On the Effects of Government Purchases and Their Transmission Mechanism*

Click here for the most updated version of this document.

Edoardo Briganti†

University of California, San Diego

This version: November 25, 2023

Abstract

I use novel data on defense contracts to study the effects of government purchases in the US and develop new stylized facts about their transmission mechanism. My methodology leverages the construction of a new quarterly series of US military prime contracts, available from 1947:1. Defense contracts: (i) are exogenous to output fluctuations; (ii) retain statistical power and robustness across various samples; (iii) accurately measure the timing of the shocks; and (iv) obviate the need for narrative analysis. My findings indicate that a positive shock to defense contracts, ordered first in a VAR, bolsters output, inventories, non-durable-plus-service consumption, hours worked, employment, labor earnings, disposable income, the price-cost markup, the product-wage, and labor productivity. I argue that the observed gains in labor productivity stem from "learning-by-doing," a feature particularly relevant to the production of military items. Further, leveraging a two-sector RBC model, I demonstrate that the learning-by-doing induced productivity enhancements in the manufacturing sector suffice to increase aggregate consumption, rationalizing the VAR evidence.

(JEL E60, E62)

---

*I am particularly thankful to Valerie Ramey, Nir Jaimovich and Jim Hamilton for their constant feedback and help. I also acknowledge helpful comments from Francesco Amodeo, Giampaolo Bonomi, Ricardo Duque Gabriel, Juan Herreño, Munseob Lee, Jacob Orchard, Giacomo Rondina, Fabian Trottner, Johannes Wieland and seminar participants at UCSD.

†PhD Candidate in Economics. Email: `ebrigant@ucsd.edu`.

I. Introduction

What is the impact of government spending (G) on consumption and GDP? Despite this being one of the classic questions in macroeconomics, there is still no consensus on the answer, with the debate centered on the assumptions and methods used to identify government spending shocks.

In this paper, I propose to identify government spending shocks using a newly constructed instrument for government spending: "defense contracts," a quarterly time series available from 1947:1, accounting for the dollar value of all US military prime contracts. I find that if government spending increases by 1$, non-durable-plus-service consumption increases by 0.12$.

Previous work in the fiscal policy literature can be divided into two main camps: the Blanchard and Perotti (2002)'s approach, or "SVAR Approach," which identifies government spending shocks by ordering G first in a VAR (henceforth, BP shocks); and (ii) the instrument approach, which (i) relies on instruments for G which measure expected defense spending and (ii) place them first in a VAR.

The first line of research, the SVAR approach, has faced criticism due to its failure to account for the anticipatory effect of government spending (Ramey (2011)). As a response, Ramey proposed the use of "defense news shocks" as an instrument for G. Simultaneously, the second line of research, the instrument approach, has been criticized for reasons such as weak instrument problems.

In this paper, my measure aims to address both of these shortcomings. Specifically, I demonstrate that defense contracts accurately measure the timing of the shocks, without missing out on any early and relevant GDP response. Furthermore, defense contracts preserve statistical power and alleviate concerns about weak instrument problems.

The literature has found positive responses of consumption in response to BP shocks but concerns arise from their timing and their potential endogeneity. In fact, G might endogenously respond to GDP within the same quarter, violating the VAR's recursive assumption (i.e., Cholesky identification). In contrast, defense contracts reflect variation in defense spending that is primarily driven by exogenous military events. Moreover, changes in government spending are anticipated, and BP shocks, constructed using government spending, are anticipated too (Ramey (2011)). In contrast, defense contracts reflect the future value of defense procurement spending, which is part of the National Income and Product Accounts (NIPA)'s measure of G, and they accurately measure the timing of the fiscal shocks. In fact, NIPA follows the accounting practice to record most defense contracts into G only after payment-on-delivery, which, in the case of several complex items such as aircraft and missiles, occurs several quarters after the beginning of production (see Brunet (2022) and Briganti and Sellemi (2023)). Therefore, G lags behind the placement of new defense orders, as measured by defense contracts. Simultaneously, NIPA keeps track of ongoing production in response to new orders using inventories. In fact, I find positive responses of inventories, displaying a faster response than G.

Considering the endogeneity and anticipation of government spending, the literature has constructed instruments which measure future changes in military spending. For example, Ramey (2011) and Ramey and Zubairy (2018) narratively construct defense news shocks, which measure future changes in military spending deemed exogenous to output variations; Fisher and Peters (2010) constructs Top 5, an index of cumulative excess returns of defense contractors; Ben Zeev and Pappa (2017) identifies defense news shocks using medium-run restrictions. A shared limitation of all these instruments is that they suffer from low statistical power in samples after the Korean War, that is, after 1953 (Ramey (2016)). In fact, results can be quite sensitive to the exclusion of the Korean War (Perotti (2014)). Even if I argue that the Korean War must be included in the baseline sample, as the US economy never turned into a war-economy and its size is dwarfed by the scale of WWII (Hickman (1955), Dupor and Guerrero (2017)), defense contracts provide results robust to the exclusion of the Korean War from the sample. Furthermore, defense contracts preserve statistical power in samples after the Korean War.

Given that defense contracts effectively address the main critiques of the literature regarding the timing of shocks and the statistical power of the instruments, my results remain resilient against these fundamental concerns and aim to provide a more conclusive answer to the longstanding questions on the effects of government purchases.

In the second part of the paper, I find that a shock to defense contracts causes an increase in labor productivity, which can, in theory, rationalize the observed rise in consumption (Devereux, Head, and Lapham (1996)). Using balance-sheet data from publicly traded defense contractors from Compustat and contract data from the Top100 companies report from the Department of Defense (DoD), I also find that lagged contracts are associated with higher labor productivity of defense contractors around the years of the Vietnam war. Christiansen and Goudie (2007) find similar results using a panel VAR from 1969 to 1996. I offer a comprehensive examination of how learning-by-doing in manufacturing and military production enhances productivity with rising production rates, offering a plausible explanation for the observed rise in labor productivity in response to contracts. To provide context, here are two interesting anecdotes: (i) the concept of learning-by-doing itself was introduced to economics through the analysis of military aircraft production data (Arrow (1962)), and (ii) even the official BEA's Government Transaction Methodology Paper acknowledges the effects of learning in generating rapid price declines for military items due to increased productivity.

Therefore, I use a two-sector RBC model with manufacturing and non-manufacturing to rationalize the empirical evidence. In the model, government purchases increase only in manufacturing, to mimic the well-known sectoral bias of procurement spending (Ramey and Shapiro (1998), Perotti (2007), Nekarda and Ramey (2011), Cox et al. (2022)). Additionally, only manufacturing production is characterized by learning-by-doing, allowing for an increase in labor productivity when production rates rise. The use of learning-by-doing is motivated by the aforementioned thorough discussion of its vast empirical evidence, found almost exclusively in manufacturing and defense production. In the model, learning can induce an increase in consumption of a magnitude similar to that empirically observed, thereby rationalizing the VAR evidence.

Related Literature

My work directly speaks to the vast literature which studies the aggregate effects of government spending. The literature can be segmented into two groups: those finding positive consumption responses with the "*SVAR approach*" and those generally finding negative responses via the "*instrument approach*."

Firstly, the SVAR approach identifies government spending shocks by ordering the NIPA measure of G first in a VAR (Fatas and Mihov (2001), Blanchard and Perotti (2002), Galí, López-Salido, and Vallés (2007), Monacelli and Perotti (2008) and Perotti (2014)). This method is based on the notion that policymakers and legislatures require more than a quarter to learn about a GDP shock, determine the appropriate fiscal response, pass these measures through the legislature, and implement them.

Secondly, the instrument approach identifies government spending shocks by constructing instruments for G using different measures of expected military spending. The instruments use (i) military spending instead of total spending to ensure that the variation is driven by exogenous military events and (ii) expected changes instead of current changes, as government spending is often anticipated several quarters in advance (Ramey (2011)). Notable examples using this approach are Ramey (1989), Ramey (1991), Ramey and Shapiro (1998), Burnside, Eichenbaum, and Fisher (2004), Eichenbaum and Fisher (2005), Fisher and Peters (2010), Ramey (2011), Barro and Redlick (2011), Ben Zeev and Pappa (2017) and Ramey and Zubairy (2018).

Advocates of the instrument approach criticize the SVAR approach, contending that because government spending is anticipated, the shocks they identify are predictable (Ramey (2011)). Proponents of the SVAR approach counter by claiming that the instruments for G often lack statistical power and are overly sensitive to the sample choice (Perotti (2014)).

I contribute to the literature by constructing a new instrument for government spending, "*defense contracts*", which addresses the central critiques from both sides of the literature. Namely, defense contracts measure expected defense (procurement) spending in line with the instrument approach, while avoiding the pitfall of low statistical power and sample choice sensitivity, thus addressing the criticism raised by proponents of the SVAR approach.

The idea of using defense contracts to identify government spending shocks builds on the recent work of Brunet (2022) and Briganti and Sellemi (2023), who propose an alternative empirical explanation for why GDP moves before G in response to a defense news shock (Ramey (2011)). They highlight the NIPA's accounting practice of recording military contracts for complex military items, such as aircraft or missiles, at the time of the payment, which occurs after delivery. NIPA keeps track of the ongoing contractors' production by recording a positive change in inventories. In fact, they find positive responses of inventories in response to fiscal shocks. *The change in inventories creates a semblance of fiscal foresight as GDP mechanically moves before G due to accounting reasons.* If the limitation of the SVAR approach is its oversight of GDP's early response, but this effect primarily reflects contractors' production in response to new contracts not yet recorded by NIPA in G, then defense contracts can account for this, as their timing is aligned with the beginning of production. This distinction clarifies why defense contracts accurately reflects the timing of the shocks and offers a significant advantage over the SVAR approach, which instead relies on the NIPA measure of G, which records military contracts with a delay.

Concerning the effects of new contracts on contractors labor productivity, my work directly relates to Christiansen and Goudie (2007), who find a positive response of labor productivity of defense contractors in response to newly awarded contracts, using a sample which starts from 1969. Since I have been able to gather all available Top 100 companies report from the Directorate for Information Operations and Reports (DIOR), I extend the analysis from 1960 to 1969, the years around the outbreak of the Vietnam war.

I also relate to the vast literature on learning-by-doing developed around manufacturing and, in particular, military programs. Notable examples are: Wright (1936), Asher (1956), Alchian (1963), Smith (1976), Gulledge and Womer (1986), Bourgoine and Collins (1982), Argote and Epple (1990), Argote, Beckman, and Epple (1990), Benkard (2000). Other relevant works which discuss labor productivity gains due to learning effects during WWII are: McGrattan and Ohanian (2010) and Ilzetzki (2023) (i.e. "*Learning-by-Necessity*").

Lastly, on the empirical side, my contribution relates to other works studying the effects of government spending in the US economy. Notable examples are classified as follows: Mountford and Uhlig (2009) (sign restrictions); Leeper, Traum, and Walker (2017) (Bayesian prior and posterior analysis of DSGE model); state/sign dependency: Auerbach and Gorodnichenko (2012), Ramey and Zubairy (2018) and Barnichon, Debortoli, and Matthes (2022); state/city level analysis: Nakamura and Steinsson (2014), Dupor and Guerrero (2017) and Auerbach, Gorodnichenko, and Murphy (2020); consumer level analysis: Giavazzi and McMahon (2012) and Coibion, Gorodnichenko, and Weber (2020); industry level analysis: Acemoglu, Akcigit, and Kerr (2016), Bouakez, Rachedi, and Santoro (2023) and Barattieri, Cacciatore, and Traum (2023).

Although my model is inspired by an empirical question — whether enhancements in labor productivity for manufacturers or contractors due to learning effects can boost aggregate labor productivity and increase consumption — it implicitly relates to the theoretical literature that develops models resulting in consumption increases following a positive government spending shock.

To the best of my knowledge, the models that yield a consumption increase in response to a positive government spending shock are limited to a few examples: Devereux, Head, and Lapham (1996) employs an RBC model with increasing returns to specialization as per Krugman (1979); Galí, López-Salido, and Vallés (2007) uses a New Keynesian (NK) model with rule-of-thumb consumers; both Monacelli and Perotti (2008) and Bilbiie (2011) apply a NK model with non-separable preferences and consumption-hours complementarities; and Jørgensen and Ravn (2022) adopts a medium-scale NK model with variable technology utilization. However, none of these models are well-suited to answer my question, as they do not feature a two-sector framework with manufacturing characterized by learning-by-doing. D'Alessandro, Fella, and Melosi (2019) integrate the learning-by-doing mechanism from Chang, Gomes, and Schorfheide (2002) into a one-sector medium-scale NK model, resulting in increased consumption. Nevertheless, there are two crucial distinctions between our models. First, my model is driven by the observation that learning-by-doing predominantly occurs in manufacturing data and military programs. Consequently, I adopt a two-sector model aligned with this observation, calibrating the learning parameters using empirical findings from military data. Second, my approach to modeling learning differs slightly from Chang, Gomes, and Schorfheide (2002). While they rely on past deviations of hours worked, I use current deviations of manufacturing output, consistent with empirical literature examining learning curves in manufacturing and military programs.

The paper is organized as follows: Section 2 delineates the construction of the new instrument for G, namely defense contracts, and discusses its benefits. Section 3 presents the empirical results. Section 4 offers a theoretical rationalization of the empirical findings. Section 5 concludes.

II. A New Instrument for G

II.a From VAR Shocks to Contracts

Measuring the aggregate effects of government spending requires the identification of government spending shocks. The conditions that ensure a valid identification of government spending shocks are (i) exogeneity to output changes and (ii) unpredictability (Ramey (2016)).

The SVAR approach identifies government spending shocks using the principle that changes in government spending at quarterly frequency are predetermined (Blanchard and Perotti (2002)). This is achieved by ordering the NIPA measure of government spending, G, first in a VAR (i.e. Cholesky identification). I will refer to these shocks, as the BP shocks.

One main concern with BP shocks is that most variation in government spending can be anticipated by economic agents. To better illustrate the point that GDP and its components can move even before any actual visible change in G, I replicate in the top-panel of Figure 1 the impulse response functions (IRFs) of GDP and G to a defense news shock, found in Ramey (2011). Firstly, defense news shock is a narratively constructed instrument for government spending, G, which measures the present discounted value of expected changes in military spending associated by exogenous military events, as predicted by the periodical *Business Week*:[1]

$$(\text{Defense News Shocks})_t = \sum_{h=0}^{H} \mathbb{E}\left(\frac{G_{t+h}^{\text{Defense}}}{(1+i_t)^h} \middle| \Omega_t^{\text{News}} \cap \Omega_t^{\text{Exogenous}} \right), \tag{1}$$

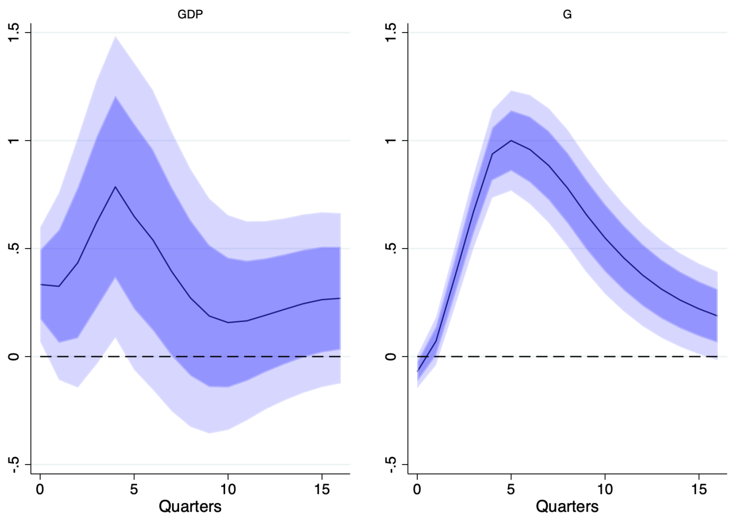
where i_t is the interest rate and Ω denotes an information set.

After a defense news shock, the IRF of G increases starting from quarter 2, while the GDP surge is immediate. The literature attributes the earlier GDP response to the Ricardian behavior of economic agents who supply more hours in anticipation of expected higher taxes needed to fund the additional expenditure (i.e. *negative income effect*).

An alternative explanation is presented by Brunet (2022) and Briganti and Sellemi (2023). Brunet suggests that the NIPA measure for G is time delayed,

[1]See Figure notes for VAR details.

