0.01 percent at H10; the corresponding 95 percent credible intervals are (0.05, 0.45), (0.26, 0.72), (0.23, 0.77), (0.09, 0.72), (-0.08, 0.66), (-0.22, 0.58), (-0.34, 0.58), (-0.42, 0.56), (-0.47, 0.53), and (-0.52, 0.54).

announcements or geopolitical events that may have caused them. The aforementioned internal validity limitations are discussed more thoroughly in appendix A.4.

6.1.3 External Validity

The ability to generalize our findings to a current European setting is limited by the extensive scale and scope of current rearmament efforts. The empirical material underlying most of our results consists of predominantly smaller and more national armament efforts, whereas the current European defense buildup is larger and more simultaneous across countries. As Furceri et al. (2026, p. 8) note, historical estimates may therefore not apply fully to today. For example, the multipliers may be lower if monetary policy is less accommodative, as today's large scale rearmament efforts could generate stronger inflationary pressure. Additionally, the linear VAR model we employ treats defense spending shocks as if small and large increases have the same proportional effect on output. This is a restrictive assumption, and an established critique in the fiscal multiplier literature is that such proportionality need not hold, since multipliers may vary across macroeconomic circumstances (Morelli, 2025, p. 16; Ramey, 2019, pp. 90–91).

6.2 Comparison with Literature

6.2.1 Multipliers

We find that median news shock multipliers are generally on the lower side of the recent defense spending literature, as outlined in Table 5. Ben Zeev et al. (2026), estimating defense news shocks for 16 Western European economies, find stronger and more precise effects, with baseline multipliers of 1.24 on impact, 1.21 after two years and 1.06 after four years. Our corresponding estimates for the 35-country baseline sample are lower at 0.90, 0.99, and 0.98, as well as much less precise. Similarly, our findings fall below the recent EU-27 estimates, which are based on alternative identification strategies. García-Serrador et al. (2025) use a local projections method and estimate an EU-27 defense multiplier of about 1.4 within one year, peaking at 1.6 in year two. Furceri et al. (2026) report cumulative multipliers of about 1.5 at 12 months, 1.7 after two years and 1.9 after three years using an instrumental-variables approach based on monthly defense procurement contracts. In comparison, our EU-27 news-shock estimates are 0.87 on impact and 0.91 after two years, again with wide credible intervals. Our results therefore sit closer to the lower range discussed by Ilzetzki (2025), whose calculations suggest that European defense buildups may imply multipliers around 0.6 to 1.0. The wide credible bands of our news shock multiplier estimates prevent a strong conclusion about the expansionary or contractionary effects that may be associated with defense spending.

Table 5: Comparison of Short-Run European Defense Multipliers

Study	Sample / context	Estimated multiplier
This thesis (Baseline)	35 European economies; defense news shock; Bayesian SVAR with hierarchical pooling	Impact: 0.90; Year 2: 0.99
This thesis (EU-27 sample)	EU-27; defense news shock; Bayesian SVAR with hierarchical pooling	Impact: 0.87; Year 2: 0.91
Ben Zeev et al. (2026)	16 Western European economies; defense news shock; Bayesian SVAR with hierarchical pooling	Impact: 1.24; Year 2: 1.21
Garcia-Serrador et al. (2025)	EU-27; Local projections	Year 1: approx. 1.4; Year 2: approx. 1.6
Furceri et al. (2026)	EU-27; monthly defense procurement contracts; instrumental variables	12 months: approx. 1.5; 24 months: approx. 1.7; 36 months: approx. 1.9
Ilzetzki (2025)	European defense-buildup scenario calculations	Implied range: 0.6 to 1.0

Notes: "This thesis" report median defense news multipliers from Table 3. Ben Zeev et al. (2026) values are their baseline Europe news shock multipliers. Garcia-Serrador et al. (2025) report approximate linear local-projection multipliers of 1.4 within one year and 1.6 in year two. Furceri et al. (2026) report instrumental-variable monthly procurement multipliers of about 1.5 at 12 months, 1.7 at 24 months, and 1.9 at 36 months. Ilzetzki (2025) gives an implied range based on a Europe-wide scenario in which defense spending rises from 2.0 to 3.5 percent of GDP and GDP rises by 0.9 to 1.5 percent.

6.2.2 Institutional and Regional Heterogeneity

The large differences in multiplier estimates, as well as deviations in the regionally divided groupings of IRFs (see Figure 14 in appendix A.2), reinforce previous findings of highly heterogenous economic responses to defense news across Europe (Ben Zeev et al., 2026, p. 2). As mentioned by Ramey (2019, p.91-92), the output response to government spending depends on a plethora of factors, all of which can plausibly vary across regional groupings. However, two distinct patterns stand out.

First, NATO responses follow the baseline sample closely, while the membership-timed non-NATO multiplier is higher in the short-run but is not matched by stronger output IRFs. This contrasts with Ben Zeev et al. (2026, pp. 11, 19), who find larger effects for NATO countries and interpret this as evidence that alliance commitments, capability targets and coordinated defense planning may make defense announcements more credible. Our results suggest that this NATO/non-NATO divide is not robust to a broader European sample. One likely reason is sample composition: Ben Zeev et al. (2026) use a 16-country Western European panel, whereas

our 35-country sample includes a wider set of European economies and classifies NATO status using membership-timed samples. Although overlapping credible bands prevent inference of a statistically significant difference, the median patterns suggest that the previously documented NATO divide weakens once recent NATO entrants and a larger set of non-NATO countries are included.

Second, Western and Central European fiscal multipliers exhibit opposite trends over their horizons. Western Europe starts with a relatively larger short-run multiplier, but the estimate declines steadily over time, while Central Europe begins with a low impact multiplier and rises to its highest values at medium horizons. The larger absolute defense spending and output responses in the Central European sample may simply be driven by a much shorter average sample length, roughly half that of Western Europe.[10] This shorter sample then places a larger relative emphasis on the more recent, large defense buildup news, leading to stronger absolute magnitude impulse responses.

These divergent fiscal multiplier patterns between the two groups are difficult to interpret conclusively. A plausible explanation consistent with the observed results would be variation in government credibility. Our Central European sample comprises several transition economies characterized by lower institutional capacity and fiscal space, particularly in the immediate post-Cold War period. If private agents lack confidence in the government's ability to execute announced fiscal plans, they may not have as large initial consumption and investment responses as those we observe in Western Europe. Consequently, the private sector only reacts with significant magnitude in a more lagged manner, once the announced defense spending materializes and is assumed to persist.

6.2.3 Import Dependence Heterogeneity

While our estimates examine heterogenous responses across groups and do not constitute causal effects, the direction of the estimates is consistent with the open-economy leakage argument (Barrell, Holland and Hurst, 2012, pp. 12-13). The expectation that fiscal multipliers should be smaller where a larger share of demand falls on imports rather than domestic production aligns with our overall findings. It also aligns qualitatively with García-Serrador et al. (2025, pp. 31-32, 39-40), who find larger defense multipliers when reliance on imported military goods is low, and with Furceri et al. (2026, p. 17), who report similar results.

[10] The Central European countries have an average estimation sample of 26.5 years, compared with 55.5 years for the Western European countries.

Compared with earlier literature, our method yields much less precise estimates. The estimates point in the same direction, but the high-import-dependence group has too imprecise credible bands to conclude that a significant difference exists. Further, the sample-split methodology required by our SVAR model is time-invariant but investigates variables that may change substantially over time. This limits robust inference. Figure 13 in appendix A.2 shows the IRFs do not indicate substantial differences between how the macroeconomic variables in the high- and low-import-dependence countries respond to defense news shocks. Overall, the results presented should be interpreted merely as suggestive evidence that corroborates earlier findings.

6.3 Theoretical Implications

The IRFs show an expansion of real activity immediately following defense news shocks, with responses staying significantly above zero over multiple horizons. Output rises on impact and remains significantly above its counterfactual path for several years, while consumption, investment, and employment also display positive responses (the significance of investment and employment is concentrated in the earlier horizons). Real wages initially respond weakly and are slightly negative on impact. They subsequently rise and become significantly positive after the first period, suggesting that the stronger labor demand eventually places upward pressure on real wages.

The positive investment and consumption responses offer some empirical background against which to evaluate the previously discussed theories. A standard neoclassical model would suggest that the wealth effect of anticipated future taxation crowds out the private sector, as Ricardian consumers anticipate future taxation and therefore reduce current consumption (Woodford, 2011, pp. 2–5). This explanation appears inconsistent with our findings, which instead shows a significant and persistent rise in consumption occurring immediately in response to the defense news shocks. The predominant effect therefore appears to be that consumers respond to defense news as a signal of higher future income and demand, while any consumption-reducing effect from anticipated future taxation is either weak or outweighed by this channel. Our results are instead consistent with an aggregate demand shock. The increase in consumption suggests that private demand expands alongside expected future government demand, while the rise in employment is consistent with firms increasing production and labor demand.

Finally, our results align with literature on fiscal foresight. As seen in Figure 3, output, consumption, and investment respond immediately to defense news shocks, even though defense spending only materializes at later horizons. By extending the sample to a broader set of European economies, our results provide further evidence on the importance of accounting for these anticipation effects.

7 Conclusion

7.1 Summary of Findings

This thesis investigated the macroeconomic consequences of anticipated future defense spending in Europe. Using a novel data set of 35 European economies spanning seven decades, we estimated a Bayesian structural vector autoregression (SVAR) model with hierarchical pooling. By identifying defense news shocks through maximum forecast error variance over a five-year horizon, we accounted for the fiscal foresight of economic agents.

Our impulse response functions indicate that defense news shocks generate an immediate and significant response on several macroeconomic variables in our model. Specifically, output, consumption, investment, and employment rise on impact and remain above their counterfactual paths for several years. Furthermore, the identified defense news shocks explain approximately 36 percent of defense spending variation over a ten-year horizon. However, our estimated fiscal multipliers are generally lower than those found in recent literature and exhibit insignificant credible intervals. For the baseline panel of 35 European countries, the median cumulative news multiplier is 0.90 on impact, rising to 0.99 in year two, before declining to 0.71 by year ten. Because the credible bands include zero across most horizons, these estimates cannot be interpreted as definitive evidence of a positive output effect. Therefore, our fiscal multipliers do not indicate that the true Europe-wide average effect stimulates output enough to make government defense spending partially self-financing.