(a) Anticipation Effect of G



(b) Response of Inventories

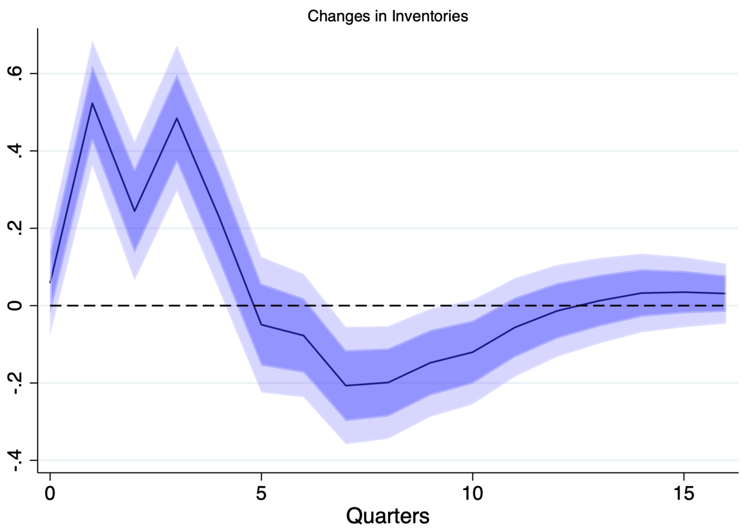


Figure 1: Response to a Defense News Shock

Notes: IRFs of GDP, G and Inventories with respect to a defense news shock (updated series of Ramey and Zubairy (2018)). Sample: 1947:1 to 2015:4. Values normalized by the peak response of G. IRFs are obtained from a VAR with defense news shocks, G, GDP, tax receipts, business sector hours worked, and the 3-month T-Bill rate, inventories and a quadratic trend. Real variables are divided by real potential GDP, estimated with a sixth degree polynomial fit to log-real GDP (see Ramey (2016)).

as NIPA logs government contracts only upon delivery of the purchased items. In particular, when the government commissions new aircraft, multiple contracts are secured for various components like the airframe, engine, and communication systems. These components are manufactured for multiple aircraft, assembled, and only then delivered to the government. The entire process spans several quarters, with contractors compensated after-delivery. As NIPA logs these payments, a timing disparity arises between the beginning of production and the recording of contracts into G by NIPA.[2] Both Brunet (2022) and Briganti and Sellemi (2023) document empirically the time lag. Furthermore, Brunet posits that NIPA accounts for ongoing production using inventories, a claim for which Briganti and Sellemi (2023) provide corroborative evidence. The bottom-panel of Figure 1 exhibits the IRF of inventories following a defense news shock, mirroring the findings of Briganti and Sellemi (2023).

Notice that inventories exhibit a positive response to a defense news shock, peaking at horizon 1 — prior to the initial response of G. Briganti and Sellemi (2023) conclude that the detected anticipation effect with defense news shocks can be attributed to inventory responses, which mirror the deferred production of defense items in G, stemming from two primary factors: (i) the convention of compensating contractors after-delivery and (ii) the duration required to fabricate complex defense items, like military aircraft.

II.b Construction and Benefits of Defense Contracts

In this section I detail the construction of a new quarterly variable of defense contracts and illustrate its advantages compared to other established instruments for government spending.

II.b.i Construction of Defense Contracts

First, data on defense contracts is recorded when a firm is awarded a new prime contract award from the DoD. For example, imagine Boeing wins a new contract worth 800 million dollars for new military aircraft in quarter t, then Boeing will commence the airframe production within the quarter following the award date, since this moment denotes the end of any demand-uncertainty for Boeing.[3] Payments are processed as batches of parts are sequentially delivered from Boeing to the DoD. For instance, Auerbach, Gorodnichenko, and Murphy (2020) convert lumpy contract data from FPDS-NG into spending data by evenly distributing the value of a contract over its duration. In this example, if the contract to Boeing

[2]See the F-15 example at page II-33 of the Government Transaction Methodology paper and the “timing difference” paragraph at page II-11 of the same.

[3]I will address the potential anticipatory behavior of contractors in the subsequent section.

has duration of 2 years, NIPA spending will contain 100 million dollars in each subsequent quarter according to their logic.

Hence, defense *contracts* can be perceived as a weighted average of the current and future values of NIPA defense procurement *spending*, which is a component of G:

$$\left(\text{Defense Contracts}\right)_t = \sum_{h=0}^{H} \psi_h \cdot G_{t+h}^{\text{Defense Procurement}} \tag{2}$$

As a result, the values of future NIPA defense procurement *spending* are known to firms in advance at time t, since the value of lumpy contracts is incorporated by NIPA into G with a delay. Most importantly, as noted in Brunet (2022) and Briganti and Sellemi (2023), when production-deliveries take longer than one quarter, NIPA keeps track of ongoing production using inventories, not defense procurement spending (i.e. G).

Furthermore, comparing Equations (1) and (2) reveals the similarities and differences between defense news shocks and defense contracts. First, their sole commonality is that they both measure future military spending. However, defense news shocks measure the expectations of overall future defense spending related to exogenous military events as forecasted by the news. In contrast, defense contracts mirror the present value of awarded contracts, which are recorded in G after a delay due to NIPA's accounting methods. In this regard, the two variables differ significantly in the way the measure future military spending.

In what follows I describe the data sources of defense contracts.

(i) BCD: The first data source comes from contract data used in Ramey (1989), which is originally from the periodical Business Condition Digest, or "*BCD*".[4] It contains monthly data of prime military contracts from January 1951 until November 1988.[5]

A data limitation arises from the fact that the BCD series starts in 1951, during the Korean War, the largest military shock in the post-WWII sample. Therefore, I use NIPA data on defense procurement spending, constructed as in Cox et al. (2022).[6] Given that NIPA data lags behind defense contracts, I use both

[4] Military contracts became part of the set of instruments known as the Hall-Ramey Instruments (from **hall˙invariance˙1989** and Ramey (1989)).

[5] The series was then discontinued and migrated on the Survey of Current Business (SCB). However, data on SCB is only available from January 1990 until September 1995 with a systematic omission of the fourth quarter of the year. Data on SCB is very noisy and because of the systematic omission, less reliable. Therefore, I will not use this data source.

[6] Sum of NIPA defense intermediate goods and services purchased plus defense gross investments on structure, equipment and software.

contemporaneous and future defense procurement spending to forecast current defense procurement contracts. In particular, I estimate the following equation via OLS, spanning from 1951:1 to 1980:4:

$$\underbrace{\text{BCD}_t}_{\text{Def. Prcm. Contracts}} = \kappa + \sum_{h=0}^{4} \psi_h \cdot \underbrace{\text{NIPA}_{t+h}}_{\text{Def. Prcm. Spending}} + \varepsilon_t.$$

The linear regression yields an R^2 of 80%. Using the OLS estimates and the NIPA data on defense procurement spending from 1947:1 to 1950:4, I predict the defense procurement contracts for that time frame. This data is referred to as "*BCD Extrapolated*"

(ii) FPSR: The second data source originates from the annual Official Federal Procurement Summary Report ("*FPSR*"), produced by DIOR. It contains data on both annual and quarterly federal procurement contracts, and is available starting from the inception of the Federal Procurement Data System (FPDS) in the first quarter of 1981 and ends in the third quarter of 2003. The annual reports present the value of military prime contract awards by fiscal year. They also feature bar charts illustrating quarterly values of total federal procurement contracts: federal defense plus federal non-defense. To remove the non-defense component from the quarterly values, I adjust the quarterly data so that the average of the quarterly values within each fiscal year is equal to the official annual annual values of military data. This adjustment is innocuous; in Online Appendix A, I demonstrate that the fluctuations in federal procurement largely stem from its defense component. Specifically, I reveal that (i) approximately 80% of the federal procurement during those years is attributable to military procurement, and (ii) the quarterly federal values, when aggregated by fiscal year, correlate strongly with the annual military values.

(iii) FPDS-NG: After the fourth quarter of 2000, all daily federal procurement transactions are observed from the Next Generation of FPDS (or "*FPDS-NG*"). I aggregate all new defense contracts and defense contracts modifications by quarter. Since the FPDS-NG data is very noisy, and most noise comes from contract modifications, I add the original new defense contracts to "smoothed" defense contracts modifications.[7] Therefore, the high-frequency variation in the FPDS-NG data comes from newly-awarded contracts and not contract modifications.

[7] Examples of contract modifications are funding only actions, request for extra work, options exercise, cancellations of some work.

Assembling the Series: The top panel of Figure 2 shows real prime contract measures along with the Ramey and Shapiro (1998)'s war dates augmented with the 9/11, the Budget Control Act of 2011 and the election of President Trump.

Notice how well the measures overlap in the 80s and at the beginning of the 2000s, indicating a remarkable consistency across the different data sources.

I append all the data from the four sources to construct a new quarterly variable of defense contracts from 1947:1 until 2019:4. In particular, the series is made of data from (i) BCD-extrapolated from 1947:1 until 1950:4, (ii) BCD from 1951:1 until 1980:4, (iii) FPSR from 1981:4 until 2003:3 and (iv) FPDS-NG from 2004:1 onward. Henceforth, I will refer to this variable as simply "*(defense) contracts*". The bottom panel of Figure 2 shows the newly constructed (real) defense contracts and (real) defense procurement spending from the NIPA data.[8]

Time Variation: It is recognized that the primary variations in Government spending (G) are attributed to the military events of the 20th century (Hall (2009) and Ramey (2016)). This is portrayed in Figure 3, displaying the newly constructed variable defense contracts as a share of GDP.

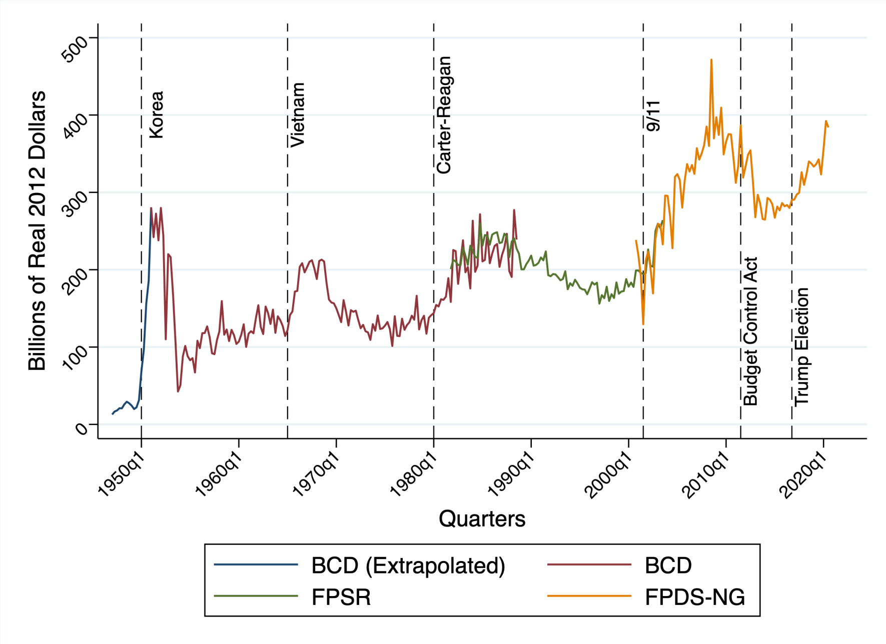
Major shifts in defense expenditure relative to GDP size are mainly linked to war events in the 20th century, with the Korean War standing out as the most substantial shock in the post-WWII era. However, it is worth noting that the Korean War did not transform the US economy into a war-centric one, and its relative size pales when compared to WWII. Consequently, the Korean War provides a natural experiment to study the impact of government spending and should be part of the baseline sample, since it did not transform the US economy into a war economy (see Hickman (1955) and Dupor and Guerrero (2017)).[9]

The variation of military spending relative to GDP in the early decades of the 21st century is clearly smaller than the variation of the second half of the 20th century. Therefore, I primarily use the 1947:1-2000:4 sample for yielding the most accurate estimates. For robustness, I will also check results with a sample which excludes the Korean War (1954:1 to 2000:4), as sensitivity of results to the exclusion of the Korean war is a well-known fact in the literature (Perotti (2014), Ramey (2016)). Similarly, I will check the robustness of my results over the full-sample (1947:1 - 2019:4). Overall, my results are robust across the three samples.

[8] Price deflator is an average of price indices for NIPA intermediate goods and services purchased and government gross investments.

[9] I am aware that price controls were introduced at the end of January 1951, however, their effect was very limited (Hickman (1955)).

(a) Military Prime Contracts Measures



(b) Military Contracts Vs Military Spending

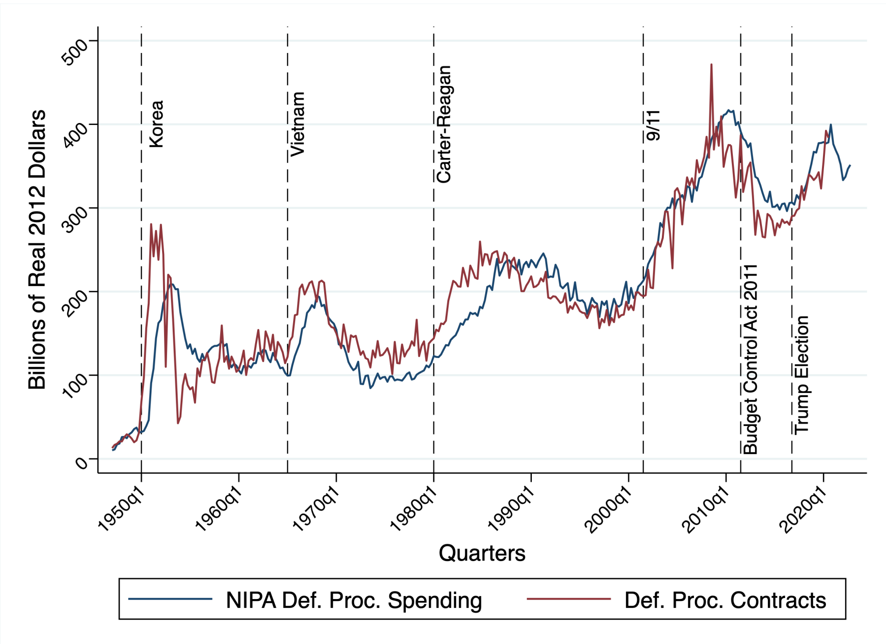


Figure 2: Defense Procurement

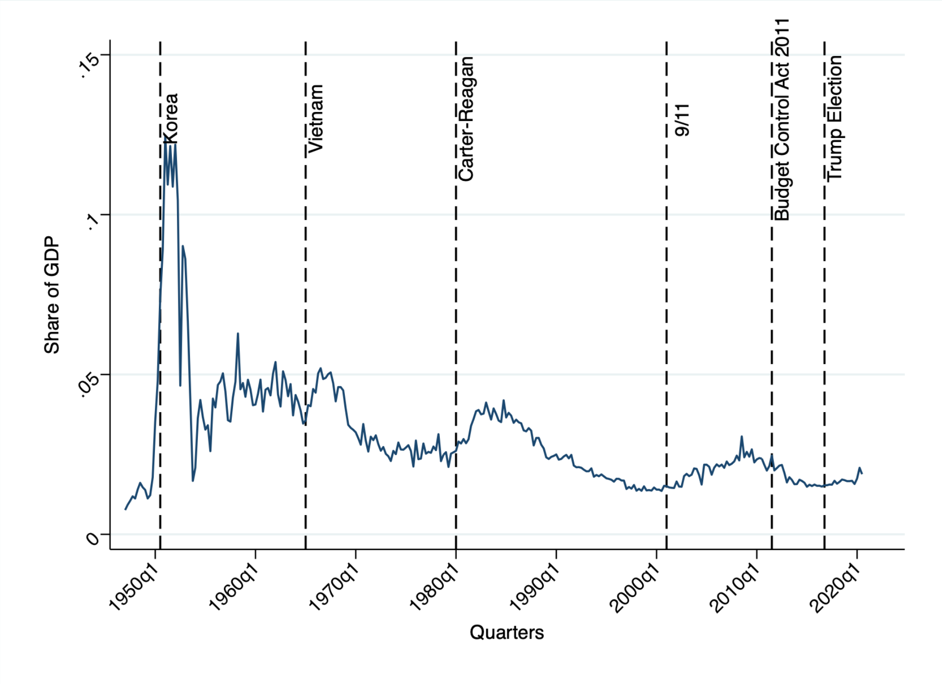


Figure 3: Defense Contracts as Share of GDP

II.b.ii Advantages of Contracts

In this section, I argue that defense contracts accurately measure the timing of shocks, addressing a limitation of the SVAR approach. Secondly, I maintain that defense contracts can serve as an instrument for G, given that the variable is (i) exogenous and (ii) relevant. Lastly, I highlight that defense contracts do not necessitate a narrative analysis, offering a readily accessible method for estimating the effects of government spending in countries that maintain records of military contracts.

Measurement Delay in NIPA: The previous section discussed how the early response of GDP relative to G in response to a defense news shock can be reconducted to the early response of inventories. In turn, the response of inventories captures defense production which does not show up in G yet. In fact, G accounts for the dollar value of the contracts only after payments to contractors, which happen after delivery of the item goods. Since it takes time to produce a defense item, such as a guided missile or an aircraft, G will be delayed relative to contracts (see Brunet (2022) and Briganti and Sellemi (2023)).

In fact, the bottom panel of figure 2 shows that defense contracts lead defense procurement spending from the NIPA, a component of G. I quantify the delay using a lead-lag correlation map. In particular, Figure 4 plots the correlation map between quarterly year-to-year changes of defense procurement contracts and spending.[10]

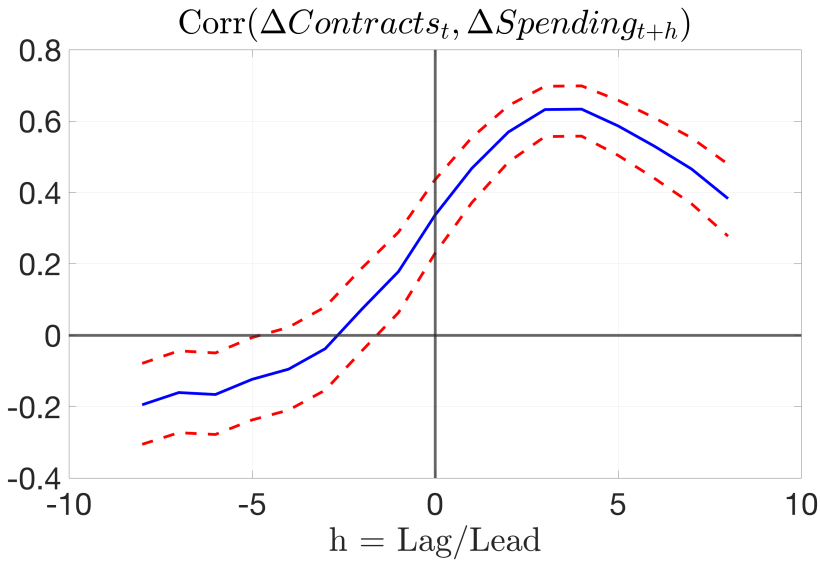


Figure 4: Lead-Lag Correlation Map Between Contracts/Spending

Notes: sample goes from 1948:1 to 2019:4. Here $\Delta_4 x_t$ means $x_t - x_{t-4}$. The price deflator used is the one of Intermediate goods and services purchased by the government, available from NIPA.