7.2 Conclusions

Despite the high posterior uncertainty in our multiplier estimates, the qualitative responses of macroeconomic variables offer important theoretical implications. The immediate and persistent rise in private consumption following a defense news shock contradicts standard neoclassical predictions, which suggest that the anticipation of future taxation should crowd out current consumption through a negative wealth effect. Instead, our findings are more consistent with an aggregate demand channel, where private demand expands alongside expected future government demand. Furthermore, our expanded 35-country panel reveals that broadening the geographic scope beyond Western Europe introduces substantial cross-country heterogeneity. This heterogeneity, particularly from shorter and more volatile data series in newer country additions, is the primary driver behind the wider credible bands observed in this study. The import-dependence results provide an additional source of heterogeneity. The aggregate import-intensity split indicates that low-import-intensity economies experience positive and economically

meaningful defense news multipliers over short and medium horizons, while high-import-intensity economies display lower median multipliers and substantially wider credible intervals.

7.3 Future Research and Policy Recommendations

Future research should extend this analysis in three main directions. First, narrative or event-based measures of European defense news should be constructed to validate the shock identification method and findings from current SVAR models. Second, non-linear SVAR specifications should be explored. Third, European countries should begin aggregating quarterly defense spending data, like the U.S. This would allow researchers to study macroeconomic effects of defense news shocks with greater precision.

Finally, our main findings suggest that policymakers ought to use prudence when interpreting the macroeconomic benefits of rearmament. The heterogeneous responses across Europe imply that results observed in one grouping of countries may not generalize well to others. The uncertainty surrounding our estimates similarly makes it difficult to draw firm conclusions about the magnitude of crowding-in effects. As a result, our findings suggest that policymakers should exercise caution when accounting for crowding-in effects to offset the fiscal costs of higher defense spending.

References

Antolin-Diaz, J. and Surico, P. (2025) 'The Long-Run Effects of Government Spending', American Economic Review, 115(7), pp. 2376-2413. doi: 10.1257/aer.20231278.

Arias, J.E., Rubio-Ramirez, J.F. and Waggoner, D.F. (2025) 'Uniform Priors for Impulse Responses', Econometrica, 93(2), pp. 695-718. doi: 10.3982/ECTA21101.

Auclert, A., Rognlie, M. and Straub, L. (2024) 'The Intertemporal Keynesian Cross', Journal of Political Economy, 132(12), pp. 4068-4121. doi: 10.1086/732531.

Bank of Albania (2026). *Interest rates*. [online] Available at: https://www.bankofalbania.org/Markets/Interest_rates/ (Accessed 5 May 2026).

Barrell, R., Holland, D. and Hurst, I. (2012) 'Fiscal Consolidation: Part 2. Fiscal Multipliers and Fiscal Consolidations', OECD Economics Department Working Papers, No. 933. Paris: OECD Publishing. doi: 10.1787/5k9fdf6bs78r-en.

Barro, R.J. and Redlick, C.J. (2011) 'Macroeconomic Effects from Government Purchases and Taxes', Quarterly Journal of Economics, 126(1), pp. 51-102. doi: 10.1093/qje/qjq002.

Beaudry, P. and Portier, F. (2006) 'Stock Prices, News, and Economic Fluctuations', American Economic Review, 96(4), pp. 1293-1307. doi: 10.1257/aer.96.4.1293.

Ben Zeev, N. and Pappa, E. (2017) 'Chronicle of a War Foretold: The Macroeconomic Effects of Anticipated Defence Spending Shocks', Economic Journal, 127(603), pp. 1568-1597. doi: 10.1111/ecoj.12349.

Ben Zeev, N., Pappa, E. and Scola Gagliardi, E. (2026) The Offensive Power of Defense News in Europe. Working paper, 26 February.

Blanchard, O. and Perotti, R. (2002) 'An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output', Quarterly Journal of Economics, 117(4), pp. 1329-1368. doi: 10.1162/003355302320935043.

Bokan, N., Jacquinot, P., Lalik, M., Muller, G., Priftis, R. and Rigato, R. (2025) 'Macroeconomic impacts of higher defense spending: a model-based assessment', ECB Economic Bulletin, 6/2025. Available at: https://www.ecb.europa.eu/press/economic-bulletin/articles/2025/html/ecb.ebart202506_01~d41c118e13.en.html (Accessed: 7 May 2026).

Born, B., Juessen, F. and Muller, G.J. (2013) 'Exchange rate regimes and fiscal multipliers', Journal of Economic Dynamics and Control, 37(2), pp. 446-465. doi: 10.1016/j.jedc.2012.09.014.

Brinca, P., Faria-e-Castro, M., Ferreira, M.H., Holter, H.A. and Nobrega, V. (2025) 'The nonlinear effects of fiscal policy', Journal of Public Economics, 252, article 105517. doi: 10.1016/j.jpubeco.2025.105517.

Caldara, D. and Iacoviello, M. (2022) 'Measuring geopolitical risk', American Economic Review, 112(4), pp. 1194-1225. doi: 10.1257/aer.20191823. Data available at: https://www.matteoiacoviello.com/gpr.htm (Accessed: 7 May 2026).

CEIC Data (2026). *Global Economic Data, Indicators, Charts & Forecasts | CEIC*. [online] Available at: https://www.ceicdata.com/en (Accessed 5 May 2026).

Chao, J.C. and Phillips, P.C.B. (2002) 'Jeffreys prior analysis of the simultaneous equations model in the case with n + 1 endogenous variables', Journal of Econometrics, 111(2), pp. 251-283. doi: 10.1016/S0304-4076(02)00106-9.

Corsetti, G., Meier, A. and Muller, G.J. (2012) 'What Determines Government Spending Multipliers?', Economic Policy, 27(72), pp. 521-565. doi: 10.1111/j.1468-0327.2012.00295.x.

Dunne, J.P., Smith, R.P. and Willenbockel, D. (2005) 'Models of military expenditure and growth: a critical review', Defence and Peace Economics, 16(6), pp. 449-461. doi: 10.1080/10242690500167791.

European Central Bank (2026) ECB Data Portal: Euribor 3-month - historical close, average of observations through period, Euro area, monthly. Available at: https://data.ecb.europa.eu/data/datasets/FM (Accessed: 11 May 2026).

European Commission (2025) Future of European defense. Available at: https://commission.europa.eu/topics/defense/future-european-defense_en (Accessed: 7 May 2026).

European Commission (2026) AMECO database. Available at: https://economy-finance.ec.europa.eu/economic-research-and-databases/economic-databases/ameco-database_en (Accessed: 7 May 2026).

Eurostat (2025) General government expenditure by function (COFOG) (gov_10a_exp): Reference metadata. Available at: https://webgate.ec.europa.eu/eurostat/cache/metadata/en/gov_10a_exp_esms.htm (Accessed: 7 May 2026).

Federal Reserve Bank of St. Louis (n.d.) Spot Crude Oil Price: West Texas Intermediate (WTI) (WTISPLC) (dataset). FRED. Available at: https://fred.stlouisfed.org/series/WTISPLC (Accessed: 7 May 2026).

Fisher, J.D.M. and Peters, R. (2010) 'Using Stock Returns to Identify Government Spending Shocks', Economic Journal, 120(544), pp. 414-436. doi: 10.1111/j.1468-0297.2010.02355.x.

Furceri, D., Juarros, P., Mishra, S., Nguyen, A.D.M., Pessoa, A.S. and Sollaci, A. (2026) Macroeconomic Impacts of EU Defense Spending. IMF Working Paper No. 2026/053. Washington, DC: International Monetary Fund.

Garcia-Serrador, A., Sarasa-Flores, D. and Ulloa, C.A. (2025) Buy Guns or Buy Roses?: EU Defense Spending Fiscal Multipliers. BBVA Research Working Paper No. 25/06. Available at: https://www.bbvaresearch.com/wp-content/uploads/2025/06/Buy-Guns-or-Buy-Roses-EU-Defense spending-Fiscal-Multipliers-WP-25-06-1.pdf (Accessed: 7 May 2026).

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B. (2013) Bayesian Data Analysis. 3rd edn. Boca Raton, FL: CRC Press.

Ilzetzki, E. (2025) Guns and Growth: The Economic Consequences of Defense Buildups. Kiel Report No. 2. Kiel: Kiel Institute for the World Economy. Available at: https://www.ifw-kiel.de/publications/guns-and-growth-the-economic-consequences-of-defense-buildups-33747/ (Accessed: 7 May 2026).

Ilzetzki, E., Mendoza, E.G. and Vegh, C.A. (2013) 'How big (small?) are fiscal multipliers?', Journal of Monetary Economics, 60(2), pp. 239-254. doi: 10.1016/j.jmoneco.2012.10.011.

International Monetary Fund (2013) 'Dancing Together? Spillovers, Common Shocks, and the Role of Financial and Trade Linkages', in World Economic Outlook: Transitions and Tensions. October 2013. Washington, DC: International Monetary Fund, pp. 81-112. Available at: https://www.imf.org/~/media/Websites/IMF/Imported/external/pubs/ft/weo/2013/02/pdf/_c3pdf.ashx (Accessed: 11 May 2026).

Kilian, L. and Lutkepohl, H. (2017) Structural Vector Autoregressive Analysis. Cambridge: Cambridge University Press. doi: 10.1017/9781108164818.

Knight, M., Loayza, N. and Villanueva, D. (1996) 'The Peace Dividend: Military Spending Cuts and Economic Growth', IMF Staff Papers, 43(1), pp. 1-37.

Morelli, F. (2025) 'The Evolution of Spending Multiplier Economics: From Neoclassical to New Keynesian and Beyond', Journal of Economic Surveys, article joes.70034. doi: 10.1111/joes.70034.

Nakamura, E. and Steinsson, J. (2014) 'Fiscal Stimulus in a Monetary Union: Evidence from US Regions', American Economic Review, 104(3), pp. 753-792. doi: 10.1257/aer.104.3.753.

NATO (2024) Defence Expenditure of NATO Countries (2014-2024). Brussels: North Atlantic Treaty Organization. Available at: https://www.nato.int/content/dam/nato/legacy-wcm/media_pdf/2024/6/pdf/240617-def-exp-2024-en.pdf (Accessed: 11 May 2026).

NATO (2025) The Hague Summit Declaration, 25 June. Available at: https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2025/06/25/the-hague-summit-declaration (Accessed: 7 May 2026).

National Consortium for the Study of Terrorism and Responses to Terrorism (START) (2022) Global Terrorism Database, 1970-2020 (data file). Available at: https://www.start.umd.edu/gtd (Accessed: 7 May 2026).

OECD (2026) OECD Data Explorer. Available at: https://data-explorer.oecd.org/ (Accessed: 7 May 2026).

Olejnik, L.W. (2023) 'Economic growth and military expenditure in the countries on NATO's Eastern flank in 1999-2021', Eesti Pank Working Paper Series, No. 2/2023. doi: 10.23656/25045520/022023/0199.

Our World in Data (2023) Terrorism deaths (dataset), based on START Global Terrorism Database, with major processing by Our World in Data. Available at: https://ourworldindata.org/grapher/terrorism-deaths (Accessed: 7 May 2026).

Ramey, V.A. (2011) 'Identifying Government Spending Shocks: It's all in the Timing', Quarterly Journal of Economics, 126(1), pp. 1-50. doi: 10.1093/qje/qjq008.

Ramey, V.A. (2016) 'Macroeconomic Shocks and Their Propagation', in Taylor, J.B. and Uhlig, H. (eds) Handbook of Macroeconomics. Vol. 2. Amsterdam: Elsevier, pp. 71-162. doi: 10.1016/bs.hesmac.2016.03.003.

Ramey, V.A. (2019) 'Ten Years After the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?', Journal of Economic Perspectives, 33(2), pp. 89-114. doi: 10.1257/jep.33.2.89.

Ramey, V.A. and Shapiro, M.D. (1998) 'Costly Capital Reallocation and the Effects of Government Spending', Carnegie-Rochester Conference Series on Public Policy, 48, pp. 145-194. doi: 10.1016/S0167-2231(98)00020-7.

Ramey, V.A. and Zubairy, S. (2018) 'Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data', Journal of Political Economy, 126(2), pp. 850-901. doi: 10.1086/696277.

Rubio-Ramirez, J.F., Waggoner, D.F. and Zha, T. (2010) 'Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference', Review of Economic Studies, 77(2), pp. 665-696. doi: 10.1111/j.1467-937X.2009.00578.x.

Shieh, J.Y., Lai, C.C. and Chang, W.Y. (2002) 'The Impact of Military Burden on Long-Run Growth and Welfare', Journal of Development Economics, 68(2), pp. 443-454. doi: 10.1016/S0304-3878(02)00022-6.

Sims, C.A. (1980) 'Macroeconomics and Reality', Econometrica, 48(1), pp. 1-48. doi: 10.2307/1912017.

Sims, C.A. and Zha, T. (1998) 'Bayesian Methods for Dynamic Multivariate Models', International Economic Review, 39(4), pp. 949-968. doi: 10.2307/2527347.

SIPRI (2026) SIPRI Military Expenditure Database. Stockholm International Peace Research Institute. Available at: https://www.sipri.org/databases/milex (Accessed: 11 May 2026). doi: 10.55163/CQGC9685.

SIPRI (n.d.) Definition of military expenditure. Available at: https://www.sipri.org/databases/milex/definitions (Accessed: 7 May 2026).

Stock, J.H. and Watson, M.W. (2001) 'Vector Autoregressions', Journal of Economic Perspectives, 15(4), pp. 101-115. doi: 10.1257/jep.15.4.101.

Uhlig, H. (2005) 'What are the effects of monetary policy on output? Results from an agnostic identification procedure', Journal of Monetary Economics, 52(2), pp. 381-419. doi: 10.1016/j.jmoneco.2004.05.007.

United Nations Department of Economic and Social Affairs, Statistics Division (UN DESA/UNSD) (n.d.) United Nations Comtrade Database. Available at: https://comtradeplus.un.org/ (Accessed: 11 May 2026).

Woodford, M. (2011) 'Simple Analytics of the Government Expenditure Multiplier', American Economic Journal: Macroeconomics, 3(1), pp. 1-35. doi: 10.1257/mac.3.1.1.

World Bank (2026a) Imports of goods and services (% of GDP) (NE.IMP.GNFS.ZS) (dataset). World Development Indicators. Available at: https://data.worldbank.org/indicator/NE.IMP.GNFS.ZS (Accessed: 11 May 2026).

World Bank (2026b) Military expenditure (% of GDP) (MS.MIL.XPND.GD.ZS) (dataset). World Development Indicators, based on SIPRI Military Expenditure Database. Available at: https://data.worldbank.org/indicator/MS.MIL.XPND.GD.ZS (Accessed: 7 May 2026).

Appendices

Appendix A: Supplementary Materials

A.1 Robustness

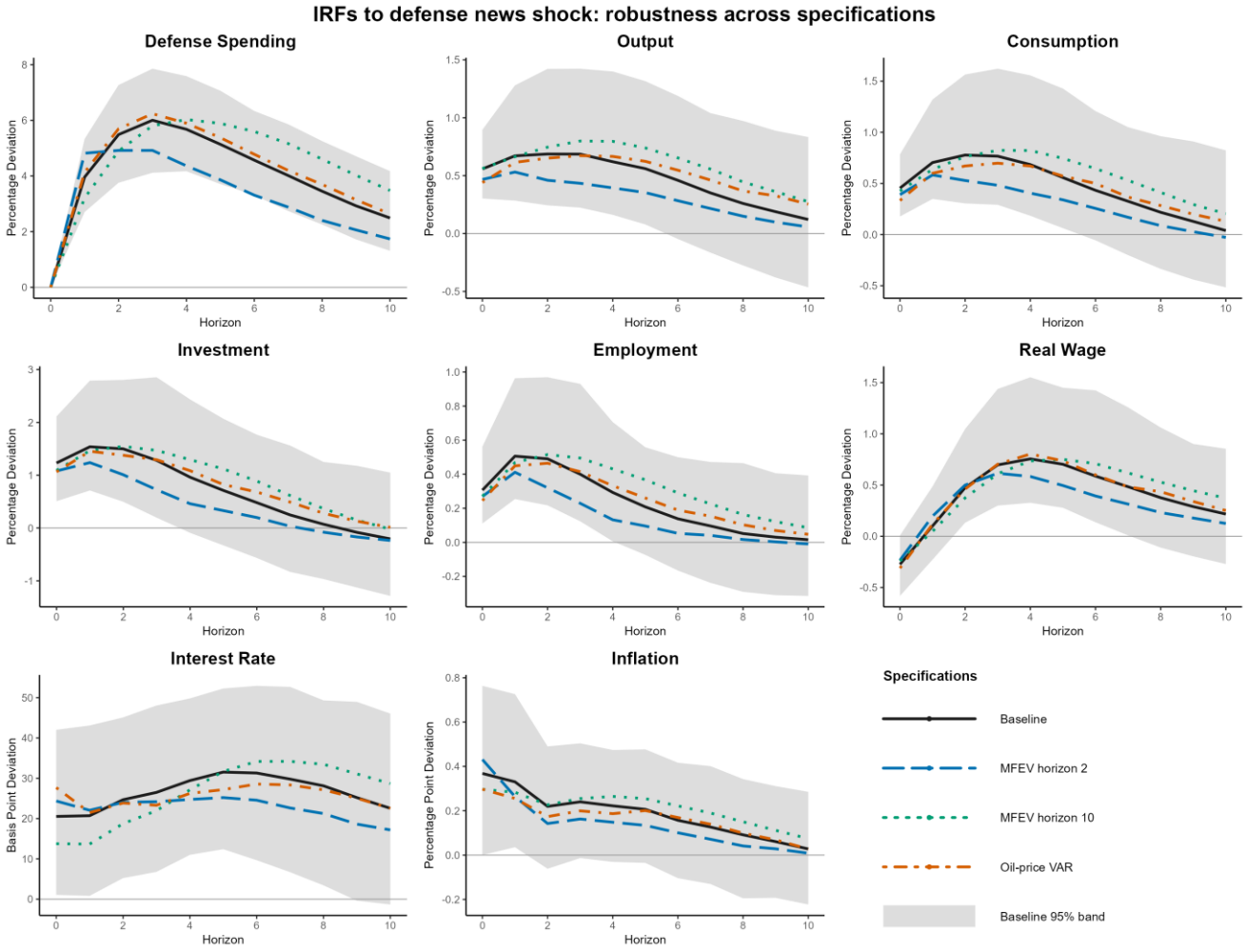
We conduct numerous robustness tests, the results of which are reported and discussed below.

A.1.1 Truncation Horizon

First, we examine the sensitivity of our results to the choice of the truncation horizon, that is, the horizon used in the medium-run forecast error variance (MFEV) maximization identification method. In the baseline specification, the defense news shock is identified by maximizing its contribution to the forecast error variance of defense spending over a five-year horizon. Re-estimating the model with the $H = 10$ horizon yields output multipliers that are very close to the baseline. As seen in Figure 7, IRFs follow very similar patterns across both robustness and baseline horizons. Using a shorter horizon of $H = 2$ produces somewhat smaller multipliers, although qualitatively like the baseline estimates. Overall, our results remain robust to identification horizon specifications.