Since the correlation map is positive in the North-East quadrant, that is, when spending is delayed, changes in defense procurement contracts anticipate changes in defense procurement spending, as measured by NIPA. In particular, the correlation map spikes at 3 quarters, suggesting a similar average time-delay between contracts and payments. The result replicates if I use first-differences instead of quarterly year-to-year changes and if I look at different time-periods. Robustness checks are reported in Online Appendix A.1.

Another way to formally observe the time delay in G relative to defense contracts is by means of Granger-causality tests. Firstly, I construct BP shocks to government spending by ordering the NIPA measure of G first in a VAR; then, I construct shocks to defense contracts by augmenting the previous VAR with defense contracts ordered first; finally, I conduct Granger Causality tests to see whether one predicts the other and viceversa. Results are reported in Table 1.

[10] Briganti and Sellemi (2023) use lead-lag correlation to study the time mismatch between BCD prime contracts and NIPA spending and find a delay of about 3 quarters between the two.

Table 1: Do Defense Contract Shocks Granger Cause BP Shocks?

Predicted	*Predictor*	*F*	*Pvalue*	*Sample*
Defense Contract Shocks	BP Shocks	1.34	22.25%	1947:1 - 2019:4
Defense Contract Shocks	BP Shocks	1.13	34.47%	1951:1 -2019:4
BP Shocks	Defense Contract Shocks	7.7	0.00%	1947:1 - 2019:4
BP Shocks	Defense Contract Shocks	3.86	0.00%	1951:1 - 2019:4

Notes: Granger Causality test is a Wald test on the 8 lags of the predictor while controlling for 4 lags of the predicted variable. BP shocks are constructed as OLS residual of a regression of G on four lags of G, GDP, Hours worked in the private sector, 3 Months T-Bill rate. Shocks to defense contracts are obtained as OLS residuals of a regression of defense contracts on four lags of defense contracts, G, GDP, Hours worked in the private sector, 3 Months T-Bill rate. All nominal variables are in logs of real per-capita values while hours are in logs.

The top panel of the table shows that BP shocks fail to predict shocks to defense contracts. On the contrary, shocks to defense contracts do predict the BP shocks., as visible from the bottom panel of the table. Results are robust to the exclusion of the outbreak of the Korean war, 1950:3, which uses extrapolated contract data from NIPA defense procurement spending. I interpret this result as a consequence of the delay in recording military contracts into NIPA measure of government spending, as evident from Figure 2 and the lead-lag correlation map.

Defense contracts anticipate NIPA spending, however, I still have to rule out the possibility that contractors systematically anticipate future contracts and begin production even before new contracts awards. Therefore, I carry out two tests.

The first test boils down in augmenting the VAR with Ramey (2011)'s defense news shocks ordered first and investigate the response of contracts and inventories. In fact, if contractors systematically anticipate new contracts, I would expect inventories to significantly respond to a defense news shock even before any significant response of defense contracts. If this is the case, even though contracts predate NIPA spending, they would still miss part of the early response of inventories, which reflects contractors' production.

Therefore, I augment the VAR with the Ramey (2011)'s defense news shocks and inventories. I then look at the IRFs of contracts and inventories to a defense news shock, ordered first. The IRFs are showed in Figure 5.

Notice that the response of contracts (blue line) is largely positive already on impact, indicating that when a defense news shock occurs, newly awarded contracts are disbursed *within the quarter* of the shock's occurrence. In contrast, inventories increase only from quarter 1, that is, *after* the initial positive response of contracts. In other words, concurrently with a positive defense news shock, the Department of Defense promptly awards new prime contracts and contractors

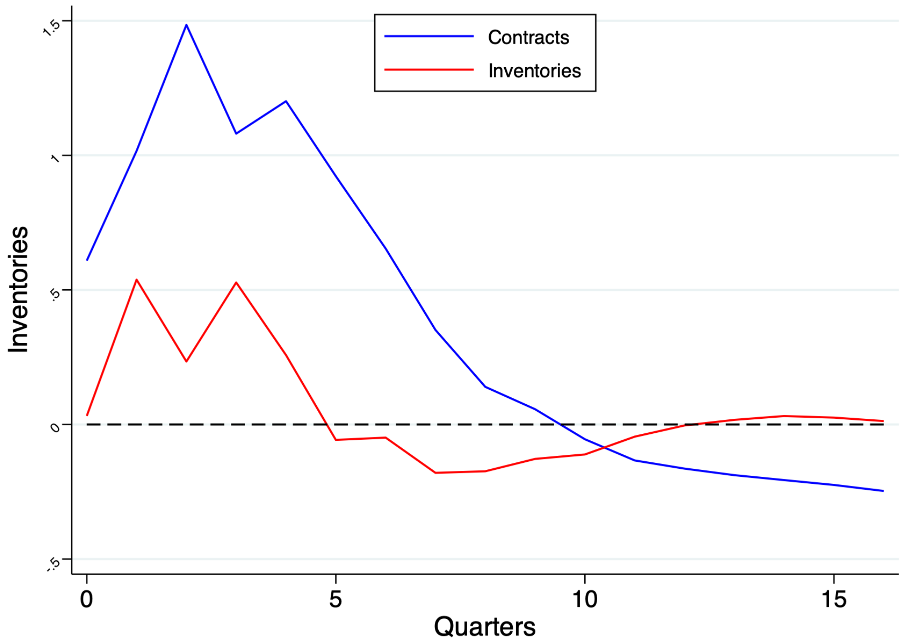


Figure 5: Do Contracts Miss Inventories After a Def. News Shock?

Notes: Point estimates of IRFs of defense contracts and inventories to a defense news shock. Sample goes from 1947:1 to 2019:4. IRFs are obtained from the same VAR of Figure 1

appear to increase production, as monitored by inventories, only after the receipt of those newly awarded contracts. Therefore, defense contracts do not miss any part of the response of inventories which occurs before the change in G.

The second test I carry out answers the question: does the stock market anticipate new defense contracts? Therefore, I construct an equally weighted portfolio of stock prices from four major defense contractors: Boeing, Northrop-Grumman, Lockheed-Martin, and Raytheon which goes from 1947:1 till 2001:4.[11] Then, I calculate the cumulative excess quarterly year-to-year returns for this defense sector portfolio, using the S&P500 as a benchmark index; this approach is inspired by the work of Fisher and Peters (2010). I will refer to this variable as the "*Top 4*" index.

I then use a VAR with Top 4 index, defense contracts, G, GDP, hours worked in the private sector, TB3 and total NIPA tax receipts. I look at the IRF of Top 4 index in response to a shock to contracts as well as the IRF of contracts in response

[11]The companies' choice is motivated by their large dependence of their total revenues on government purchases as well as their constant presence in the list of Top 100 defense contractors.

to a shock to the Top 4 index. I choose the sample from 1954:1 to 2001:4, which excludes the Korean war as done in Fisher and Peters (2010), who noted that the overall stock-market response at the start of the Korean war was significantly dampened by the profit tax increase.

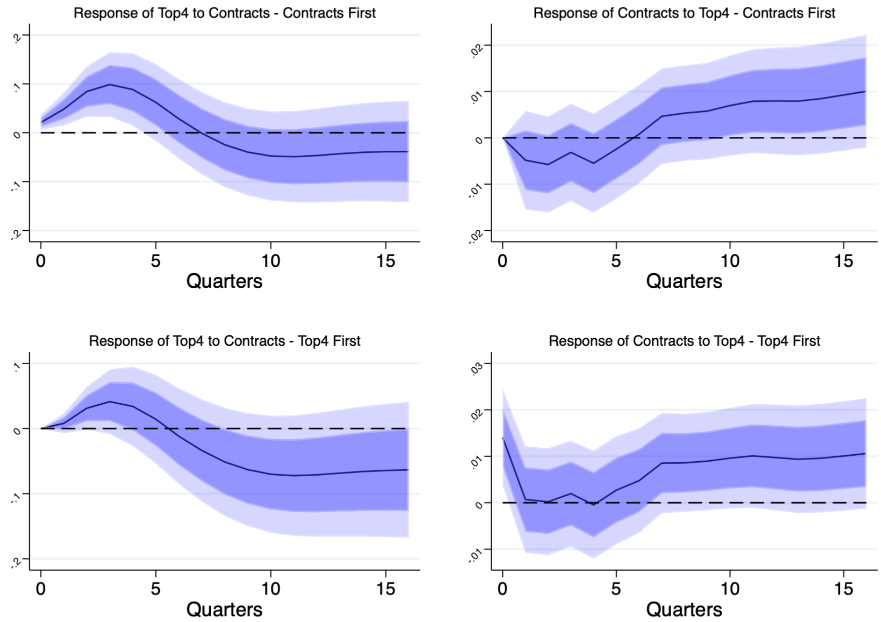


Figure 6: Are Contracts Predicted by the Stock Market?

Notes: Sample is: 1954:1 to 2000:4. Confidence bands are 68% and 90%. Nominal variables are in logs of real per capita values, while hours are in logs. The deflator is the GDP price deflator. *Top 4* accounts for the cumulative excess returns of an equally weighted portfolio of the stocks of Lockheed-Martin, Raytheon (now RTX), Northrop-Grumman and the Boeing Company; I use the S&P500 as a benchmark.

The top-left panel shows the IRF of the Top 4 index in response to a shock to defense contracts when defense contracts are ordered first in the VAR. The response is precisely estimated and indicates that the portfolio gradually increases its value, peaking 3 quarters after the shock to defense contracts. Conversely, the top-right panel illustrates the IRF of contracts to a shock to the Top 4 excess returns, with defense contracts still ordered first in the VAR. The results suggest that after a positive shock to the Top 4 index, defense contracts do not increase, and the response is imprecisely estimated.

Next, I assume the Top 4 index to be predetermined to contracts by ordering it first in the VAR. A positive shock to defense contracts has a similar effect on the Top 4, although the IRF is now less precisely estimated (as seen in the bottom-left panel). The response of defense contracts to a shock to the Top 4 index, when ordered first (bottom-right panel), is also not precisely estimated. However, by design, defense contracts respond positively on impact, aligning with the positive response of the Top 4 index to a shock to defense contracts (as shown in the top-right panel).

Given the absence of a significant delayed response of contracts to a shock to the Top 4 index, and considering the contrasting observation, that the Top 4 has a precisely estimated delayed response to contracts, I conclude that defense contracts are not anticipated by the stock market, or else, the stock market responds to shocks to defense contracts.

In summary, defense contracts provide an accurate measurement of the timing of fiscal shocks, in contrast to the delayed measure of G constructed by NIPA. Defense contracts indeed anticipate spending by three to four quarters, a result of NIPA's accounting practice of recording contracts after-payment-on-delivery for most complex defense items. It is therefore unsurprising that shocks to defense contracts Granger-cause BP shocks but not viceversa. Moreover, production appears to begin following the award of new contracts, as evidenced by the quicker response of contracts compared to inventories in response to defense news shocks. Further supporting the argument that contracts are not anticipated at a quarterly frequency, the stock prices of major defense contractors exhibit a delayed response to a shock to defense contracts, while the contracts themselves do not respond to shocks to stock prices.

Having established that defense contracts accurately measure the timing of shocks, I will now focus on their exogeneity and statistical power as an instrument for G.

Exogeneity: The first condition for a valid instrument is exogeneity. Defense contracts capture variations in future military procurement spending; in turn, military spending is primarily influenced by exogenous events. However, new contracts might have been awarded in response to a recession. Similarly, higher-than-usual deficits, resulting from slower economic growth, could have prompted endogenous reductions in defense procurement spending. To address these concerns, I conduct two Granger Causality Tests.

I construct shocks to defense contracts as OLS residuals from a regression of defense contracts on four lags of defense contracts, G, GDP, hours worked in the private sector, and TB3. Essentially, this is a VAR with contracts, GDP, G, hours, and TB3, and contracts are ordered first. All nominal variables are expressed as logs of real per capita values, with hours in logs.

Table 2 shows the results of the Granger causality tests.

Table 2: Granger Causality Test

Predicted	*Predictor*	*F*	*Pvalue*	*Sample*
Defense Contract Shocks	NBER Recessions	1.16	32.06%	1947:1 - 2019:4
Defense Contract Shocks	NBER Recessions	0.59	78.32%	1951:1 - 2019:4
Defense Contract Shocks	Deficit	0.51	84.69%	1947:1 - 2019:4
Defense Contract Shocks	Deficit	0.99	44.19%	1951:1 - 2019:4

Notes: Granger Causality test is a Wald test on the 8 lags of the predictor while controlling for 4 lags of the predicted variable. I construct deficit as the difference between government total expenditures less government total receipts (NIPA Table 3.1, Line 43); I use the NBER based recession indicator to identify a recession in a quarter.

The results indicate that neither a recession nor deficit have predictive power on a shock to defense contracts. Results are also shown for the sample starting from 1951:1, which misses the beginning of the Korean war, as data from 1947:1 to 1950:4 is extrapolated.

Furthermore, notice that these results complement the ones of Cox et al. (2022), who show that federal procurement spending was not affected by endogenous counter-cyclical policies like the American Recovery and Reinvestment Act of 2009 and the COVID relief packages.

Statistical Power: I now turn attention to the statistical power of defense contracts. I do so by estimating a quarterly VAR which includes defense contracts, G, GDP, total hours worked in the private business sector and the 3 months T-Bill rate. I order defense contracts first in the VAR and look at the IRFs of the NIPA measure of G to a shock to defense contracts, using different samples. Results are reported in Figure 7.

Observing the left panels, it is clear that when the sample incorporates the Korean War, defense contracts capture a considerable proportion of the variation in G. Following the critique of Perotti (2014), I also ensure that once the Korean War is removed from the sample, the instrument preserves its statistical power. In the right panels, where the sample starts from 1954, defense contracts still capture a considerable variation in G. This represents a great advantage compared to the other instruments for G, like defense news shocks (Ramey (2011), Fisher and Peters (2010)'s shocks of cumulative excess returns of top defense contractors and Ben Zeev and Pappa (2017)'s shocks of expected defense spending obtained via medium horizon restrictions.

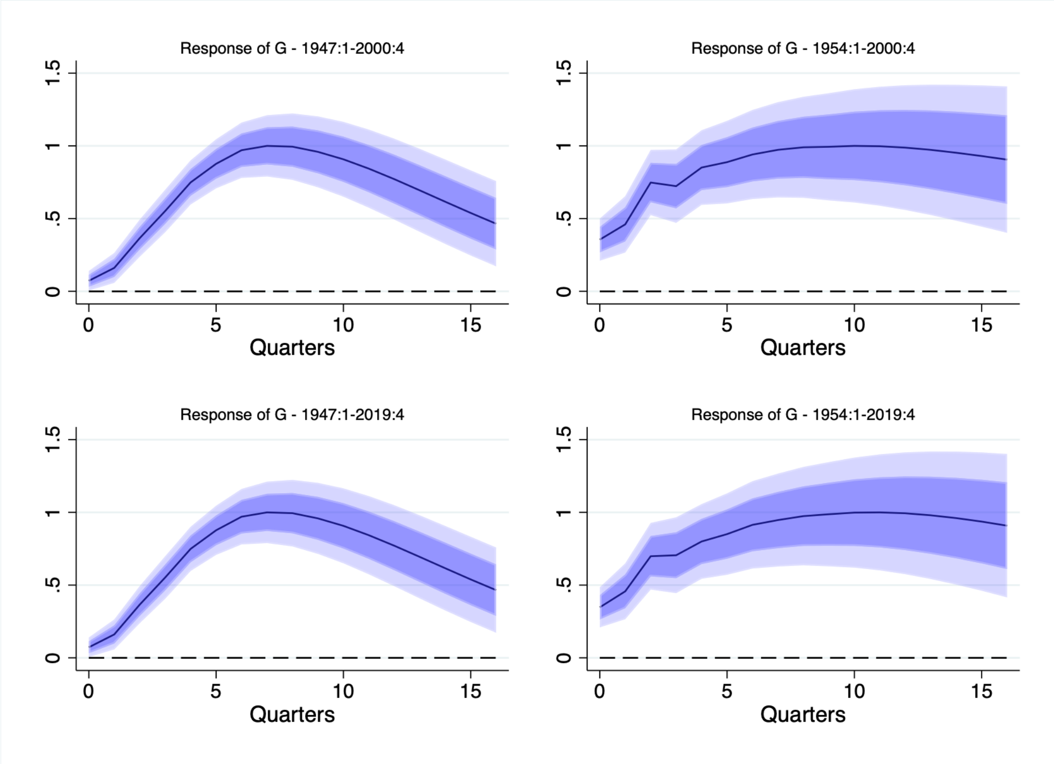


Figure 7: Response of G to Fiscal Shocks (Instrument's Power)

Notes: Confidence bands are 68% and 90%. Deflator is the GDP price deflator.

Avoiding Narrative Analysis: Finally, an advantage of defense contracts is that it does not require a narrative analysis. Even though narrative analysis are extremely useful to increase the stock of institutional knowledge about specific aspects of the economy, they are the result of subjective judgment of the researcher. On the contrary, defense contracts come from official government data.

Moreover, data on procurement contracts is becoming more and more available in most OECD countries, thus extending the applicability of this methodology to a panel of countries.[12]

[12]For example: Brazil (see Ferraz, Finan, and Szerman (2015)), South Korea (see Lee (2017)), Austria (see Gugler, Weichselbaumer, and Zulehner (2020)), France (see Pinardon-Touati (2022)), Portugal (see Gabriel (2022)), Spain (see Gugler, Weichselbaumer, and Zulehner (2020)).

III. Empirical Results

In this section, I explore the effects of shocks to defense contracts, identified by ordering defense contracts first in a VAR.

GDP Breakdown: First, I break down the response of GDP in Fixed Investments, Inventories, Durable Consumption, Non-Durable-plus-Service Consumption and G.[13] The VAR always includes defense contracts, G, GDP, hours worked in the private sector and the TB3, then rotates in and out a variable of interest. Nominal variables are in logs of real GDP per capita, while hours are in logs. Figure 8 displays the IRFs thus calculated for the components of GDP in response to a shock to defense contracts for the sample spanning from 1947:1 to 2000:4.

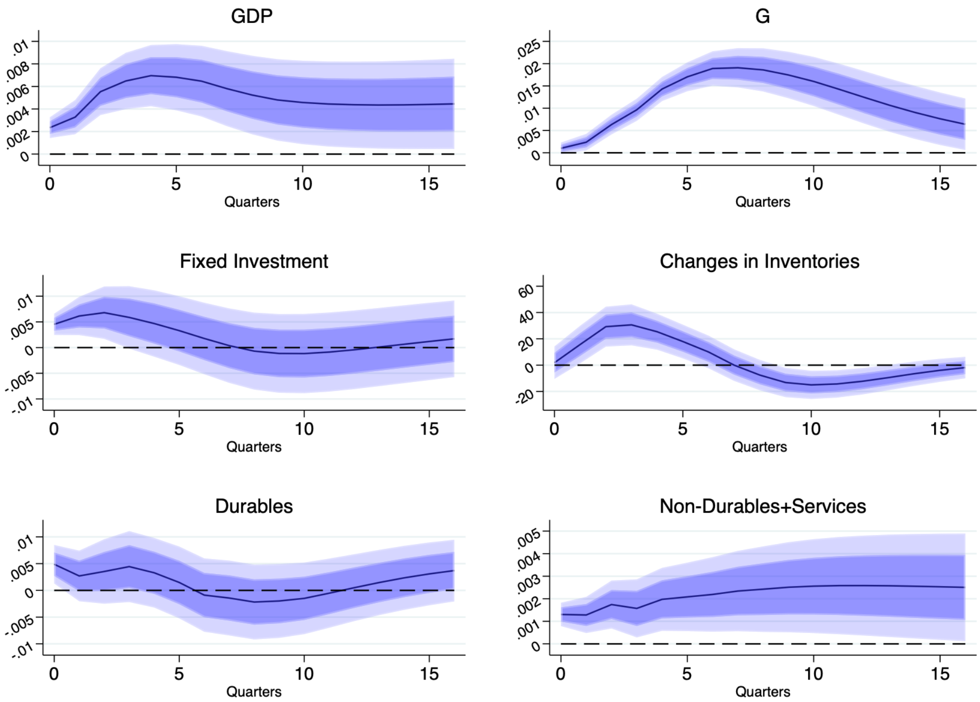


Figure 8: IRFs of GDP Components to Defense Contracts

Notes: Sample goes from 1947:1 to 2000:4. Confidence bands are 68% and 90%. Nominal variables are in logs of real GDP per capita, while hours are in logs. Price deflator is the GDP price deflator.

[13]The response of net-export and import is analyzed in the Online Appendix B.3 and it is non significant at any horizons and any sample.

The top-left panel shows that GDP increases on impact and spikes at 3 quarters from the shock, then starts falling down to zero. The top-right panel displays how government spending slowly increases, spiking 6-7 quarters after the shock. The middle right panel shows the response of inventories, which increases rapidly, spiking after 2-3 quarters from the shock. Despite being the most volatile component of GDP, I find that inventories respond strongly and significantly. This result complement the findings of Brunet (2022) and Briganti and Sellemi (2023). Note that the unit of inventories is different from the one of the other components of GDP,; in fact, inventories can take on negative values, therefore, I use real changes per capita instead of logs. A VAR specification with the real variables divided by real potential GDP - Gordon and Krenn (2010) transformation - leaves the result unaffected.

The bottom-right panel shows the response of *non-durables-plus-service consumption*, which is positive and significant. The positive response of (non-durable-plus-service) consumption, is an important result. In fact, proponents of the instrument approach, have consistently found negative responses of consumption.[14] At the same time, proponents of the SVAR approach, who employ BP shocks, consistently find positive responses of consumption, even if almost all works report 68% confidence bands. However, proponents of the instrument approach are not convinced by these findings since the timing of BP shocks is delayed (Ramey (2011)). Similarly, proponents of the SVAR approach, highlight that the results obtained with the military instruments are sensitive to the exclusion of the Korean war from the sample (Perotti (2014)) and lack statistical power in samples after the Korean war (Ramey (2016)). Therefore, the literature has not reached an agreement about the effects of fiscal shocks on consumption.

However, defense contracts accurately measure the timing of the shocks, addressing the limitation of the SVAR approach, and preserve statistical power after the Korean War, addressing the limitation of the instrument approach. Additionally, the response of non-durable-plus-service consumption, which accounts for 83% of aggregate consumption on average, is robust to the exclusion of the Korean War. Since defense contracts successfully address the major concerns of the most widely employed methods to study the effects of government spending, I argue that my results provide a more conclusive answer to the long-debated question of the effect of government spending on consumption.

[14] A notable exception is Fisher and Peters (2010) who find positive responses of consumption in sample which excludes the Korean war. However, their instrument lacks statistical power (Ramey (2016)) and the reported IRF of consumption is barely significant at 68% confidence level.

Fixed Investments and Durables: The response of fixed investments and durables deserves a separate discussion. First, the middle-left panel of Figure 8 shows the response of fixed investments. The response is positive on impact and then falls to zero. Second, the bottom left panel portrays the response of durable consumption, which accounts for 17% of total consumption, on average. The response is positive and significant only on impact.

Upon excluding the Korean War from the sample, I observe a noticeable variability in the responses of durables and fixed investments. This phenomenon extends to BP shocks and defense news shocks, which demonstrate an even more pronounced sensitivity in their results. The response of fixed investments and durable consumption for the sample 1954:1 to 2000:4 is showed in Figure 9.

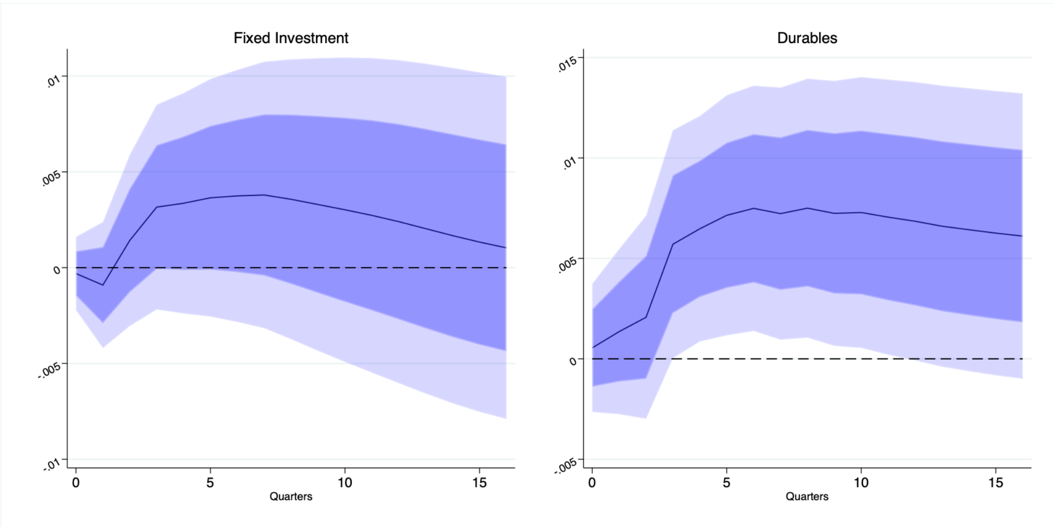


Figure 9: Fixed Investments and Durables after Korea

Notes: Sample goes from 1954:1 to 2000:4. Confidence bands are 68% and 90%. Nominal variables are in logs of real GDP per capita, while hours are in logs. Price deflator is the GDP price deflator.

The left panel of Figure 9 shows a persistent increase in fixed investments after a positive shock to defense contracts, but significant only at 68%. The right panel shows that the response of durable consumption also increases and becomes significant even at 90% confidence.

The sensitivity of these two GDP components to the exclusion of the Korean War from the sample has been underscored in Perotti (2014) and Ramey (2016). Specifically, during the onset of the Korean War in the last two quarters of 1950, consumers, with the fresh memory of WWII rationing, rushed to purchase durable goods in large quantities (refer to Ginsburg (1952), Hickman (1955), and Ramey

(2016) for more details).[15] The introduction of Regulations X and W aimed to mitigate the inflationary pressure resulting from the buying wave, adversely impacting the medium-term response of residential investments and home-building, as well as the consumption of durable goods like furniture and household equipment. Consequently, the responses of fixed investments and durable goods are sensitive to the exclusion of the Korean War from the analysis.

To summarize, in response to a positive shock to defense contracts, GDP, G, inventories and non-durables-plus-service consumption increase significantly, while the response of fixed investments and durables also exhibit positive responses after the Korean war.

Robustness: The results for GDP, G, inventories and non-durables-plus-service consumption are robust to using the samples 1954:1-2000:4 (without Korean war) and 1947:1-2019:4 (full sample). Moreover, the inclusion of a tax control or a quadratic time-trend also do not affect the baseline results reported here.

Lastly, given that the direct estimation of fiscal multipliers requires real variables to be transformed using the Gordon and Krenn (2010)'s transformation (Ramey (2016)), I check the results using a VAR where the real variables are divided by real potential GDP.[16] The estimated multipliers has the interpretation of the ratio of the area under the IRF of GDP and the area under the IRF of G. Results are robust to using this specification.

All these robustness checks are reported in Appendix A.2.

Fiscal Multiplier: In the Online Appendix B.1, I illustrate and derive my estimate of the cumulative fiscal multiplier, obtained with LP-IV (see Ramey (2016)). My baseline estimate for the 4-years GDP multiplier is 0.92, while the non-durable-plus-service consumption multiplier is 0.12. This means that if government spending increases by 1$, non-durable-plus service consumption increase by 0.12$. Results and discussions on fiscal multipliers are remanded to the Online Appendix B.

III.a Wages, Hours, Employment, Production Earning and Income

In this section I explore the effect of a shock to defense contracts on the product wage, hours, employment, production earnings and income.

[15] This mechanism of intertemporal shift of consumption of durable goods is explored also in McKay and Wieland (2021) in the context of raising interest rates.

[16] In particular, the VAR includes defense contracts, GDP, G, Hours worked in the private sector, 3 months T-Bill rate. Nominal variables are deflated by the GDP price deflator and divided by potential GDP. Potential GDP is estimated with a 6-degree polynomial fit to log of real GDP.

Product Wage

The fiscal policy literature has paid significant attention to the response of the product wage to fiscal shocks. According to economic theory a fiscal shock is associated with higher taxes, which trigger a negative wealth effect, shifting the labor supply curve to the right. As long as labor demand remains constant, the product wage is expected to decrease, resulting in a reduction in consumption within standard DSGE models.[17] Conversely, if labor demand also shifts to the right, it creates the possibility for the product wage to increase, fulfilling a necessary condition for a rise in consumption within DSGE models.

However, the literature has not reached a consensus on the effects of fiscal shocks on the product wage. For example, Ben Zeev and Pappa (2017) find negative responses of the manufacturing product wage using the instrument approach, while Monacelli and Perotti (2008) find positive responses using the SVAR approach.

In light of the low statistical power of the instruments for G and the delayed timing of BP shocks, I add to the literature by also exploring the response of product wages, at different aggregation levels, using shocks to defense contracts, which preserve power across samples and accurately measure the timing of the shocks. In particular, I look at four measures of product wage. First, I use the average hourly wage of aircraft manufacturing from the discontinued BLS data, divided by the Producer Price Index (PPI) of durable manufacturing. Secondly, I construct two measures of manufacturing hourly product wage. Specifically, I use the hourly earnings of production workers in manufacturing from the BLS, divided by the PPI of manufacturing. Additionally, I use the NIPA total wages and salaries in manufacturing divided by total hours of manufacturing production workers from the BLS, divided by the manufacturing PPI. Lastly, I construct the product hourly wage in the private economy by dividing the NIPA wages and salaries in the private sector by the total hours worked in the private sector, then deflated with the GDP price deflator.

As in previous sections, I rotate in and out the log of the four measures of product wages in the baseline VAR augmented with the tax control. Figure 10 illustrates the IRFs to a positive shock to defense contracts for the sample 1947:1-2000:4.

[17]The product wage is preferred over the consumption wage based on the theoretical results presented by Ramey and Shapiro (1998). In their two-sector model, an increase in defense purchases increase the relative price of manufacturing goods, subsequently lowering the manufacturing product wage while increasing the overall consumption wage. As a result, the product wage is the appropriate metric to determine whether labor demand or labor supply experiences a more pronounced shift following a government spending shock.

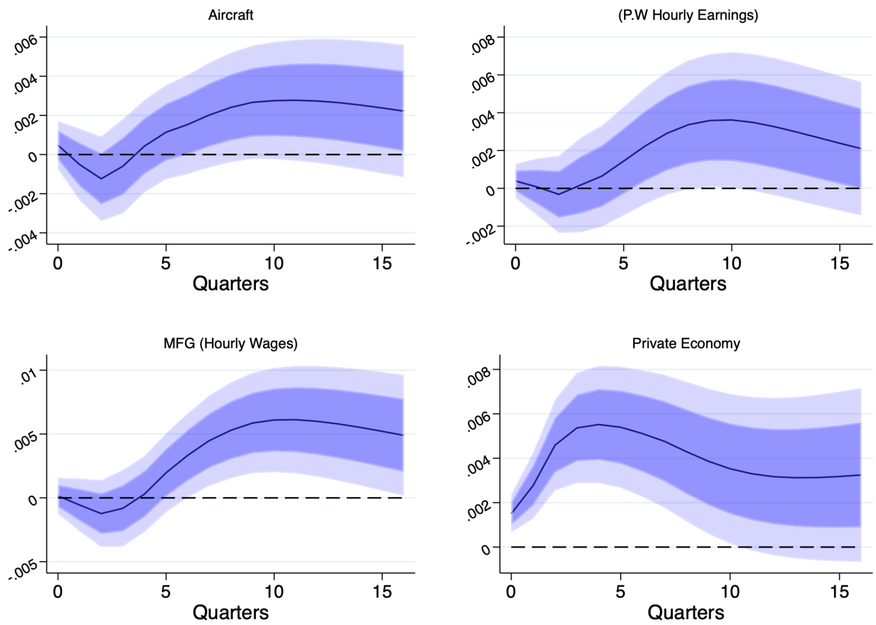


Figure 10: Response of (Product) Wage to Contracts

Notes: Sample goes from 1947:1 to 2000:4. Confidence bands are 68% and 90%. The VAR contains defense contracts, G, GDP, hours worked in the private sector, the TB3, total tax receipts and a sixth rotating variable of interest. Nominal variables are in logs of real per capita GDP. Deflator is the GDP price deflator. Hourly product wages are in logs, as well as hours.

The real product wage reveals a near-zero response in the short horizons, gradually rising in medium to long horizons.

Robustness: Similar outcomes are apparent in other samples. Most importantly, the results without the tax control look qualitatively identical but it is estimated with less precision: the IRFs are not statistically significant, except for the private economy case. I also replicate the VAR not in logs but with the variables Gordon-Krenn transformed, with and without taxes. The results are very similar and also indicate to a delayed positive response. All robustness checks are detailed in the Online Appendix D.2.

Hours, Employment, Earnings and Income

I now turn my attention to the responses of hours, employment, production earnings, and disposable income. To this end, I augment the baseline VAR from the previous section with various outcomes of interest: the log of average weekly hours worked, the log of employment, the log of total hours worked, and the log of real total earnings per capita. I study their response at three different levels of aggregation: (i) aircraft manufacturing, (ii) total manufacturing, and (iii) the total private economy. I focus on the responses in aircraft manufacturing and total manufacturing due to the pronounced sectoral bias of defense purchases toward these industries (see Ramey and Shapiro (1998), Perotti (2007), Nekarda and Ramey (2011), and Cox et al. (2022)). At the same time, examining the response in the total private economy is crucial, because sectoral reallocation of workers and crowding-out effects might negate, in the aggregate, the positive effects observed in directly affected industries.

Figure 11 shows the IRFs of these variables with respect to contracts: first column is aircraft manufacturing, followed by manufacturing and private economy respectively.

Aircraft Manufacturing: data from aircraft manufacturing is available from 1939 at monthly frequency from the BLS' discontinued database.