A.1.2 Oil-Price Control

We also assess whether fluctuations in oil prices drive our results. As the oil price is sensitive to geopolitical events and conflicts, they may coincide with defense news announcements. Oil price fluctuations could then affect the responses of variables such as inflation, output, and spending. To test this, we include annual average WTI oil prices in the VAR. The data is sourced from Federal Reserve Bank of St. Louis and is available for the entirety of our sample period. As seen in Figure 7, including oil price controls in the VAR model does not materially alter either the multipliers or the responses of the other macroeconomic variables, suggesting that oil prices are not the driving force behind our findings.

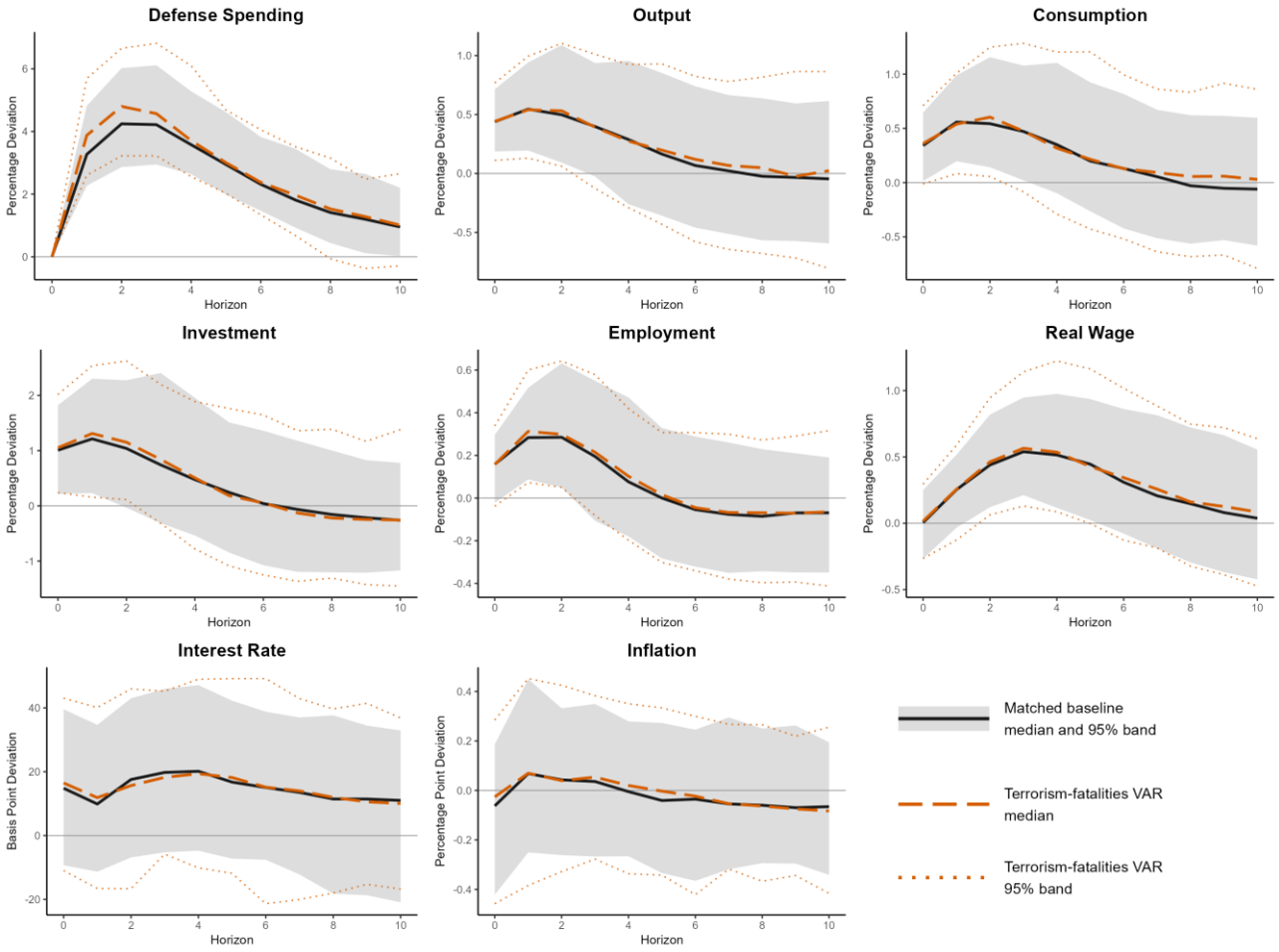


[Note] Lines show median pooled impulse responses to a defense news shock. The grey band is the baseline 95% credible band. The truncation-horizon checks change the MFEV horizon used to identify the defense news shock; the oil-price VAR adds annual WTI oil prices as an endogenous control. Log-variable responses are reported as percentage deviations; interest rates are basis-point deviations and inflation is percentage-point deviation. Raw model horizons 0 through 10 are plotted.

Figure 7: IRFs to defense news shocks: robustness across specifications.

A.1.3 Terrorism Control

Terrorism can plausibly affect both defense and security spending, as well as macroeconomic outcomes. The latter could plausibly occur through, for example, lower market confidence or disrupt tourism. As such, we control for terrorist attacks by adding this as a variable in our VAR. This allows terrorism-related shocks to enter directly, rather than being absorbed into the identified defense news shock. If terrorism were materially driving the baseline results, we would expect the estimated defense news IRFs and multipliers to change noticeably once terrorist fatalities are included. The data is sourced from the Global Terrorism Database and covers confirmed deaths from terrorist attacks by country, over the period 1970 to 2021, recording an accumulated total of 14 148 fatalities in our sample. The data is concentrated to a few countries, with Türkiye accounting for approximately 50 percent of all fatalities, followed by the United Kingdom and Spain. As seen in Figure 8, the inclusion of terrorist attacks does not materially change our IRFs.



[Note] Lines show median pooled impulse responses to a defense news shock. The grey band is the matched-baseline 95% credible band; dotted colored lines are the robustness 95% credible band. Both specifications use the same 30 countries and the same terrorism-data country-year support; the robustness line adds log(1+terrorism fatalities) as a ninth endogenous variable. Log-variable responses are reported as percentage deviations; interest rates are basis-point deviations and inflation is percentage-point deviation. Raw model horizons 0 through 10 are plotted.

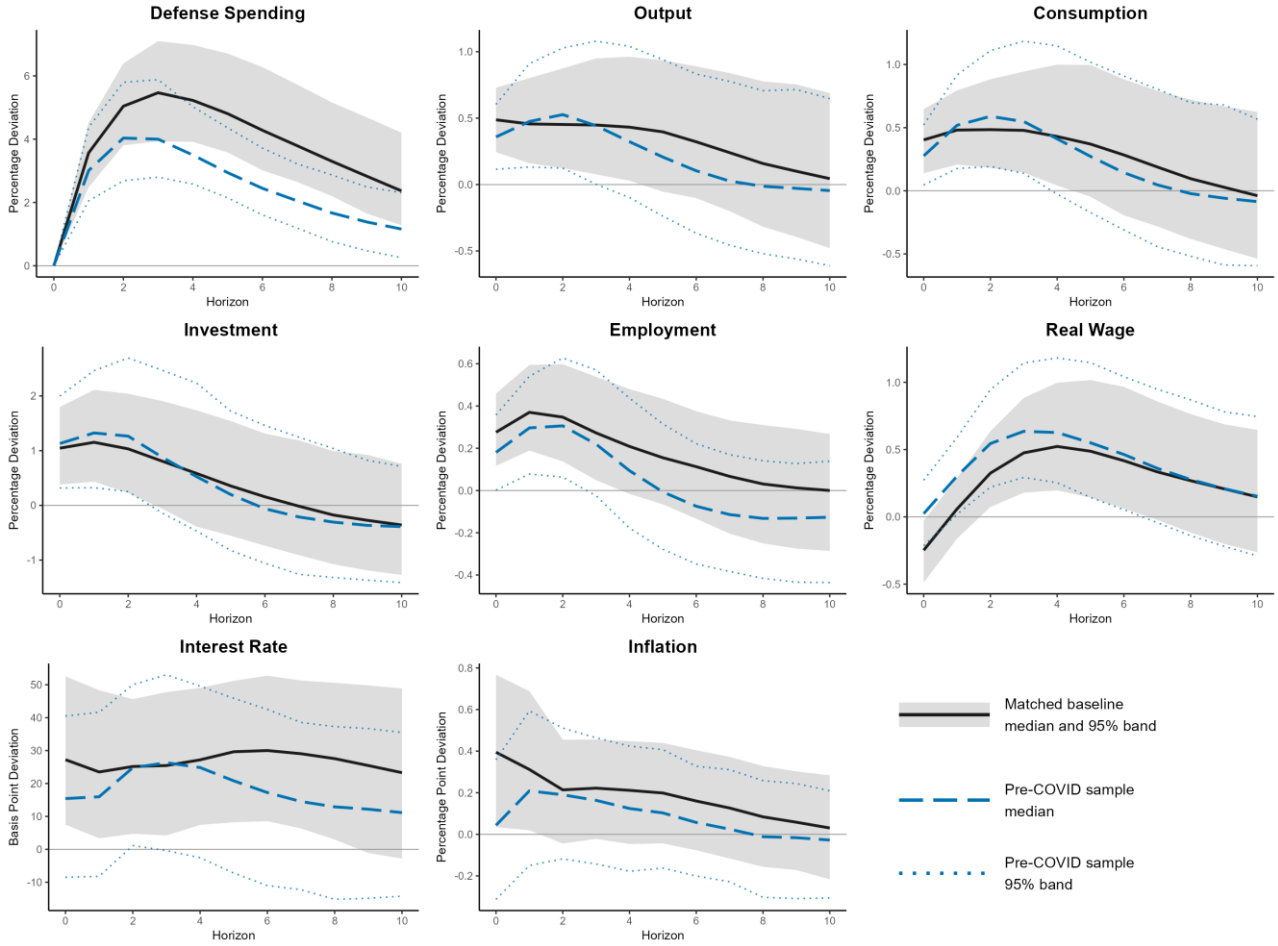
Figure 8: Terrorism matched comparison: impulse responses and pooled news multipliers.