Firstly, weekly hours are measured by weekly hours of production workers, and respond positively to a shock to government contracts. As noted by Bils and Cho (1994) and Fernald (2012), average weekly hours worked is an excellent proxy for the intensity of capital utilization. Therefore, its immediate response indicates a rapidly increased production after a shock to defense contracts. Secondly, employment is thousands of production workers and also responds positively, with a delayed response, probably reflecting labor market frictions. Thirdly, total hours worked is the product of weekly hours worked and number of production workers, and responds positively. Lastly, I calculate total earnings, derived from multiplying the hourly wage of production workers with the total hours worked. The response of this variable is positive.[18]

Manufacturing: Similar to the aircraft sector, the total manufacturing data is obtained from the BLS's discontinued monthly data. Like the aircraft data, all variables show a positive response to an increase in defense contracts.

This pronounced positive reaction in the manufacturing sector indicates that government funds positively impact industries outside of just aircraft manufacturing, for at least three reasons. First, the government's demand extends to a range of products including motor and space vehicles, ships, IT equipment, ammunition, and clothing (Ramey and Shapiro (1998), Perotti (2007), Nekarda and Ramey (2011) and Cox et al. (2022)). Second, as highlighted by Gulledge and Womer

[18]Total earnings of production workers in aircraft manufacturing account for 0.5% of potential GDP, on average.

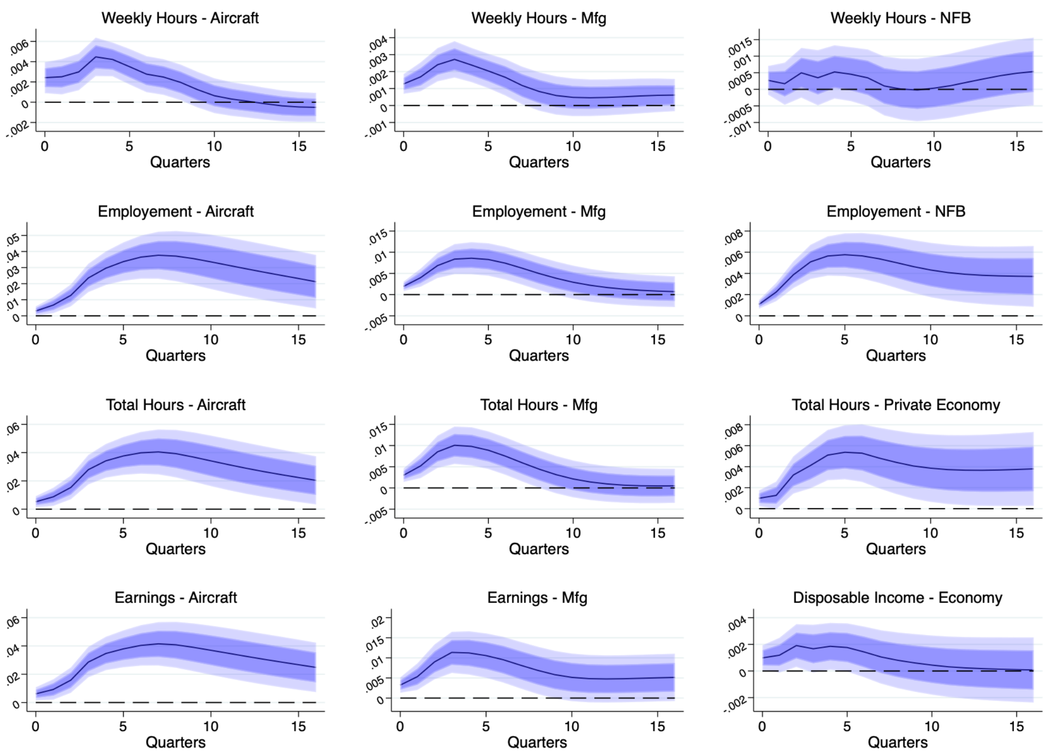


Figure 11: Response of Hours, Employment and Income to Contracts

Notes: Sample goes from 1947:1 to 2000:4. Confidence bands are 68% and 90%. The VAR contains defense contracts, G, GDP, hours worked in the private sector, the TB3 and a sixth rotating variable of interest. Nominal variables are in logs of real per capita values, while hours is in logs. Price deflator is the GDP price deflator. The rotating variables weekly hours, hours and employment are expressed in logs (when aggregate hours are analyzed, the VAR only has 5 variables). Total production earnings and disposable income are in logs of real per capita values.

(1986) and the Top 100 companies reports, subcontracting was common, meaning that prime military contractors often delegate work to smaller, specialized firms, thereby distributing the initial increase in demand to various other players in the industry. Lastly, the input-output relationships in the manufacturing network amplify the effects of a spike in defense contract demands, causing a ripple effect upstream in the production network (Acemoglu, Akcigit, and Kerr (2016)). For example, the production of the F-4 Phantom II during the Vietnam war, involved General Electric, which manufactured engines, while other firms such as Alcoa, provided crucial materials like aluminum and titanium.

Aggregate: The last column presents the results for the entire private economy. First, average weekly hours worked in the non-farm business (NFB) sector is constructed by dividing total hours worked of all employees by the total number of employees. The IRF of this variable portrays positive, albeit non-significant, point estimates. Second, the data exhibits a consistent positive response in the total number of NFB employees across different samples, dispelling the concerns of potential crowding-out effects on employment. Third, I examine the total hours worked within the private sector, which serves as a baseline variable in my VAR analysis. Although the response indicates a positive trend, it is significant only on impact. Other samples echo this finding, pointing to a generally mild positive reaction in total hours worked. Lastly, production workers data for the economy is available only from 1964, therefore, I use NIPA's disposable income, given also my interest in the response of consumption. Disposable income has a robust positive response, which is consistent with the rise in aggregate consumption.

Robustness: All results are robust to the exclusion of the Korean war, sample 1954:1 to 2000:4 and the inclusion of the most recent years, sample from 1947:1 to 2019:4. I also control for a quadratic trend as well as taxes and the results are unaffected. Furthermore, I carry out the analysis for all three samples using a VAR with the nominal variables Gordon-Krenn transformed, the results are also the same. All the robustness checks are illustrated in the Online Appendix D.1.

In summary, the response of product wages, hours, employment, earnings and disposable income to an increase in defense contracts is positive and consistent with the observed rise in non-durable-plus-service consumption. The positive response of total hours and employment is not new, since similar responses have been consistently found in the literature. However, here I analyze these variables at different aggregate levels; notice that the magnitude of the response decreases with the aggregation level, indicating that the effects of a shock to defense contracts are more concentrated in the sectors directly affected by the shock. The response of weekly average hours of production is - to the best of my knowledge - new in the context of time series regressions.[19] and serves as a proxy for capital utilization/production rates. Production earnings and disposable income are studied to substantiate the observed positive response of consumption. I am not aware of other time series studies which analyzed the response of production earnings in aircraft and total manufacturing; on the contrary, the response of disposable income is analyzed in Galí and Monacelli (2008). They find a positive and significant response of both disposable income and consumption; however, they use BP shocks and display IRFs with 68% confidence bands.

[19] Nekarda and Ramey (2011) study weekly hours in a cross-sectional industry-level framework.

III.b Markup and Labor Productivity

In the previous section the observed joint rise in hours worked and the product wage in response to a shock to defense contracts indicate that labor demand shifts to the right. According to economic theory, labor demand can increase due to sticky prices and/or an increase in labor productivity.

To better illustrate this point, consider the standard textbook New Keynesian model (Chapter 3 of Galí (2015)). Here, the aggregate real marginal cost (MC_t^r) is equal to the real product wage (W_t^r) over the marginal product of labor (MPN_t):

$$\mathrm{MC}_t^r = \frac{W_t^r}{\mathrm{MPN}_t} \implies \hat{mc}_t^r = \hat{w}_t^r - \hat{mpn}_t$$

where the ˆ notation denotes percent deviations from the steady state. By definition, the price-cost markup is the ratio of prices over marginal cost, therefore, using the hat notation, the markup, $\hat{\mu}_t$ is the negative of the real marginal cost: $\hat{\mu}_t = -\hat{mc}_t^r$. Using the two expressions allows me to write the real product wage as a function of the marginal product of labor and the price-cost markup:

$$\text{Labor Demand: } \hat{w}_t^r = \hat{mpn}_t - \hat{\mu}_t.$$

Now, from the household problem, the labor-leisure intratemporal condition is:

$$\underbrace{-u_N(C_t, N_t)}_{\psi \cdot N_t^{\varphi}} = \underbrace{u_C(C_t, N_t)}_{C_t^{-\sigma}} \cdot W_t^r \implies \text{Labor Supply: } \hat{w}_t^r = \varphi \cdot \hat{n}_t + \sigma \cdot \hat{c}_t$$

where the parametric form of the marginal disutility of labor, $-u_N$, and the one of the marginal utility of consumption, u_C, follows, for the sake of the illustration, by assuming a simple isoelastic and separable utility function.[20]

Putting everything together yields:

$$\varphi \cdot \hat{n}_t \uparrow + \sigma \cdot \hat{c}_t \uparrow = \hat{mpn}_t - \hat{\mu}_t.$$

When G rises above its steady state level, hours increase due to a negative wealth effect resulting from heightened lump-sum taxes.[21] In this context, consumption will increase if any of the following scenarios occur: (i) labor productivity rises, (ii) the price-cost markup decreases, (iii) both labor productivity rises and the price-cost markup decreases, or (iv) the price-cost markup increases but at a slower rate than the rise in labor productivity.

This simple theoretical example aims to highlight the importance of examining the responses of both the price-cost markup and labor productivity in order to rationalize the observed increase in consumption.

[20] For example: $U(C_t, N_t) = (C_t^{1-\sigma}/(1-\sigma) - \psi \cdot N_t^{1+\varphi}/(1+\varphi)$.

[21] I acknowledge that in the presence of dividends, the representative household experiences a substantial additional negative income effect through the decrease in profits, which is caused by the conditionally counter-cyclical markup, as noted by Broer and Krusell (2021).

Price-Cost Markup:

I investigate the response of the price-cost markup by rotating in and out four measures of the markup in a VAR. Firstly, I construct the markup in the manufacturing sector as in Monacelli and Perotti (2008).[22] Secondly, I use the negative of the log share of labor income in the non-financial-corporate-business (NFCB) sector, also analyzed by Monacelli and Perotti (2008). The third and fourth measures are the negative of the log-share of labor income in the economy and in the non-farm-business (NFB) sector, taken from Nekarda and Ramey (2020)'s online database.[23]

Figure 12 shows the responses of these four measures of the markup to a positive shock to defense contracts for sample 1947:1 to 2001:4. The VAR employed here mimics the one of Nekarda and Ramey (2020), it contains defense contracts, G and GDP in logs of real per capital values, the log of the GDP price deflator and the TB3.

The markup exhibits a positive response at short horizons, then diminishes and turns negative across all measures. The positive response of the markup is consistent with the findings of Nekarda and Ramey (2020) who use defense news shocks on a sample from 1947:1 to 2017:4. I extend their results to samples without the Korean war and for the manufacturing price-cost markup measure. On the contrary, Monacelli and Perotti (2008) find negative responses of the markup using BP shocks. Therefore, in Online Appendix C I look at the response of the four measures of the markup in response to a BP shock, mimicking the approach of Monacelli and Perotti. I find that the response of the markup in NFCB and manufacturing aligns with those found in their paper, with the markup declining. Yet, these findings are not robust when the Korean war is excluded from the sample; in this case, the markup does not show a significant response. Furthermore, when the markup is quantified as the negative of the log share of labor income in the economy and the NFB, BP shocks lead to positive responses of the markup.

Robustness: I also look at the response of the mark-up in other samples (1954:1-2000:4 and 1947:1-2019:4) and find similar results. All robustness checks as well as the results with bP shocks are reported in Online Appendix C.

Overall, since defense contracts accurately measure the timing of the shocks

[22] I follow the indications in the appendix of their paper and construct the markup by taking the log of the ratio of manufacturing national income less capital consumption adjustment and manufacturing wages.

[23] I am aware of the recent criticisms moved towards mark-up measures based on the log-shares of variable input of production (see Bond et al. (2021)). However, in the absence of better measures of the price-cost markup, I comply with what the fiscal policy literature has used so far.

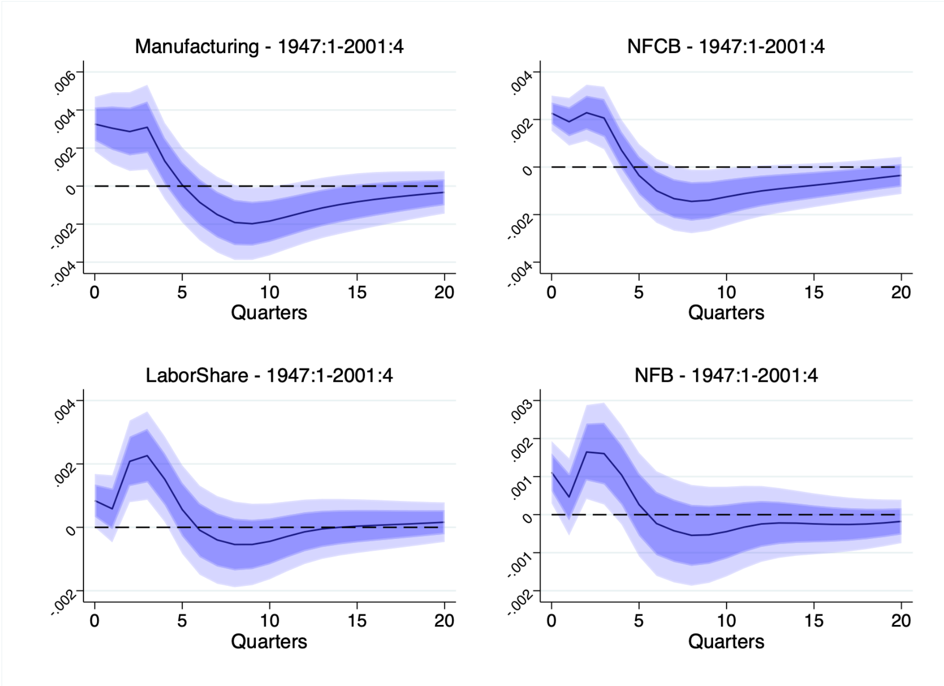


Figure 12: Response of Markups to Contracts

Notes: Sample goes from 1947:1 to 2000:4. Confidence bands are 68% and 90%. Price deflator is the GDP price deflator.

and provide results robust to the exclusion of the Korean war, I argue that the price-cost markup increases after a positive fiscal shock.

Labor Productivity: I now turn attention to the response of labor productivity. More specifically, I look at the response of the the log of output-per-hour (OpH) in the private business sector, a measure of labor productivity, and Total Factor Productivity (TFP), as measured by Fernald (2012). I rotate in and out those variables in my baseline VAR, with the log of real per capita values of defense contracts, GDP and G, the log of hours in the private sector and the TB3.

Figure 13 shows the response of OpH and TFP to a positive shock to defense contracts using the baseline sample 1947:1 to 2000:4.

Both variables display a significant positive response.

Robustness: In the Online Appendix D.3 I also study the response of OpH in the NFB and the NFCB. The results indicate a positive response at short horizons.

I also replicate the analysis using the other samples for TFP and OpH in NFB,

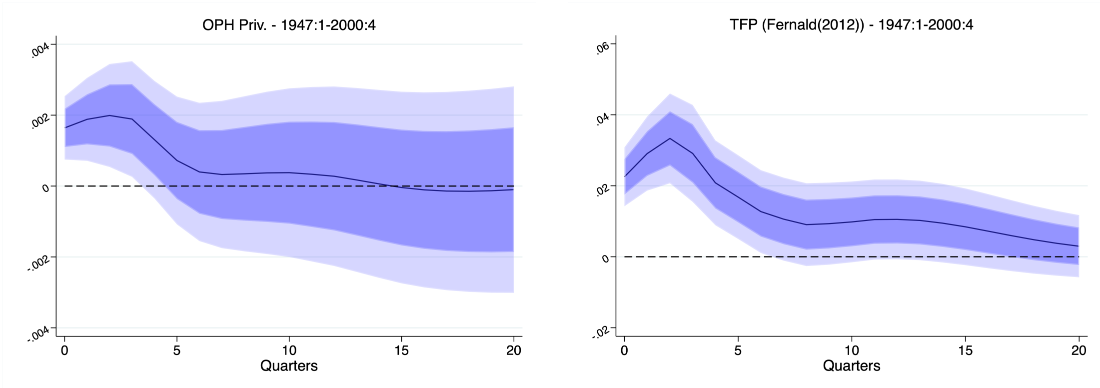


Figure 13: Response of Productivity to Contracts

Notes: Left panel is the response of the log of output-per-hour (OpH) in the private sector, a measure of labor productivity. Right panel shows the response of TFP as measured by the cumulative quarterly growth rates of TFP as measured by Fernald (2012). Sample goes from 1947:1 to 2000:4. Confidence bands are 68% and 90%. Price deflator is the GDP price deflator.

NFCB and the private sector, finding very similar results. The major discrepancy is when the Korean war is removed from the sample: all three measures of OpH as well as TFP display a slower and hump-shaped response, which preserves statistical significance. I also find that the inclusion of a tax control (log of real total tax receipts per capita) does not affect the results, or else, it makes the responses even more significant and positive.

All these robustness checks are documented in the Online Appendix D.3.