A.1.4 Excluding the Post-COVID 19 Period

We investigate whether the baseline results are driven by macroeconomic fluctuations associated with the COVID-19 pandemic, which resulted in large economic consequences (see Figure 9). By excluding the post-2019 period, we also demonstrate the sensitivity of our findings to the 2022 Russian invasion of Ukraine and associated inflationary pressures from energy prices. Because the post-2019 restriction also makes several short country panels drop out of estimation due to too few observations, we compare the pre-COVID estimates to a baseline that is re-estimated on the same 30-country sample, ensuring that the comparison is not driven by changes in country composition.

This sample restriction yields median multipliers that are consistently lower than the full-sample baseline, suggesting that the post-2019 observations positively contribute to multiplier estimates. For example, at H1 the median multiplier falls from 0.84 in the matched baseline to only 0.66 in

the pre-COVID-19 sample.[11] While credibility intervals remain large, this gives some indication that the recent period drives a relatively large share of the multiplier medians. For IRFs, all retain their broad shape in the pre-COVID-19 period, except for inflation which becomes much lower on impact, suggesting that the five excluded years have a substantial effect on the initial inflationary pressure associated with our defense news shocks in the baseline sample.



[Note] Lines show median pooled impulse responses to a defense news shock. The grey band is the matched-baseline 95% credible band; dotted colored lines are the robustness 95% credible band. The matched baseline drops the five countries skipped in the pre-COVID run; the robustness line additionally drops 2020 and later years. Log-variable responses are reported as percentage deviations; interest rates are basis-point deviations and inflation is percentage-point deviation. Raw model horizons 0 through 10 are plotted.

Figure 9: Pre-COVID matched comparison: impulse responses to a defense news shock.

A.1.5 Alternative Defense Spending Definition

We consider the robustness of our results to the definition of defense spending. Our baseline model uses data from SIPRI, which follows the military expenditure definition adopted by NATO. It therefore captures different spending components than the classification used in the Eurostat COFOG GF02 defense data set. Important differences are the inclusion of civil defense expenditure in COFOG GF02, and the exclusion of military expenditure and healthcare. (SIPRI, n.d.; Eurostat, 2025). We therefore evaluate differences in results by estimating a new model that

[11] The corresponding 95 percent credible bands are (-0.39, 3.21) for the matched full-sample baseline and (-3.82, 5.15) for the pre-COVID sample. At horizons 2, 4, 8, and 10, the matched-baseline medians are 0.84, 0.76, 0.61, and 0.52, compared with 0.62, 0.63, 0.40, and 0.25 in the pre-COVID sample.

instead uses COFOG data. The COFOG dataset is more limited than that provided by SIPRI, covering at most 29 countries over 1990-2025 in our matched support, though most country coverage begins in 1995 and ends in 2024; 28 countries are estimable in the COFOG robustness model.

To ensure comparability, we specify an additional model where the panel covers exactly the countries and years of the COFOG sample but uses SIPRI data for defense spending. As seen in Figure 10, the IRFs of both samples exhibit qualitatively similar patterns. The COFOG-definition model does, however, produce larger output multipliers, particularly at short horizons, relative to our comparison baseline sample. Figure 10 displays these two specifications using the yellow and green lines. The matched SIPRI-baseline sample gives multipliers of 0.63, 0.70, 0.95, 0.93, and 0.69 at horizons 1, 2, 4, 8, and 10, respectively, while the COFOG-definition model gives corresponding multipliers of 1.64, 1.76, 1.58, 1.08, and 0.75. Following Ben Zeev et al. (2026, p. 16), one plausible explanation is that COFOG and SIPRI/NATO-style defense spending differ in expenditure composition. In their interpretation, the larger short-run multiplier can partly reflect that the COFOG definition captures a narrower, and potentially more capital- or equipment-intensive set of expenditures, while SIPRI/NATO-style military expenditure covers relatively more labor costs.

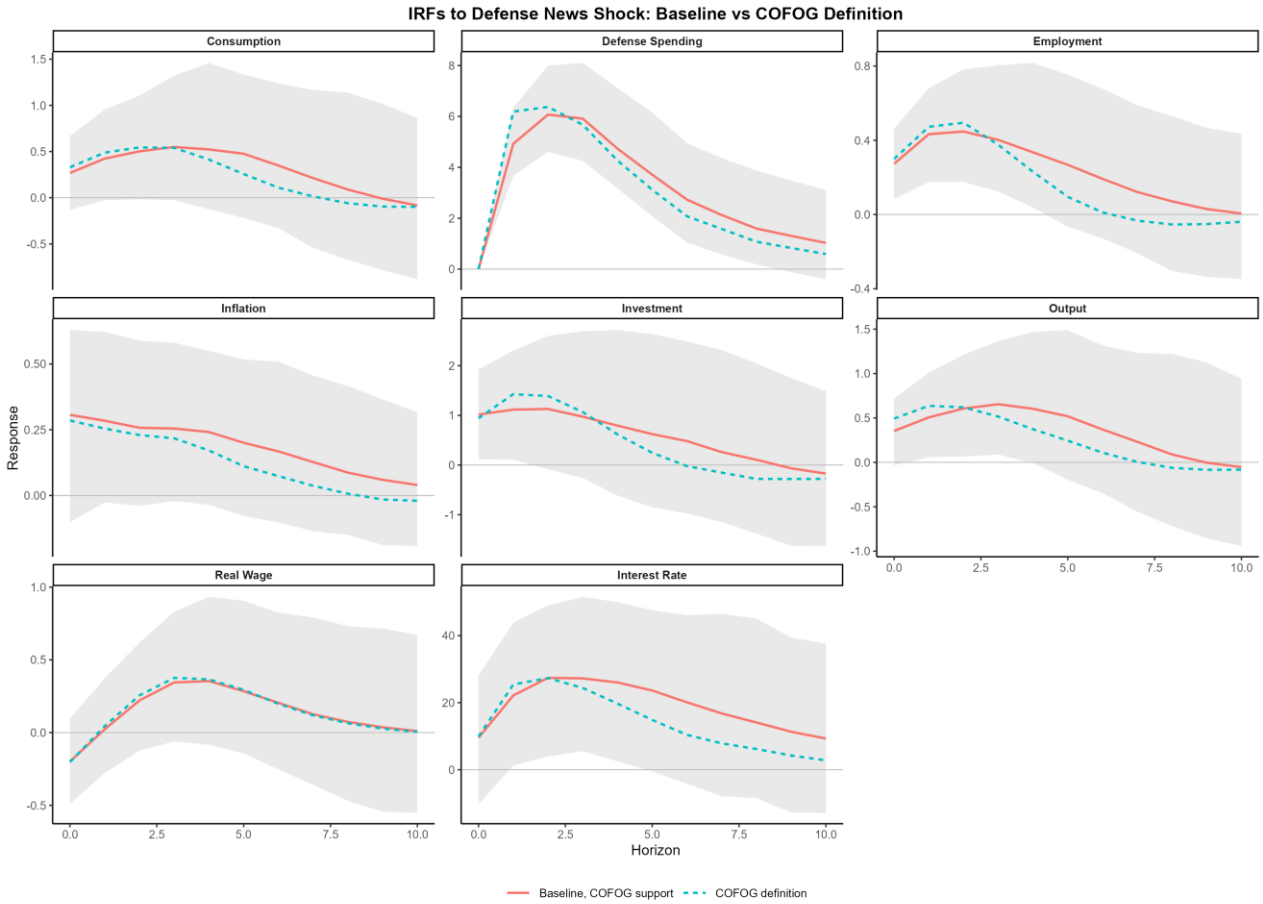


Figure 10: Impulse responses for the baseline and COFOG defense spending definitions.

A.1.6 Geopolitical risk index VAR

It is reasonable to expect announcements of new defense spending to occur in conjunction with heightened geopolitical risk (GPR). This does not conflict with the interpretation of defense news as exogenous, since geopolitical events are one plausible source of variation in expected future defense spending. The concern is instead that our results may capture the direct macroeconomic effects of geopolitical events, such as market uncertainty, rather than the effects of anticipated defense spending itself.

To assess if such a confounding influence is driving the results in our sample, we conduct an additional robustness check using the global Caldara-Iacoviello GPR index (Caldara and Iacoviello, 2022). Since this is a global index, it accounts for common geopolitical-risk episodes rather than country-specific ones. Including the GPR index in the VAR, in addition to the other baseline variables, allows us to plot the impulse response of GPR to defense spending news (see Figure 11). The response of GPR is positive on impact, suggesting that the identified defense news shocks may occur around geopolitical events. However, the response is not statistically significant and dissipates much more rapidly than the effect on the other macroeconomic variables.

Further, as seen in Figure 12, the IRFs are virtually the same whether GPR is included or not. Taken together, this suggests that our baseline results are not primarily driven by the direct effects of measured geopolitical risk, while not ruling out that geopolitical events motivate defense announcements.