Overall, the findings presented in this section suggest a boost in productivity following a positive shock to defense contracts. Notice that the rise in productivity is not necessarily evidence of increasing return to scale. In fact, Basu and Fernald (1997) propose that government spending shocks prompt a reallocation of resources, particularly labor, toward manufacturing. If the manufacturing sector is more productive than other sectors, this shift could lead to an apparent increase in overall productivity, creating the illusion of increasing returns to scale.[24]

While sectoral reallocation might explain increases in productivity at the aggregate level, it does not align with the findings of Christiansen and Goudie (2007). They merge contract data from the Top 100 companies' annual Department of Defense reports with balance sheet information from Compustat, constructing an an-

[24]It is noteworthy that their research specifically identifies durable manufacturing, the main beneficiary of defense spending, as the sole sector exhibiting increasing returns to scale, even when accounting for sectoral reallocation effects.

nual panel database of publicly traded defense contractors. By estimating a panel VAR from 1969 to 1996, they observe positive effects of new contract awards on sales, employment, and sales-per-employee, a proxy for labor productivity. Their results suggest that defense contracts do, in fact, stimulate productivity gains at the firm level.

In the subsequent section, I propose that a significant source of these productivity gains for manufacturers and contractors is attributable to learning-by-doing.

III.c Showcase: The Vietnam War

I now briefly discuss the empirical evidence during the Vietnam war in order to conceptualize the VAR findings in the context of a major exogenous shock to defense spending in the post WWII sample.

The February attack of 1965 marked the beginning of the US military's escalation in the Vietnam war (Ramey and Shapiro (1998)). Figure 2 illustrates in blue the quarterly year-to-year growth rates of defense contracts. A substantial increase was witnessed in the second quarter of 1965, reaching its peak in the final quarter of that year. The top-left panel shows the growth rates of non-durable-plus-service consumption while the top-right panel illustrates output-per-hour growth rates in the private sector. Both variables exhibit a surge concurrent with the increase in defense contracts.[25]

Analysis of the Survey of Current Business (SCB) during the Vietnam War era highlights that the uptick in defense production was a substantial economic stimulus at that time. From the SCB of January 1967:

"*Heavy defense purchases last year accounted for most of the rise in Federal outlays from 1965 to 1966 and were the dominant stimulus to rising activity in the second half of the year. [...] As in the summer months, government purchases continued to be a major stimulus to the rise in production*".

Simultaneously, a surge in personal income, largely attributed to escalated production, seemingly fostered the observed growth in consumption during those periods, as reported in the SCB January 1967:

"*Another large increase in personal income accompanied the continued strong advance in economic activity in 1966. The flow of income reflected essentially the*

[25] Real service and non-durable consumption per capita during the Korean and Vietnam wars display consistent above-trend values (trend estimated with either Hamilton (2018)'s filter or a polynomial filter). Additional details are available in the Online Appendix A.3.

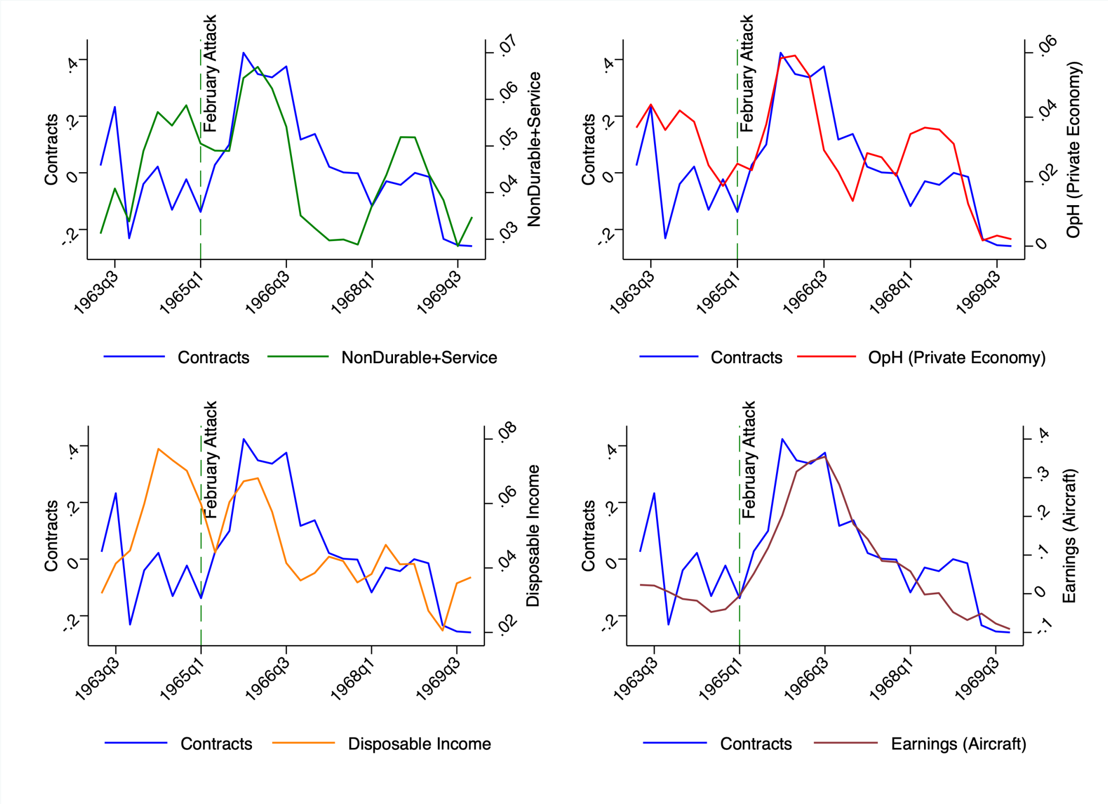


Figure 14: Growth Rates of Macro Aggregates During Vietnam

Notes: The GDP price deflator is used to deflate nominal variables. Quarterly year-to-year growth rates are calculated as $(x_t - x_{t-4})/x_{t-4}$.

large rise in earnings from current production...".

The bottom-left panel delineates NIPA's disposable income growth rates, illustrating a notable increase concurrent with the heightened contracts in 1965 and 1966, consistently with the SCB narrative.

In the same period, the Top100 companies reports show that major Vietnam war contractors were predominantly aircraft and parts manufacturers.[26] Consequently, the bottom-right panel showcases the earnings growth rates in the aircraft manufacturing sector, indicating a steep ascent beginning in the first quarter of 1965, stemming from increased production. Owing to their substantial engagement

[26]In FY1967, the top 7 contractors were McDonnell-Douglas Aircraft, General Dynamics, Lockheed Aircraft, General Electric, United Aircraft, Boeing and North American Aviation, collectively accounting for 25% of total defense contracts. General Dynamics was primarily selling aircraft, such as the F-111. While General Electric was producing aircraft engines.

with government acquisitions, examining the aircraft manufacturing dynamics can offer critical insights into the direct effects of government purchases. This aspect is further explored in Figure 15, which shows labor market variables within aircraft manufacturing during the Vietnam war.

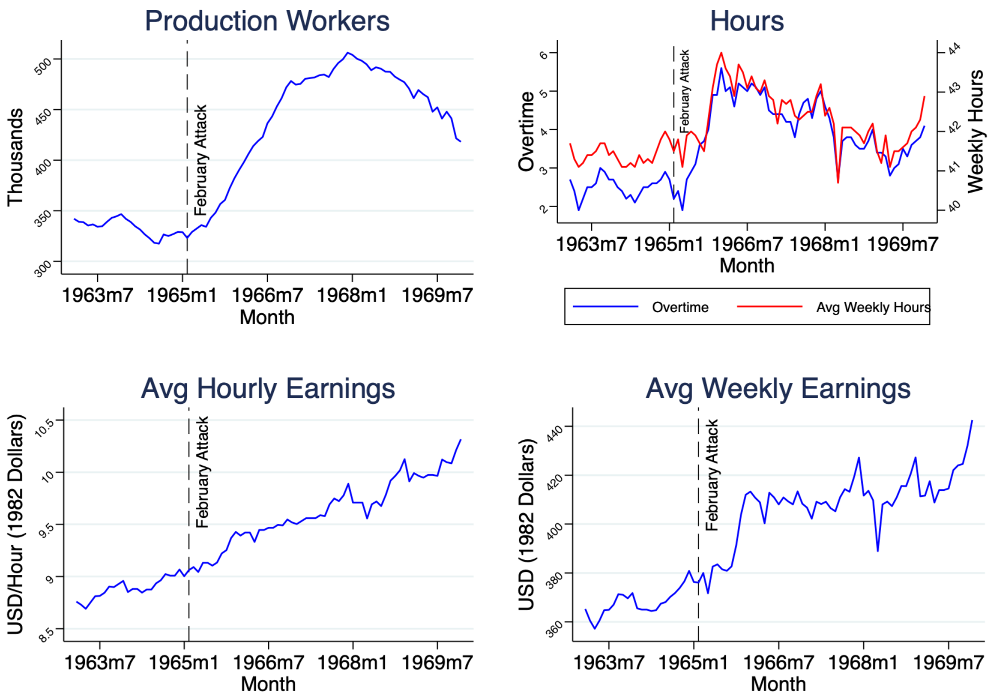


Figure 15: Aircraft Manufacturing in the Vietnam War

Notes: The data, obtained from the BLS's discontinued database, originates from the Current Employment Statistics survey. Price deflator of average hourly and weekly earnings is the PPI of durable goods (1982=100).

Following the February attack, the industry witnessed an uptick in hiring, as evidenced in the top-left panel of the graph.[27] Between January 1965 and January 1968, the sector's workforce expanded by over 150,000 individuals, representing substantial portions of the *total* production workforce in the private and manufacturing sectors.[28] The escalation in production necessitated increased working hours, including overtime, as indicated in the top-right panel. This period also saw a modest rise in real average hourly earnings (highlighted in the bottom-left

[27]Production workers represent approximately 82% of total employment.

[28]More precisely, those *changes* represented about 0.4% and 1.1% of *total* production workers of January 1965 of the private and the manufacturing sector, respectively.

panel). Consequently, real average weekly earnings soared, exhibiting an 11.4% year-on-year increase in the first quarter of 1966, significantly exceeding the 3% annual growth observed from 1960 to 1965.

In summary, this evidence corroborates the broader narrative presented in the SCB's January 1967 edition and confirms the expansionary direct effects of government purchases. The events of the Vietnam war are also consistent with the VAR evidence and help conceptualize those findings.

Labor Productivity and Manufacturing: The top-right panel of Figure 14 shows a contemporaneous rise in defense contracts and labor productivity, in line with the VAR evidence portrayed in Figure 13.

Why did labor productivity increase during the Vietnam War? If labor productivity was directly affected by increased military contracts, I would expect to observe productivity gains within the sector most affected by those contracts: the manufacturing sector. How much of an increase in labor productivity in the manufacturing sector is required to explain the rise in output per hour for the entire private sector experienced in those years?

To answer this question, consider the following back-of-the-envelope calculation. Think of an economy divided into two sectors: manufacturing (mfg) and non-manufacturing (non-mfg). The percentage increase in aggregate output-per-hour can be represented by the following equation:[29]

$$\Delta\%\mathrm{OpH}_t = \frac{Y_t^{\text{non-mfg}}}{Y_t} \cdot \Delta\%\mathrm{OpH}_t^{\text{non-mfg}} + \frac{Y_t^{\text{mfg}}}{Y_t} \cdot \Delta\%\mathrm{OpH}_t^{\text{mfg}} + \frac{\mathrm{OpH}_t^{\text{mfg}} - \mathrm{OpH}_t^{\text{non-mfg}}}{\mathrm{OpH}_t} \cdot d\left(\frac{N_t^{\text{mfg}}}{N_t}\right)$$

where Y denotes output, N employment and OpH, output-per-hour. Using average data from 1960 to 1965 (Pre-Vietnam) and the change around the outbreak of the Vietnam war, I have:

$$\underbrace{\Delta\%\mathrm{OpH}_{\text{Vietnam}}}_{\approx 6\%} = \underbrace{\left(1 - \frac{Y_{\text{Vietnam}}^{\text{mfg}}}{Y_{\text{Vietnam}}}\right)}_{\approx 1-40.9\%} \cdot \underbrace{\Delta\%\mathrm{OpH}_{\text{Vietnam}}^{\text{non-mfg}}}_{\approx 3.5\%(\text{Pre-Vietnam Avg})} + \underbrace{\frac{Y_{\text{Vietnam}}^{\text{mfg}}}{Y_{\text{Vietnam}}}}_{\approx 40.9\%(\text{1965 Use Tab.})} \cdot \underbrace{\Delta\%\mathrm{OpH}_{\text{Vietnam}}^{\text{mfg}}}_{\implies \approx 9.0\%} + \ldots$$

$$\ldots + \underbrace{\left(\frac{\mathrm{OpH}_t^{\text{mfg}} - \mathrm{OpH}_t^{\text{non-mfg}}}{\mathrm{OpH}_t}\right)}_{\approx \frac{30-22}{24.7} \text{ Compustat Sales-per-Employee}} \cdot \underbrace{d\left(\frac{N_t^{\text{mfg}}}{N_t}\right)}_{\approx 0.8\% \text{ Reallocation}}$$

[29] This follows from the definition of OpH and a log-linearization of the equation: $\mathrm{OpH}_t = \frac{N^{\text{non-mfg}}}{N} \cdot \mathrm{OpH}_t^{\text{non-mfg}} + \frac{N^{\text{mfg}}}{N} \cdot \mathrm{OpH}_t^{\text{mfg}}$.

In 1965, manufacturing output contributed 40.9% to the total private industry output.[30] The share of manufacturing employees grew by 0.8% during the Vietnam war. OpH before the Vietnam war is approximated with the average of sales of employee of all publicly traded firms from Compustat over 1960 through 1965. The same approximation is adopted for the manufacturing and non-manufacturing output-per-hour values.[31] OpH surged by approximately 6% in the Vietnam war's initial phase. Lastly, assuming that OpH in the non-manufacturing sector was growing at the same pace of the pre-Vietnam average of OpH in the private sector, i.e. 3.5%, the above formula suggests that a 9.0% boost in manufacturing output-per-hour was necessary to generate the observed 6% increase in aggregate OpH in the initial phase of the Vietnam war. Therefore, productivity increases in manufacturing are quantitatively capable of explaining aggregate fluctuations in total labor productivity.

Labor Productivity and Defense Contractors: If the rise in total labor productivity was driven by productivity gains in the manufacturing sector triggered by the effects of defense contracts, I would also expect to observe a rise in labor productivity of defense contractors. Therefore, similarly to what done in Christiansen and Goudie (2007), I combine balance-sheet data from the Annual Fundamental segment of Compustat with contract data from the Top 100 companies report for the period spanning from 1960 to 1971. I construct changes in sales per employee, a proxy for labor productivity, denoted as $\Delta\text{SpE}_{i,t}$, and changes in government contracts, expressed as $\Delta G_{i,t}$, where i represents a firm and t a year. I use OLS to estimate the following equation:

$$\Delta\text{SpE}_{i,t} = \lambda_i + \rho \cdot \Delta\text{SpE}_{i,t-1} + \sum_{h=-1}^{1} \beta_h \cdot \Delta G_{i,t-h} + \varepsilon_{i,t}, \quad t = 1960, ..., 1971$$

where λ_i represents a firm fixed effect, with sales per employee measured in \$ per employee, and government contracts quantified in millions of dollars. Estimation results are reported in Table 3.

[30] Refer to the last row of NIPA's Use table Before Redefinitions of 1965.

[31] Average sales-per-employee over the years 1960 through 1965 were 30,000\$ in manufacturing, 22,000\$ in non-manufacturing and 24,700\$ for all private firms covered by Compustat.

Table 3: Does Contractors Productivity Increase When Contracts Increase?

	Dependent: ΔSpE_t															
Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
$\Delta G_{i,t}$	-1.49			0.11	-1.34			0.65	-0.31			1.18	-0.61			1.16
	(2.04)			(2.01)	(1.56)			(1.40)	(1.35)			(1.55)	(1.05)			(1.17)
$\Delta G_{i,t+1}$		-1.96		-1.07		-1.67		-0.61		-0.59		1.18		-1.19		-0.02
		(1.73)		(1.91)		(1.24)		(1.30)		(1.33)		(1.55)		(1.00)		(1.11)
$\Delta G_{i,t-1}$			3.38*	4.44**			3.23**	4.61***			2.05	3.85**			2.55**	4.32***
			(2.01)	(2.15)			(1.51)	(1.52)			(1.42)	(1.67)			(1.08)	(1.27)
Aircraft Only	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
Reliance Weighted	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes
N (Contractors)	9	9	9	9	9	9	9	9	55	55	55	55	55	55	55	55
T (Years)	11	10	10	9	11	10	10	9	11	10	10	9	11	10	10	9
Observations	99	90	90	81	99	90	90	81	605	550	550	495	605	550	550	495

Notes: *** denotes 1% significance level. ** denotes 5% significance level. * denotes 10% significance level. All regressions include firm fixed effects. The first eight columns of the table show the results for firms operating in NAICS 3364, Aircraft Manufacturing and Parts. The last eight columns present results for all publicly traded defense contractors. The results shown in columns four to eight and thirteen to sixteen are weighted based on a firm's reliance on government contracts. The reliance of contractor i is $\text{Reliance}_i = \sum_t \text{Sales}_{i,t} / \sum_t \text{G}_{i,t}$. The median reliance is 20%, with a maximum of 100% of sales to the government. This weighting is made to acknowledge that companies more involved with government contracts might experience a greater impact from government purchases.

Notice that future and contemporaneous changes in contracts are not associated with changes in sales-per-employee. On the contrary, lagged changes in government contracts are strongly positively correlated with changes of sales-per-employee. Overall, the results are robust across different specifications: I check for aircraft manufacturers only and/or weigh observations by their average reliance of sales on government purchases (see Table 3 notes).