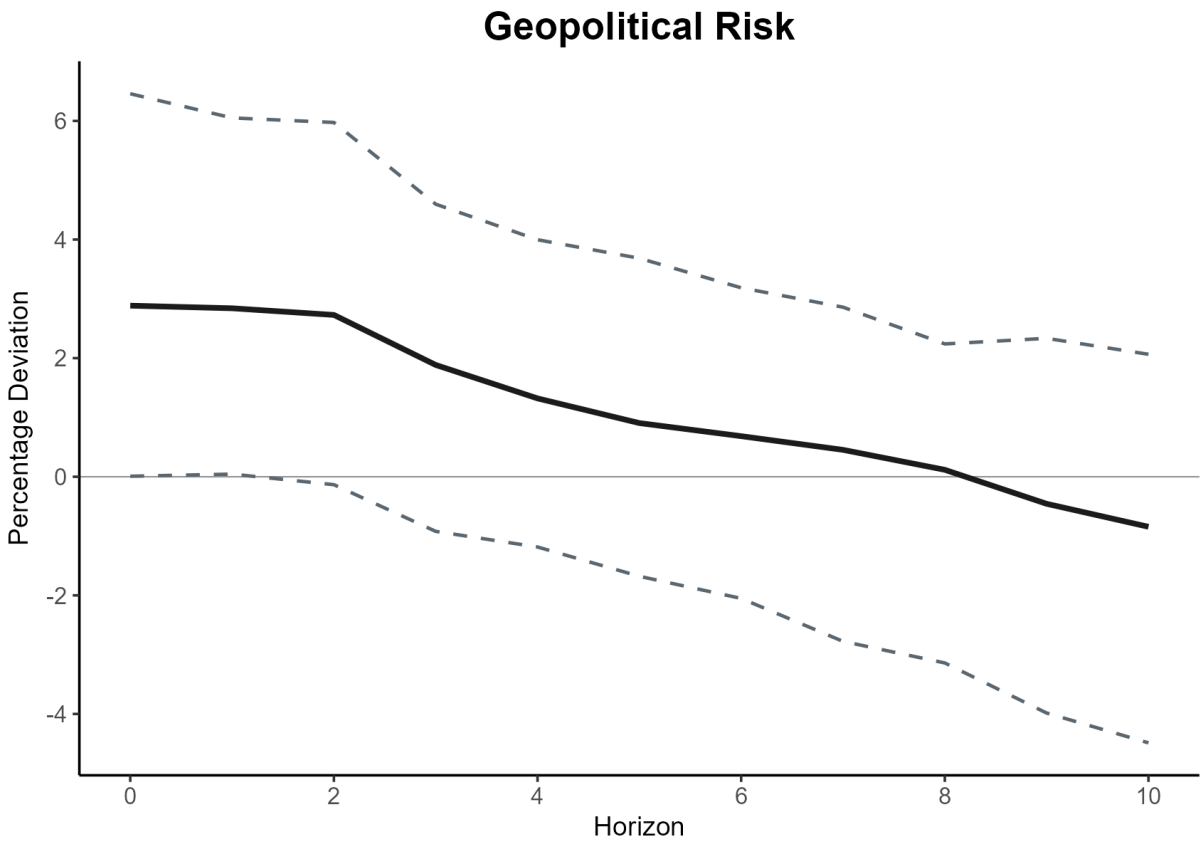
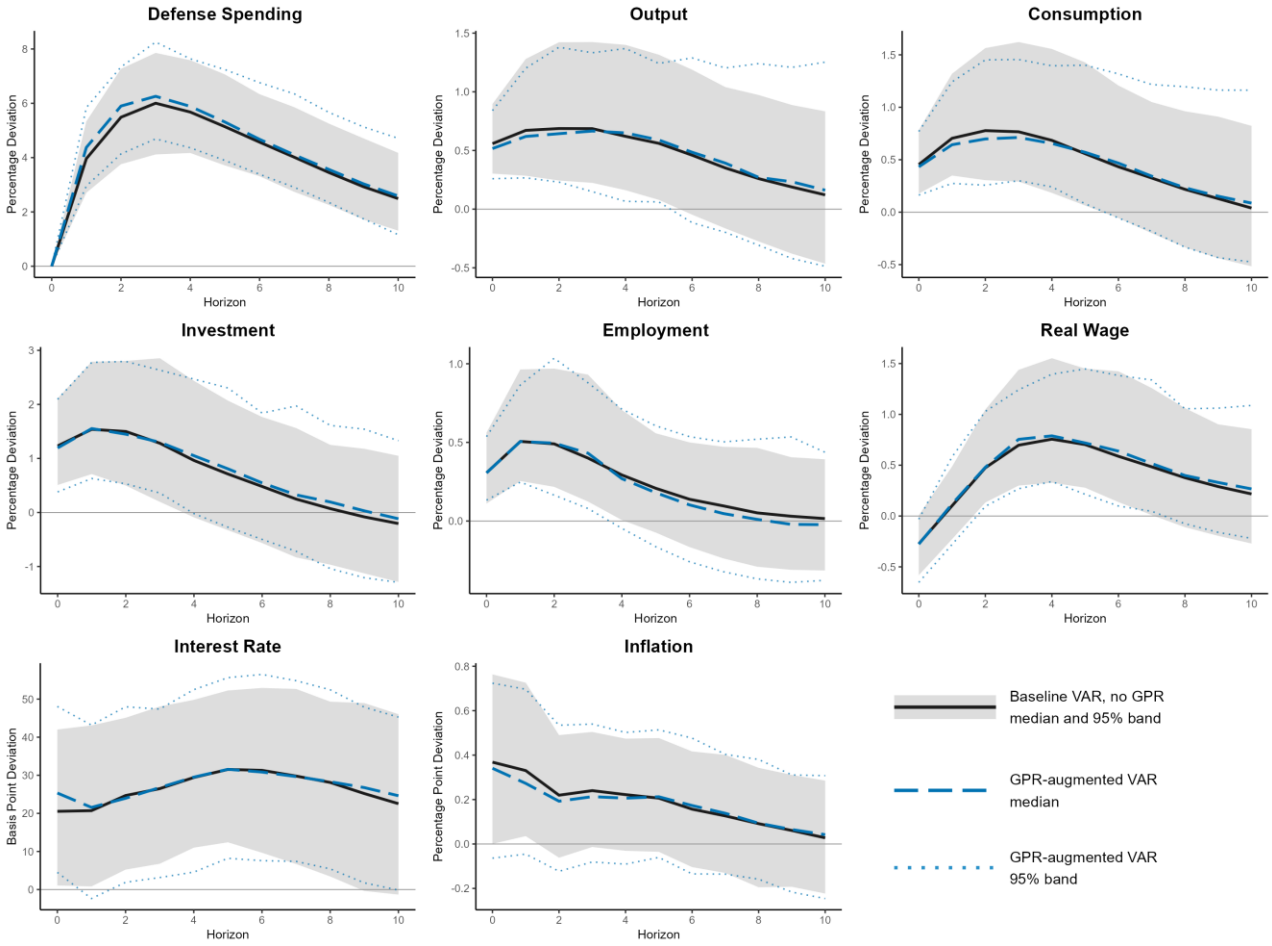


Figure 11: Geopolitical Risk.



[Note] Lines show median impulse responses to a defense news shock from the baseline VAR without GPR and from the GPR-augmented VAR. The grey band is the baseline no-GPR 95% credible band; dotted blue lines are the GPR-augmented 95% credible band. GPR is included as an additional endogenous variable in the GPR-augmented VAR but is omitted from this figure. Log-variable responses are reported as percentage deviations; interest rates are basis-point deviations and inflation is percentage-point deviation. Raw model horizons 0 through 10 are plotted.

Figure 12: GPR-augmented VAR impulse responses.

A.2 Import dependence and Regional IRFs

Median responses differ in some panels, particularly for output and private-demand components, but both groups generally remain close to the baseline response and within the baseline 95 percent credible band. The clearest visual separation of IRFs in Figure 13 and 14 appears in interest rates. In both import dependence measure groups, the low-import-dependence group has a larger immediate interest rate response. Under the aggregate import-intensity split, the high-import group rises gradually and overtakes the low-import group after the first few horizons, while under the military-import-reliance split, the low-reliance group remains higher throughout horizons. The clear differences in interest rate response patterns should not be overinterpreted. We see no obvious economic rationale for why this variable should differ in the context of import dependence. It is plausible that it captures incidental differences between the countries assigned to each group. The aggregate-import-intensity split also illustrates that larger output responses do not necessarily imply larger multipliers. The multiplier is calculated in relation to how much defense spending itself rises in response to the shock.

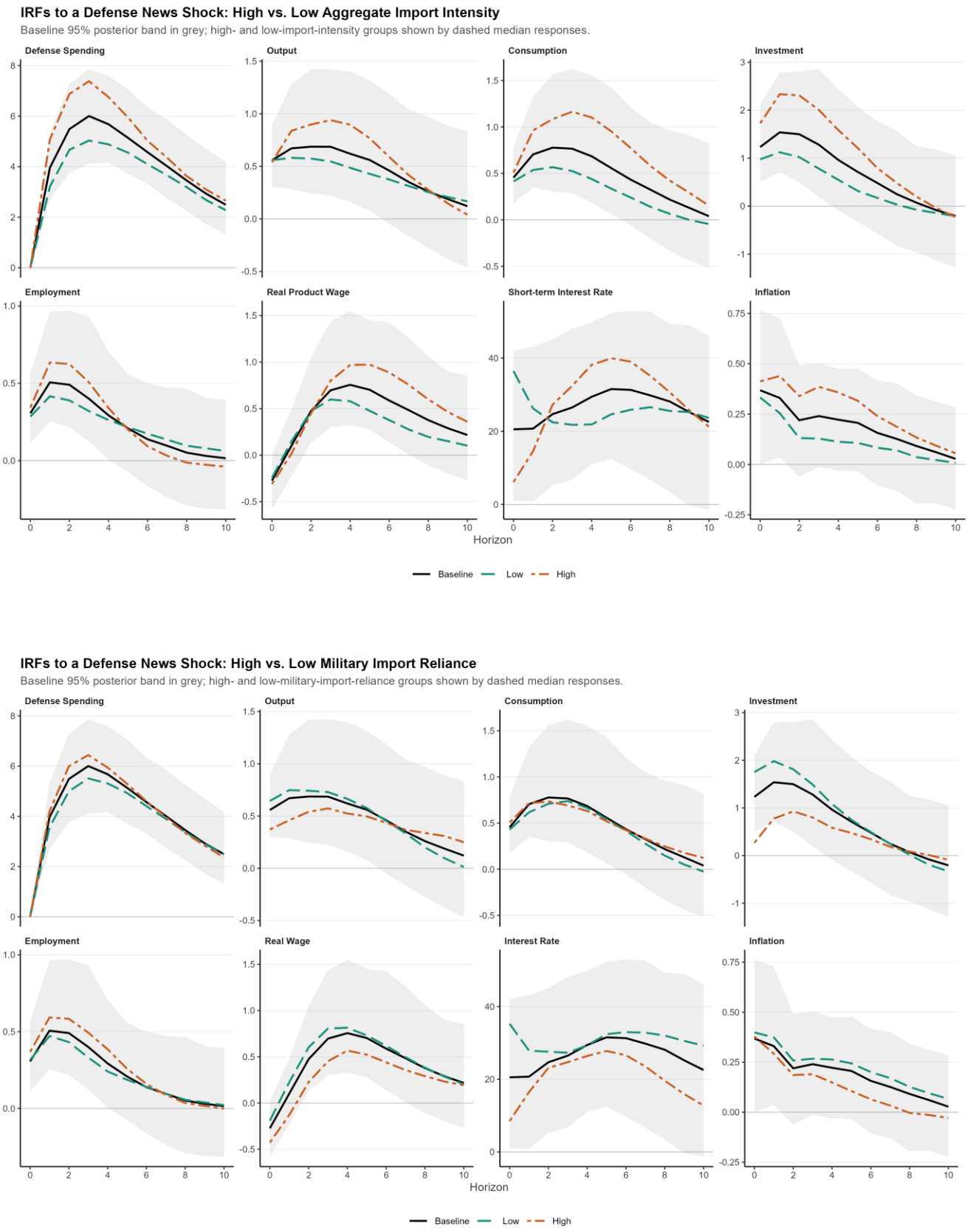
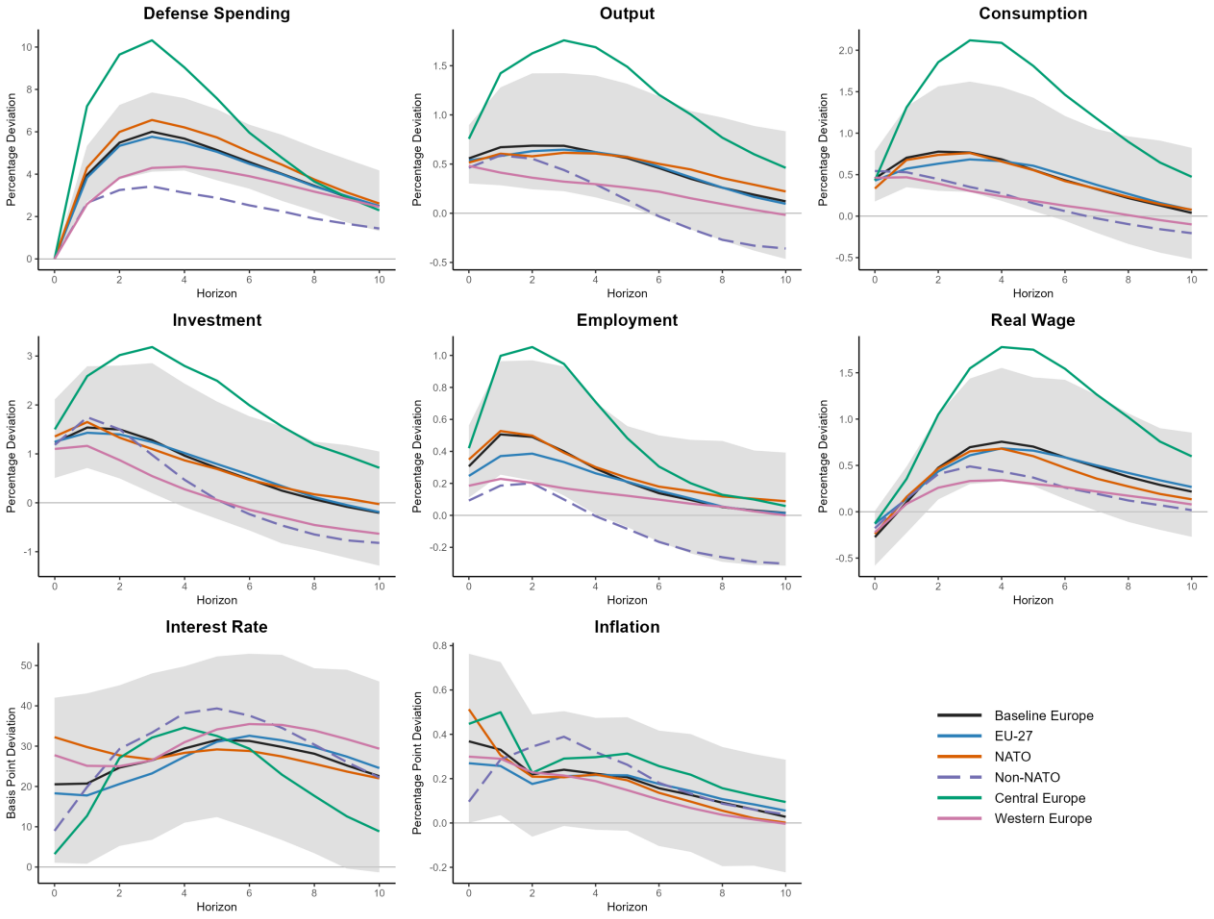


Figure 13: IRFs to a Defense News Shock: High vs. Low Import Dependence.



[Note] Lines show median impulse responses to a defence news shock by subgroup. The grey band is the baseline Europe 95% posterior interval. Log-variable responses are percentage deviations; interest rates are basis-point deviations and inflation is a percentage-point deviation. Raw model horizons 0 through 10 are plotted.

Figure 14: Regional impulse responses to a defense news shock.

A.3 Data

This section details the primary sources, harmonization procedures, and novel extraction techniques used to construct the unbalanced 35-country panel covering the period 1961-2025. Our primary objective in data compilation was to maximize the time series length for each country while staying faithful to variable definition consistency across sources and countries.

A.3.1 Defense Spending

We depart from Ben Zeev et al. (2026) by measuring defense spending in real local currency rather than in real U.S. dollars. Since the VARs are estimated for each country separately, identification relies on within-country variation. Using real local currency ensures consistency with the other macroeconomic variables of output, consumption, and investment, which are available in real local currency units. This improves economic interpretation while not materially affecting cross-country comparability, since cross-country comparison occurs after unit normalizations of country-level estimates. Taking the natural logarithm of these variables further reinforces our econometric design, as any static currency scaling factor becomes an additive constant that is

absorbed by the VAR's intercept. However, while fixed historical conversion factors are factored out, dynamic real exchange rates are not. As discussed in appendix A.4.2, excluding a dynamic exchange rate variable means open-economy price spillovers are implicitly absorbed into domestic aggregate responses, representing a formal limitation of our framework.

A.3.2 Primary Sources

The foundation of our macroeconomic data set is the European Commission's AMECO database. For most Western European economies, AMECO provides consistent uninterrupted series from the 1960s onwards. To extend the series backwards for countries where AMECO data coverage begins later (particularly for post-Soviet Union countries), we supplemented the AMECO series with historical data from the OECD (Main Economic Indicators and Economic Outlook databases) and respective National Statistical Offices (NSOs). We looked for potential structural breaks caused by using different sources for the same variable series with potentially different definitions and validated its congruence before adding new data.

A.3.3 Novel Data

As noted in section 4.1, extending the panel to 35 countries was primarily constrained by the availability of historical interest rate data for non-EU-member states, especially the Balkan countries Albania, Montenegro, North Macedonia, and Serbia. To resolve this while remaining faithful to variable definitions, we employed two Python-based pipelines. First, because the Central Bank of Albania's website is dynamically rendered, we developed a Selenium-based python web-scraping pipeline to extract daily three-month interbank money market rates. Second, we developed an image-processing pipeline that we deployed on the data aggregation platform CEIC, which displays historical graphs of many macroeconomic indicators. The image-processing pipeline reconstructed monthly time series from bar-chart images into aggregated annual values, by isolating the visual signal via color segmentation, extracting bar heights at the pixel level, and converting these into numerical values using known reference observations. The resulting series are approximate reconstructions of the underlying data rather than exact extracts. However, since Montenegro uses the Euribor, we could directly compare our extracted values with exact historical averages from the ECB. The observed measurement errors caused by the image-processing pipeline were deemed negligible.

A.3.4 Extending Western European Series

We successfully extended time series coverage for Austria, Denmark, Finland, Ireland, Spain, Sweden, and Switzerland in the baseline panel of 16 Western European countries. For Finland, Ireland, Spain, and Sweden, the interest rate constitutes the main constraining variables for

extending their time series markedly. In these countries, and many others, money market rates were either nonexistent or highly regulated prior to the financial liberalization of the 1980s and 1990s. During these earlier periods, the central bank's key policy rate served as the definitive benchmark determining the short-term cost of borrowing across the domestic economy. Therefore, using the central bank key policy rate as a proxy for the interest rate accurately reflects the true financing conditions of that era. The practice of splicing historical central bank policy rates with modern money market rates to create continuous long-run series is established in empirical macroeconomics, e.g., the *Macrohistory Database* by Jordà, Schularick, and Taylor (2017). We closely inspected if splicing key policy rates with money market rates would create structural breaks in our series, but this did not occur. Because SVAR identification relies on reduced-form innovations rather than absolute rate levels, this splicing is econometrically justifiable. Under their respective regimes, both rates proxy the marginal cost of short-term capital. Consequently, the spliced series adequately captures the dynamic monetary stance required for structural inference without introducing prohibitive estimation noise.

A.3.5 Serial Correlation

We test serial correlation at lag length $p = 1$ for defense news shocks at the country-level using the Ljung-Box test. Denmark is the only country in our sample with a reported p-value below 0.05 (0.0038). This means that at a 95 percent confidence level, we reject the null hypothesis that Denmark's defense news shocks are not serially correlated. This does not materially change the interpretation of our pooled estimates. However, it does mean that we cannot confidently interpret Denmark's defense news shocks as bringing new information to the system which economic agents contemporaneously react to.