My estimates indicate that if in year t the change in government contracts increases by one million \$, in year $t+1$ the change in output-per-employee should increase by 4.32\$ per employee, on average (see column (16) of Table 3). To give a perspective on the size of this effect, Lockheed experienced an increase in government contracts between FY1964 and FY1968 of 415 million dollars. A fluctuation of this size would predict, on average, an increase in sales per employee by 1,792 dollars, which is equal to about 7.7% of Lockheed's average sales per employee during this period.[32]

Additionally, Figure 16 depicts the distribution of total defense procurement contracts across firms, showcasing the percentages allocated to the top 5, 25, and 100 companies over time. The top 50 firms account for about half of the total. Consequently, the 55 publicly traded top 100 defense contractors analyzed in Table 3 accounted for more than half of all defense procurement contracts. This demonstrates that the productivity gains observed were not confined within the smaller entities in the military procurement sector.

[32] Values of Lockheed contracts come from the Top 100 companies report; sales-per-employee data is from the Annual Fundamental segment of Compustat, for the fiscal years 1960 through 1970.

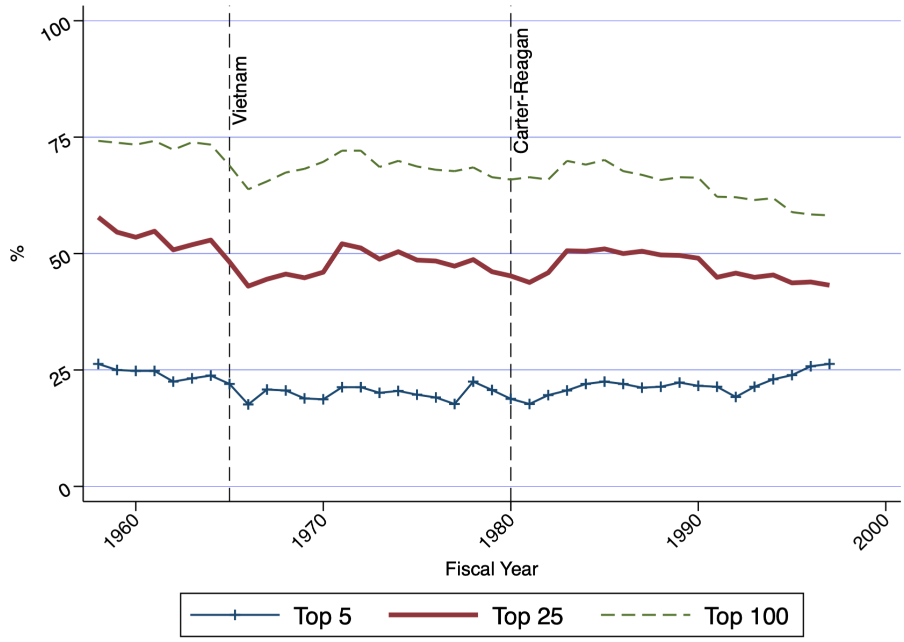


Figure 16: Granularity of Defense Procurement Contracts

Notes: Fiscal years goes from the 1st of July of year t until 30th of June of year $t + 1$, until FY1976. Afterwards, the definition of fiscal year changed: the FY starts on the 1st of October of year t and ends on the 30th of September of year $t + 1$.

The Phantom F-4 II Case and Learning-by-Doing: A potential micro-origin of the labor productivity boost experienced in war-times is learning-by-doing (McGrattan and Ohanian (2010)). A case in point is the production of the McDonnell-Douglas F-4 Phantom II, the defining aircraft of the Cold War period. Smith (1976) studied the production of this aircraft; his analysis spans 4665 airframes produced across 57 lots between 1958 and 1975. Delivery rates reached their peak during the Vietnam war with 71 airframes delivered monthly. This surge propelled McDonnell Douglas Aircraft to the premier position among the Top 100 companies.

Smith findings reveal that labor requirements decreased with both cumulative past production - a typical learning curve effect — but also with production rates, via reinforcement of labor routines. Specifically, a 1% uptick in production rates corresponded to roughly a 0.18% reduction in direct labor hours, according to his estimates (see Table 3 and Table 6 in his paper). Given that delivery rates increased by almost 300% from lot 18 to lot 30, Vietnam war build-up, labor requirement to produce the F-4 should have fallen by about 54%, and this is without accounting for the effect from cumulative past production.[33] In essence, as McDonnell Douglas ramped up production during the Vietnam war, labor productivity saw a significant rise due to learning effects.

Learning-by-Doing in Manufacturing and Military Production: It turns out that learning-by-doing is a widely observed phenomenon in the manufacturing and defense production (see Alchian (1963), Gulledge and Womer (1986), Argote and Epple (1990) and Benkard (2000)).[34] In fact, as noted by Arrow (1962), the concept of learning-by-doing originates in the context of (military) aircraft production (see Wright (1936) and Asher (1956)). Formally, it refers to an empirical regularity wherein unit costs decrease by a constant percentage as total production doubles. This decline in unit cost is attributed to the effects of learning.[35]

Learning-by-doing has been widely employed in the macroeconomic literature focused on long-run economic growth.[36] Nevertheless, learning can be rapid and its effects can have short-run implications too, as noted in Chang, Gomes, and Schorfheide (2002). For instance, the official BEA's government transaction methodology paper discusses rapidly falling prices in the context of military aircraft purchases due to learning effects (page II-66):

"*The learning curve may show steeply falling prices in the beginning years of production because of low initial labor productivity and the subsequent rapid price decline as productivity increases.*"

Learning can be fast for two reasons. First, the empirical regularity found in (military) production data, defined as "*learning curve*" is not a time-dependent concept: "*for every doubling of past production, labor requirements decrease by 20%*". Therefore, rapid learning can occur over short periods, provided production rates remain high. Evidence of a negative correlation between production rates and labor requirements has been found in many defense programs (see Smith (1976) and Bourgoine and Collins (1982) for literature review). Second, the stock of experience within a firm is subject to rapid depreciation, due to "*organizational*

[33] Notice that this is an extrapolation. Smith (1976) provides all data he used in his analysis, except for labor requirement and the estimated intercepts, which were considered proprietary and they have been masked.

[34] Evidence of learning in other manufacturing sectors: aircraft engines, machine tools, metal products, ship-building, semiconductors, refined petroleum products, power plants, chemical processing, and trucks.

[35] Specifically, they referenced: (i) job familiarization, (ii) better tool coordination, shop organization, and engineering coordination, (iii) development of efficient sub-assemblies, (iv) improved parts-supply systems, and (v) creation of more effective tools.

[36] Notable examples: Arrow (1962), Romer (1986), Lucas (1988) and Young (1991).

forgetting", see Argote and Epple (1990), Argote, Beckman, and Epple (1990), and Benkard (2000). These works have found evidence that when production halts or slows, there is a notable surge in labor hours needed per item, indicating the impact of forgetting.[37] According to Argote, Beckman, and Epple (1990), the stock of knowledge depreciates rapidly. Using data from the Liberty Ships program of WWII from 16 shipyards and 2708 ships, they estimate that only 3.2% of the initial yearly stock of knowledge persisted a year later. Less extreme depreciation rates are found in Benkard (2000) using data from the Lockheed's L-1011 commercial aircraft production, estimating a 61% yearly depreciation rate.

Finally, Ilzetzki (2023) provides evidence suggesting that military contractors during WWII experienced learning-by-doing, especially in situations where plants faced significant capacity constraints, a phenomenon referred to as "*learning-by-necessity*." In periods of military buildup, contractors encounter substantial increases in demand, placing considerable pressure on their production capacities. For instance, Figure 16 demonstrates a reduction in the share of defense procurement contracts awarded to the top 5, 25, and 100 defense contractors during both the Vietnam War and the Carter-Reagan military buildup. This pattern indirectly suggests that the leading contractors were unable to fulfill the government's entire demand during these times, necessitating that the Department of Defense seek military supplies from a broader array of firms. Should the top contractors have been operating near full capacity, the findings from Ilzetzki (2023) indicate that learning effects might have been particularly pronounced.

IV. Rationalization: Two Sector RBC with Learning

In this section I show that the empirical evidence brought forward by ordering defense contracts first in a VAR can be rationalized with a model where manufacturers feature learning-by-doing, a simple, yet empirically relevant, endogenous mechanism which rises productivity in response to extra demand from the government.

Consumption and Government Spending in Theory: The first paper to recognize that an endogenous increase in labor productivity could lead to a rise in consumption following a positive government spending shock was Devereux, Head, and Lapham (1996). In their model, productivity increases endogenously when demand rises due to increasing returns to specialization, as described by

[37] In this sense, it is remarkable the anecdote reported in Benkard (2000) at page 1049: "*In discussions with industry executives they have expressed the belief that disruptions in production, even those designed to improve efficiency, may lead to setbacks in productivity since they upset workers' routines.*"

Krugman (1979). As long as the price-cost markup exceeds 50%, an increase in government purchases leads to an uptick in consumption.

Subsequent modeling efforts to generate an increase in consumption after government spending shocks were motivated by the empirical evidence presented through the SVAR approach. A notable example is Galí, López-Salido, and Vallés (2007), which incorporates sticky prices and rule-of-thumb consumers. Both Monacelli and Perotti (2008) and Bilbiie (2011) employ sticky prices and non-separable preferences, with consumption and leisure acting as substitutes, to produce an increase in consumption.[38] Nonetheless, this class of models necessitates a countercyclical response of the markup to achieve a consumption increase, which contrasts with the evidence presented in Figure 12.

Recent studies have employed one-sector NK models with different endogenous mechanisms to boost labor productivity to achieve a positive consumption multiplier, as in Devereux, Head, and Lapham (1996). For instance, D'Alessandro, Fella, and Melosi (2019) leveraged the learning-by-doing mechanism introduced by Chang, Gomes, and Schorfheide (2002), while Jørgensen and Ravn (2022) focused on variable technology utilization.

Motivated by the abundant empirical evidence on learning-by-doing, I formulate a two-sector RBC model replicating the proportions of the manufacturing and non-manufacturing sectors. In this model, the manufacturing sector is characterized by learning-by-doing while the non-manufacturing sector is not. Unlike Chang, Gomes, and Schorfheide (2002) and D'Alessandro, Fella, and Melosi (2019), learning applies only to manufacturing, since government purchases are concentrated in manufacturing and evidence of learning is found almost exclusively in manufacturing and military production.

IV.a The Model

Households: Preferences are separable and households solve the following problem:

$$\max_{(N_t,C_{1,t},C_{2,t},u_{1,t},u_{2,t},K_{1,t},K_{2,t},I_{1,t},I_{2,t})} \sum_{t=0} \beta^t \cdot \left(\frac{\left(\tilde{C}_t - b \cdot \tilde{C}_{t-1}\right)^{1-\sigma}}{1-\sigma} - \psi \cdot \frac{N_t^{1+\varphi}}{1+\varphi} \right)$$

[38] In these models, following a positive government spending shock, hours worked rise for a given level of consumption. As a result, individuals tend to replace some of their leisure time with more consumption, prompting a leftward shift in labor supply.

subject to:

$$C_{1,t} + P_t \cdot C_{2,t} + I_{1,t} + P_t \cdot I_{2,t} = W_t \cdot N_t + \left(r^k_{1,t} \cdot u_{1,t}\ K_{1,t-1} + r^k_{2,t} \cdot u_{2,t}\ K_{2,t-1})\right) - T_t$$

$$K_{i,t} = (1 - a(u_{i,t})) \cdot K_{i,t-1} + I_{i,t} \cdot \left(1 - S(\frac{I_{i,t}}{I_{i,t-1}})\right) \quad i = 1, 2$$

with $\tilde{C}_t := C_{1,t}^{1-\phi} \cdot C_{2,t}^{\phi}$.

$C_{1,t}$ is produced by sector 1 which mimics the non-manufacturing sector of the economy. $C_{2,t}$ is the manufacturing good, produced by the manufacturing sector. The price of $C_{1,t}$ is the numeraire of the economy and P_t is the price of $C_{2,t}$ (relative to good 1). Since capital good of sector 2 has a different value of that one of sector 1, there are two rental rates of capital. Since capital goods is sector specific and cannot be shifted from one sector to another, households optimize either type of capital assets. T_t is a lump-sum tax.

Finally, households decide how much to invest in each period as well as how much to utilize capital. In particular, I have that:

- Capital utilization increases the depreciation rate of the capital stock:

$$a(u_{i,t}) = \delta + \delta_1 \cdot (u_{i,t} - 1) + \frac{\delta_2}{2} \cdot (u_{i,t} - 1)^2 \quad i = 1, 2$$

  with $\delta_1 = \frac{1-\beta}{\beta} + \delta$ to ensure that the steady state value of $u_{i,t}$ is 1 in both sectors.

- Investment adjustment costs are:

$$S(\frac{I_{i,t}}{I_{i,t-1}}) = \frac{\kappa}{2} \cdot \left(\frac{I_{i,t}}{I_{i,-1}} - 1\right)^2 \quad i = 1, 2$$

  Therefore, $S(1) = 0$ and $S'(1) = 0$. Moreover, if $\kappa = 0$, there are no adjustment costs.

Production: Production in the non-manufacturing sector occurs via a simple Cobb-Douglas technology with constant return to scale:

$$Y_{1,t} = N_{1,t}^{\alpha_1} \cdot \left(K^*_{1,t}\right)^{1-\alpha_1} \quad \text{with } K^*_{1,t} := u_{1,t} \cdot K_{1,t-1},$$

Firms maximize profits under perfect competition (take prices as given). The FOCs are:

$$[N_{1,t}]: \quad W_t = \alpha_1 \cdot \frac{Y_{1,t}}{N_{1,t}} := \mathrm{MPN}_{1,t}$$

$$[K^*_{1,t}]: \quad r^k_{1,t} = (1 - \alpha_1) \cdot \frac{Y_{1,t}}{K^*_{1,t}} := \frac{\mathrm{MPK}_{1,t}}{u_{1,t}}$$

The same production technology applies to the manufacturing sector:

$$Y_{2,t} = \left(E_t^{\theta} \cdot N_{2,t}\right)^{\alpha_2} \cdot \left(K_{2,t}^{*}\right)^{1-\alpha_2} \quad \text{with } K_{2,t}^{*} := u_{2,t} \cdot K_{2,t-1},$$

Here, E_t represents the stock of experience and θ is the learning parameter. The dynamics of experience is inspired by learning models with organizational forgetting (Argote and Epple (1990), Argote, Beckman, and Epple (1990) and Benkard (2000)). Organizational forgetting refers to the fact that the stock of experience depreciates over time due to (i) falling production rates, since typically such times are accompanied by layoffs or even (ii) normal rates, as employee turnover also lead to experience depreciation during periods of constant production.

Therefore, the dynamics of experience is equal to:

$$E_t = (1 - \delta_E) \cdot E + \underbrace{\delta_E \cdot E_{t-1}}_{\text{“Forgetting”}} + \underbrace{(Y_{2,t} - Y_2)}_{\text{“Learning”}},$$

where E is the steady state value of the stock of experience and Y_2 is the steady state value of production of manufacturing good.[39] On the contrary, Chang, Gomes, and Schorfheide (2002) uses past deviations of hours worked from steady state.

When $Y_{2,t}$ is above its steady-state level, experience accumulates, enhancing output-per-hour. This suggests that during economic booms, elevated production rates foster (i) greater reinforcement of routines among production workers and (ii) faster descent along the learning curve, thereby increasing labor productivity. Once production is back to its steady-state level, experience geometrically decays at rate δ_E. This reduction is due to organizational forgetting characterized by a slowdown in production rates, which leads to knowledge loss.

Conversely, when $Y_{2,t}$ is below the steady-state, the loss of experience triggers a decline in productivity. This can be likened to a recession period where turnover exceeds the normal rate, production is diminished, and experience is eroded due to layoffs and reduced reinforcement of routines, consequently exacerbating the recession (see Benkard (2000)).

Firms maximize profits under perfect competition (take prices as given). The FOCs are:

$$[N_{2,t}]: \quad W_t = P_t \cdot \alpha_2 \cdot \frac{Y_{2,t}}{N_{2,t}} := P_t \cdot \text{MPN}_{2,t}$$

$$[K_{2,t}^{*}]: \quad r_{2,t}^{k} = P_t \cdot (1 - \alpha_2) \cdot \frac{Y_{2,t}}{K_{2,t}^{*}} := P_t \cdot \frac{\text{MPK}_{2,t}}{u_{2,t}}$$

[39]For comparison, see Equation (2) in Argote and Epple (1990) and equation (6) in Benkard (2000).

Note that the firms do not internalize future productivity gains resulting from learning. This assumption aligns with Chang, Gomes, and Schorfheide (2002) and captures the firms' inability to foresee efficiency improvements.[40]

Markets Clearing, Aggregation and Fiscal Policy: Resources are sector specific:

$$Y_{i,t} = C_{i,t} + I_{i,t} + G_{i,t}, \quad i = 1, 2.$$

The sector specific capital accumulation equations are internalized in the household problem.[41]

The labor market clears ($N_t = N_{1,t} + N_{2,t}$), and real quantities are obtained using the price level at the beginning of the simulation, that is, the steady state value, P.[42]

Government spending is financed via lump-sum taxes. The government budget constraint is given by:

$$T_t = G_{1,t} + P_t \cdot G_{2,t}$$

Since I am interested in the effects of military purchases, which are primarily concentrated in manufacturing, I assume that government spending in non-manufacturing sector, sector 1, is constant: $G_{1,t} := G_1 = \gamma_1 \cdot Y_1$, where γ_1 is the fraction of output of sector 1 purchased by the government in steady state. On the contrary, government spending in the manufacturing sector, sector 2, is pinned down from the value of government spending in sector 1 and total (real) government spending, which is exogenous:[43]

$$G_t = (1 + \mathrm{IRF}_t^G) \cdot G$$
$$G_{2,t} = \frac{G_t - G_{1,t}}{P}$$

where P and G are the steady state values of the relative price and total government spending. IRF_t^G is an exogenous process for government spending estimated from the data: it is the estimated impulse response function of real government spending per-capita to a 1% structural shock to defense contracts.