A.4 Further Discussions on Limitations

A.4.1 Currency Conversions

A notable methodological consideration regarding our data set pertains to how Eurozone members potentially create a structural break in the AMECO series when these countries switch from local currency to the euro. To reconcile these concerns, AMECO reports a standardized retroactive conversion methodology. For all years prior to a country's adoption of the euro, AMECO converts historical national currency data using the irrevocably fixed euro conversion rate established at the time of accession. This retroactive conversion introduces a theoretical limitation: it applies a fixed, end-of-sample exchange rate to historical periods characterized by fluctuating exchange rates and differing monetary regimes. As AMECO explicitly notes, cross-country comparisons and aggregations of absolute values prior to 1999 should ideally be based on

the historical European Currency Unit (ECU) rather than the fixed-rate euro construct. Consequently, comparing raw, pre-Euro aggregate defense spending or GDP levels across countries using these series would yield historically inaccurate relative magnitudes.

However, this data limitation does not materially undermine our empirical strategy, owing to our econometric design. By applying a fixed scalar (the irrevocable conversion rate) to the entire historical series of a given country, the AMECO approach mathematically guarantees the historical continuity of the national time series. Because our SVAR systems are estimated on a country-level using log-levels transformations and per-capita adjustments, our identification relies exclusively on within-country variance over time. The fixed-rate conversion prevents an artificial structural break in the data at the year of euro adoption, ensuring that the dynamic properties and long-term variation of the macroeconomic aggregates remain intact. Furthermore, our Europe-wide estimates are derived from Bayesian hierarchical pooling of unit-normalized structural objects (IRFS, FEVs, multipliers). Because we pool these unit-independent posterior distributions rather than directly aggregating raw historical currency values, our framework is robust to the cross-country comparability limitations inherent in the AMECO source data.

A.4.2 Exchange Rate Dynamics

A further limitation of our empirical strategy pertains to the open-economy transmission of fiscal policy, specifically the endogenous response to relative exchange rates to country-level defense news shocks. In our baseline VAR, we specify defense spending and all macroeconomic aggregates in real local currency units. While this design choice preserves the internal consistency of domestic unit normalizations, it leaves the exchange rate transmission channel entirely unmodeled. When a country-level defense news shock hits an economy, it generates upward pressure on domestic short-term interest rates. This is a dynamic which we empirically confirm in our baseline impulse responses (Figure 2), which show interest rates rising by roughly 31 basis points in the medium run. According to standard macroeconomic theory, an expansionary fiscal shock accompanied by monetary policy tightening typically induces an appreciation of the domestic real exchange rate (Corsetti Meier, and Müller, 2012; Born, Juessen, and Müller, 2013).

This endogenous appreciation presents a specific complication for our estimates measured in local currency units. Because defense procurement often relies heavily on imported capital and foreign technology, a stronger domestic currency lowers the relative local currency price of imported military goods. Consequently, an exchange rate appreciation effectively increases the international purchasing power of the domestic defense budget. Alternatively, if the defense news shock hits the foreign country where military imports are acquired from, this would appreciate the relative exchange rate in the exporter country, thereby effectively decreasing the purchasing power of the

importer country. By omitting the real exchange rate from our VAR system, our model absorbs these relative price and trade-balance effects implicitly into the domestic macroeconomic aggregates rather than isolating them. Therefore, while our local currency specification accurately captures the aggregate domestic response, it limits our ability to parse out how much of the observed crowding-in effect is driven by favorable terms-of-trade shifts versus domestic aggregate demand expansion. However, if both the defense importer and exporter countries are under the same monetary policy regime, such as the European Monetary Union and Eurozone, the effect of defense news shocks on relative exchange rate appreciation is eliminated.

A.4.3 Statistical Identification

A fundamental limitation of our empirical strategy stems from the purely statistical nature of our identified defense news shocks. By defining the news shocks as the structural innovation that maximizes the forecast error variance (MFEV) of defense spending over a five-year horizon, we recover a continuous variable derived mathematically from the co-movement of the reduced-form VAR residuals. While this MFEV approach is highly efficient for capturing anticipated changes in expenditure, the resulting shock series does not map directly onto specific, observable historical events, such as public defense announcements, parliamentary budget votes, or press releases.

This statistical abstraction contrasts sharply with the narrative approach pioneered in fiscal foresight literature. Most notably, Ramey (2011) constructed a defense news shock series for the United States by systematically reading historical editions of *Business Week* and other newspaper sources. Through this narrative method, she extracted explicitly quantified, event-based estimates of changes in the expected present value of government spending. Her approach allows researchers to trace macroeconomic responses back to specific, real-world information shocks. In our context, manually constructing a comparable narrative series across 35 European countries over seven decades would be empirically intractable. Thus, while the MFEV procedure provides a mathematically rigorous workaround that successfully captures the aggregate fiscal foresight of economic agents, it operates somewhat like a "black box" model. Because the shock is an assumed structural variation derived from our collected data, we cannot definitively attribute the initial macroeconomic impulse to a specific, identifiable historical news event.

A.4.4 Linearity

A structural limitation of the baseline VAR specification is the mathematical assumption of linearity and symmetry in the transmission of fiscal shocks. The standard fiscal multiplier constructed in this framework implies that each additional unit of government defense spending generates an equal, proportional effect on aggregate output regardless of magnitude, although

declining over time. However, recent advancements in fiscal policy literature robustly demonstrate that multipliers are state-dependent and non-linear. Specifically, the size of a fiscal multiplier is highly dependent on both the magnitude and the sign of the fiscal shock (Brinca et al., 2025). From a theoretical standpoint, extrapolating our estimated multiplier to massive, historically unprecedented European rearmament programs is somewhat naive. As defense spending scales dramatically, it is likely subject to declining marginal returns. At high levels of expenditure, capacity constraints and supply-side bottlenecks begin to bind, prompting tighter monetary policy responses and heavier taxation distortions that dampen the marginal output effect of additional spending.

Appendix B: Artificial Intelligence Disclosure

In accordance with guidelines on generative artificial intelligence tools (GenAI) in thesis writing set by Stockholm School of Economics, this appendix details the use of GenAI tools during the research and writing process of this thesis.

B.1 GenAI Tools Utilized and their Application

During the development of this thesis, we utilized two distinct categories of Large Language Models (LLMs) and AI-agents as supplementary research and implementation tools:

1. Text-Generating LLMs (ChatGPT, Gemini, and Claude): These models were employed during the preliminary stages of the research to generate broad overviews of the fiscal multiplier and SVAR literature in a defense spending context. They were also used to aid our learning curve on the technical aspects of the SVAR methodology. In the later stages of the writing process, they were used as editing assistants to provide stylistic feedback and suggest structural improvements to the text.
2. Code-Generating AI-Agents (Claude Code and OpenAI Codex): The empirical foundation of this thesis required estimating complex country-level Bayesian structural vector autoregressions (SVAR) with hierarchical pooling and identifying shocks via maximum forecast error variance (MFEV). We utilized these AI coding agents as tools for implementing the defined methodology into written code inside a working R script.

B.2 Contribution to Thesis Quality

The integration of these tools systematically improved the quality of the thesis in two primary ways. First, utilizing AI coding agents enabled us to more quickly implement the Bayesian SVAR methodology with MFEV and hierarchical pooling from theory into code. This allowed us to spend much more time learning the theoretical framework underpinning Bayesian SVARs. The coding assistance also freed up time that allowed us to allocate more focus towards looking for new data sources, which enabled us to construct a novel data set of 35 European countries. Second, using LLMs for text review and writing feedback proved highly valuable, e.g., iteratively asking if the current granularity at which a specific concept was explained was sufficient depending on reader's level of expertise.

B.3 Potential Risks and Mitigation Strategies

The use of GenAI in academic research introduces significant risks, which we actively managed through critical judgement:

1. Risk of Hallucination and Theoretical Misrepresentation: LLMs are prone to generating plausible but mathematically or theoretically incorrect statements, as well as fabricating citations. *Mitigation:* We strictly treated all literature overviews provided by the models as starting points. Every cited paper (e.g., Ben Zeev et al., 2026; Ramey, 2011) was manually read and verified to the extent it was used in the thesis. All theoretical claims made by the LLMs were manually verified.
2. Risk of Unseen Errors ("Black-Box" Code): Relying on AI-agents to implement econometric models into code risks the introduction of subtle mathematical errors in e.g. complex matrix operations or posterior sampling that deviate from the methodology. This risks corrupting the results. *Mitigation:* Our main mitigation strategy pertaining to silent coding errors was to first implement the R script on the described panel of 16 Western European countries by Ben Zeev et al. (2026) and verify our results against their reported results. We computed qualitatively and quantitatively similar IRFs, FEVs, and news multipliers, strengthening the case that our script faithfully represented the methodology. However, our computed surprise multipliers were materially larger than those reported by Ben Zeev et al. (2026). As discussed in Section 5, we believe this difference is primarily driven by ambiguous data set specifications and implicit methodological assumptions not explicitly written out in the authors' methodology section. The primary example of this is the handling of posterior draws in the posterior simulation via rejection sampling of explosive draws, which is something Ben Zeev et al. (2026) never comment on, but is standard in the SVAR literature and follows economic theory as explained in section 4.6.1 Bayesian Estimation.

B.4 Insights from using GenAI in the Thesis Process

The main insights gained from implementing GenAI as a coding assistant and writing feedback partner were that, while GenAI tools have become very capable as of 2026, they still lack critical judgement, are prone to hallucination, and often fail to differentiate between relevant and irrelevant content. In the macroeconometrics domain, we must understand exactly why certain assumptions and decisions are made, such as the Cholesky decomposition or NIW priors, to correctly leverage GenAI. We learned that the effectiveness of GenAI tools is proportional to the clarity of the given instructions, reflecting the level of understanding possessed by the user.

Further, rigorousness of subsequent verification of the GenAI output was found to be crucial. Ultimately, GenAI successfully elevated our ability to push the boundaries of the level of research we could achieve within the scope of a bachelor thesis. The responsibility for economic theory, data integrity, and final conclusions remains entirely our own.