[40] I attempted to incorporate this mechanism into the firm's problem, but it led to the product wage becoming a weighted average of current and future productivity values. Consequently, in times of expanding demand and rising productivity, the firms would incur losses, as the product wages paid today would exceed that of the present value of labor productivity.

[41] Notice that in this setup it is not possible to shift capital from one sector to another as in Ramey and Shapiro (1998). This is equivalent to a situation where the cost of shifting capital is high enough to make it always sub-optimal to shift capital from one sector to another. In a model like the one of Ramey and Shapiro (1998), a cost of shifting capital of 0.50, like the one suggested in their paper, would be enough to make shifting capital always sub-optimal. Therefore, this setup can be interpreted like one in which the cost of shifting capital is simply very high.

[42] This is consistent with what done in Ramey and Shapiro (1998) (see page 163 of their paper).

[43] Real here means measured at constant, i.e. steady state, price levels.

IV.b Model Simulation

Table 4 shows the calibrated values of all parameters in the model, with their source.

Table 4: Calibration Summary

Parameter	*Description*	*Value*	*Source*
φ	Inverse Frisch	1 or 0.20	Standard Calibration or Galí, López-Salido, and Vallés (2007)
β	Discount Factor	0.985	Quarterly Calibration
δ	Capital Depreciation	0.015	Quarterly Calibration
r	Net Interest Rate	$\frac{1}{\beta} - 1$	Follows from SS
κ	Inv. Adj. Cost	5.2	Ramey (2020)
δ_1	Capital Util. (Linear)	$r + \delta$	Standard to ensure $u = 1$
δ_2	Capital Util. (Quadratic)	$2 \cdot \delta_1$	Ramey (2020): $2 \cdot \delta_1$
σ	Inverse IES	1	Jaimovich and Rebelo (2009)
b	Consumption Habit	0.71	Burnside, Eichenbaum, and Fisher (2004)
α_1	Labor Income Share of GDP	0.63	Allows aggregate labor share to be 2/3
α_2	Labor Income Share of GDP	0.75	Ramey and Shapiro (1998)
γ_2	G_2/Y_2	16.2	(Gov. Purchases of MFG Commod.)/(MFG's VA) (from 1963 Use Table)
γ_1	G_1/Y_1	0.2134	Calibrated such that $G/Y = 0.20$
ϕ	Expenditure Share of MFG consumption	0.275	Match $C_{1963}^{MFG}/C_{1963}^{PrivateNon-MFG}$ (from 1963 Use Table)
ρ_E	Forgetting	0.75^3	Argote, Beckman, and Epple (1990)
θ	Learning	0.65	Benkard (2000)
ρ_A	Persistence of Contracts	0.84	Estimated from VAR's IRF
ψ	Weight of Labor	1	-

Organizational Forgetting: according to Benkard (2000) the range of estimates of monthly depreciation rates of the stock of knowledge is between 0.75 and 0.95. For instance Argote, Beckman, and Epple (1990) estimate a monthly depreciation rate of the stock of knowledge of 75% in the Liberty Ships program of WWII. Benkard (2000) finds higher estimates for the Lockheed TriStar program and finds a monthly deprecation rate of 95%. I set a value of $\delta_E = 0.75$, consistent with the estimates of Argote, Beckman, and Epple (1990).

Learning Rate: Parameter θ determines the speed of learning: the learning rate is calculated as $1 - 2^{-\theta}$ and represents how much labor requirement would fall if cumulative past output doubled. A negative relationship between labor requirements and experience is found by log-linearizing the production function, while leaving the capital stock and output unchanged:

$$\hat{N}_{2,t} = -\theta \cdot \hat{E}_t.$$

Simple learning models regressed the log of labor-requirement per unit of output on the log of experience and the log of production rates, where experience was proxied by either the stock of cumulative past output (e.g. Smith (1976)) or the stock of discounted cumulative past production (e.g. models of organizational forgetting like Argote and Epple (1990), Darr, Argote, and Epple (1995) and Benkard (2000)).

Table 5 summarizes different estimates for the learning parameters for different military products.

Table 5: Estimates of Learning Models

Dependent Variable: Labor Requirement		Experience (-θ)	Production Rate	Learning Rate $1-2^{-\theta}$
Smith (1976):	Aircraft (F-4)	-0.26	-0.17	16.5%
Benkard (2000):	Aircraft (without OF)	-0.35	-0.05	21.5%
	Aircraft (with OF)	-0.65	-0.86	36.3%
	Aircraft (with OF and Spillover)	-0.63	-0.89	35.4%
Gulledge and Womer (1986):				
	Aircraft A	-0.45	-0.04	26.9%
	Aircraft C	-0.37	-0.33	22.8%
	Aircraft D	-0.18	-0.56	11.8%
	Aircraft E	-0.14	-0.57	9.5%
	Aircraft F	-0.21	-0.80	13.4%
	Aircraft G	-0.25	-0.30	16.0%
	Aircraft H	-0.43	-0.13	25.6%
	Helicopter	-0.25	-0.16	16.2%
	Jet Engine A	-0.42	-0.12	25.0%
	Jet Engine B	-0.49	-0.16	28.6%
	Missile G&C	-0.12	-0.75	8.1%
	Ordnance Item	-0.18	-0.04	11.9%
	RadarSet A	-0.10	-0.17	6.9%
	Radar Set B	-0.02	-0.13	1.1%
Mean		-0.31	-0.35	18.5%
St.Dev		0.18	0.31	9.9%

Notes: Results from Gulledge and Womer (1986) are taken from Table 7.1 at page 124. They present estimates of two parameters: γ and δ, which map into the parameters of a regression of unit labor cost/requirement on cumulative past output, β_1, and production rates, β_2. The mapping between parameters is presented at page 120, after Equation 7.4: $\beta_1 = -\gamma \cdot \delta$ and $\beta_2 = \gamma - 1$. Here θ corresponds to their β_1 and estimates of the production rates corresponds to their β_2. Value of Smith (1976) are taken from Table 3 at page 66. Values from Benkard (2000) are taken from regressions (3) and (9); the production rates parameters correspond to his γ_0 estimates minus 1, like in Gulledge and Womer (1986). OF means "Organizational Forgetting" model.

The first column of Table 5 report estimates of θ: percent changes in labor requirement on percent changes in experience.

Models with organizational forgetting (OF) obtain higher estimates of θ: Benkard (2000) finds values around 0.64 (model with/without knowledge spillover), while Argote and Epple (1990) find a value of 0.65.[44]

Since Benkard (2000) estimates are obtained from a model which features organizational forgetting and the parameters of his model are estimated via GMM, using several instruments to rule out endogeneity problems like reverse causality (see Ilzetzki (2023)), I prefer to set the value of θ to match the learning rate estimated by Benkard (2000): $\theta = 0.65$.

Additionally, the second column of the Table shows the estimates of the coefficient in front of the log of production rates. The estimate is negative, suggesting increasing return to scale: higher production rates lead to lower labor requirements. Finally, the last column shows the implied estimates of the learning ratio, which are consistent with a 20% learning curve, on average. The bottom line of

[44] Argote and Epple (1990) estimate a production function, rather than a production frontier, and that's why their estimate are not reported in Table 5.

Table 5 is that evidence of learning-by-doing and increasing returns to scale has been documented in the production of several military programs.

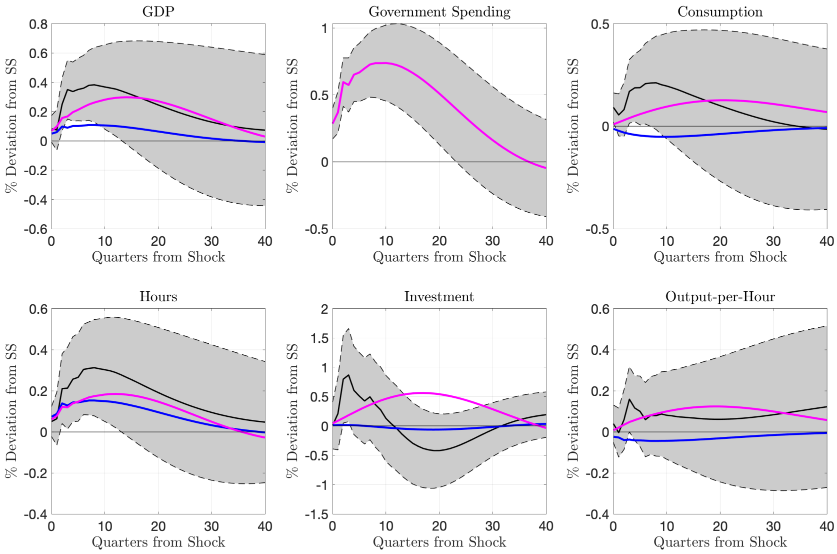
Results: I simulate the effect of an increase in government spending by feeding the estimated impulse response function of government spending to a 1% structural shock to defense contracts into the model.

The top panel of Figure 17 shows the results of a perfect foresight simulation for different values of θ. The blue solid line assumes no learning, $\theta = 0$; the magenta solid line assumes that manufacturing production is characterized by learning, with $\theta = 0.65$. Both cases assume a high Frisch elasticity of labor supply: $1/\varphi = 1/0.20$, consistent with that used in Galí, López-Salido, and Vallés (2007). Finally, the dark line presents the estimated impulse response function of several variables to a shock to defense contracts, along with the 90% confidence bands.

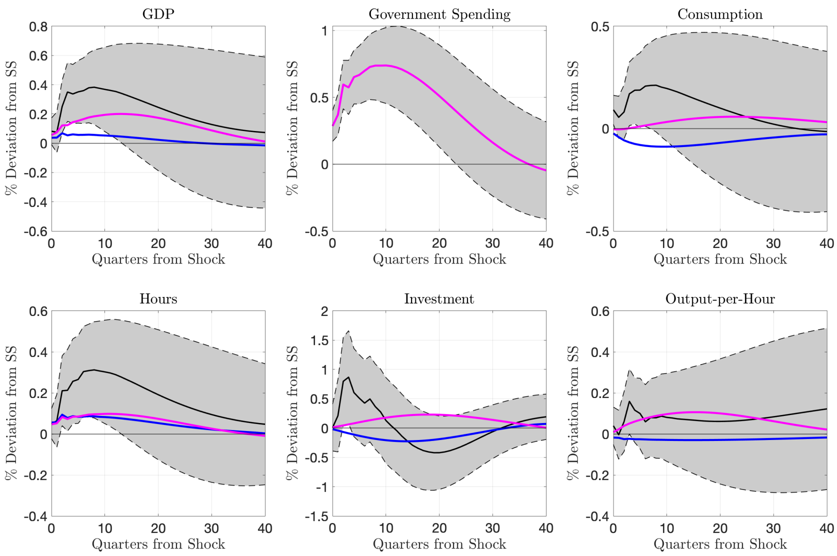
Since the model does not distinguish between durable and non-durable-plus-service consumption, I compare the model's and the empirical impulse response functions using the post-Korean War sample from 1954:Q1 to 2000:Q4. This approach helps to avoid the problem of underestimating the response of total consumption, due to the peculiar response of durable consumption during the Korean War, as discussed earlier.

The top-left panel illustrates the response of real GDP, which rises in all scenarios due to the increase in G, depicted in the top-middle panel. The bottom-left panel indicates that hours worked increase across all cases, a typical outcome stemming from the negative income effect of government spending: the necessity of higher lump-sum taxes to fund government purchases induces households to supply more labor. Even if I set a high value of the Frisch elasticity of labor supply, the response of hours in the model fall short of the observed empirical one. The bottom-middle figure displays the response of investment, which increases when learning is active, as households capitalize on the sustained productivity boost by augmenting the capital stock. In contrast, in the absence of learning, investment declines, mirroring the typical response in RBC models (refer to Baxter and King (1993) and Ramey and Shapiro (1998)). The bottom-right figure showcases the response of output-per-hour, or labor productivity, which decreases in the absence of learning due to increased hours worked and diminishing marginal returns. Conversely, output-per-hour rises when learning is operative.

Concluding, the top-right figure presents the response of consumption. In absence of learning, consumption declines as government purchases increase. This behavior is typical in standard RBC models with government spending, where consumption multipliers are negative without increasing returns to scale (see Proposition 2 in Bilbiie (2011)). In contrast, when the model incorporates learning,



(a) HIGH ELASTICITY (GALÍ, LÓPEZ-SALIDO, AND VALLÉS (2007): $\varphi = 0.20$)



(b) LOW ELASTICITY ($\varphi = 1$)

Figure 17: EMPIRICAL VS MODEL RESPONSES

Notes: Blue line: no learning ($\theta = 0$). Magenta line: learning ($\theta = 0.65$). Dark line: empirical IRF with 90% confidence bands (sample is 1954:1 to 2000:4). Real variables in the model are obtained by summing sectoral values at constant prices.

consumption increases by a magnitude similar to the empirical one. As long as production remains above the steady state, accumulated experience enhances labor productivity and real wages (as seen in the bottom-right panel), resulting in a net positive effect on consumption when the increase in labor earnings compensates for the adverse effects of higher taxes.

Lastly, the bottom panel of Figure 17 reports the results of a simulation when the Frisch elasticity is equal to one. The results are qualitatively identical to the high-elasticity case, however, the magnitude of the responses are now smaller due to the lower response of hours worked (bottom-left panel).

V. Conclusion

In this paper, I introduce a new quarterly time series, defense contracts, that measures the dollar value of all prime contract awards from the Department of Defense. I use it as an instrument for G to establish stylized facts of government spending.

Defense contracts overcome the major limitations of current methods to estimate the effects of government purchases. First, they accurately measure the timing of the shocks. In fact, since NIPA records military contracts into G with a delay, contracts lead G. This property of defense contracts represents a great advantage relative to the SVAR approach, which uses BP shocks. Second, defense contracts mainly capture fluctuations in military spending, which are driven by exogenous military events. Third, defense contracts retain statistical power as an instrument for G, even when data related to the Korean War is omitted from the sample. This addresses a major concern associated with the currently available instruments for measuring government spending, as noted by Perotti (2014) and Ramey (2016). Lastly, using data on defense contracts eliminates the necessity for narrative analysis. This methodology can be readily adopted for studies in various other countries that keep records of military contracts.

I find that a shock to defense contracts triggers a positive response in GDP, G, inventories, non-durable-plus-service consumption, hours worked, employment, production earnings, disposable income, the product wage, the price-cost markup, and labor productivity. Employing firm-level data, I demonstrate that lagged values of defense contracts correlate with increases in labor productivity among major defense contractors. Extensive evidence of productivity gains is found in both manufacturing and military production data, generally linked to learning effects. Consequently, I use a two-sector RBC model that simulates the proportions of the manufacturing and non-manufacturing sectors, with the former exhibiting learning-by-doing. In this model, a shock to defense contracts leads to an increase in government purchases from the manufacturing sector. This, in turn, boosts manufacturing production, enhancing labor productivity through learning-by-doing. As a result, the product wage increases, paving the way for a positive response in aggregate consumption, rationalizing the findings from the VAR analysis.

It should be noted that while learning-by-doing is a distinguishing feature of manufacturing, and particularly of defense production, it is not yet clear whether this transmission mechanism is characteristic of fiscal shocks exclusively or if it can influence other types of demand shocks. For instance, Christiano, Eichenbaum, and Evans (2005) observe a positive impact on labor productivity following an expansionary monetary policy shock. Learning-by-doing could potentially contribute to this increase in labor productivity, provided that the monetary expansion spurs manufacturing production, the sector most affected by learning. A potential propagation channel might be the automobile industry: lower interest rates encourage consumers to augment their current demand for vehicles, albeit to the detriment of future demand, thereby boosting present-day automobile consumption and production (see McKay and Wieland (2021)). While this exploration goes beyond the scope of the paper, it remains the subject of future research.

Bibliography

Acemoglu, **Daron**, **Akcigit**, **Ufuk**, and **Kerr**, **William** (2016). "Networks and the Macroeconomy: An Empirical Exploration". *NBER Macroeconomics Annual* 30.2015, p. 63. DOI: https://doi.org/10.1086/685961.

Alchian, **Armen** (Oct. 1963). "Reliability of Progress Curves in Airframe Production". *Econometrica* 31.4, p. 679. ISSN: 00129682. DOI: 10.2307/1909166.

Argote, **Linda**, **Beckman**, **Sara L**, and **Epple**, **Dennis** (Feb. 1990). "The Persistence and Transfer of Learning in Industrial Settings". *Management Science* 36.2, pp. 140–154.

Argote, **Linda** and **Epple**, **Dennis** (Feb. 1990). "Learning Curves in Manufacturing". *Science* 247.4945, pp. 920–924. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.247.4945.920.

Arrow, **Kenneth** (1962). "The Economic Implications of Learning by Doing". *The Review of Economic Studies* 29.3, pp. 155–173.

Asher, **Harold** (July 1956). *Cost-Quantity Relationships in the Airframe Industry*. RAND Corportation Report.

Auerbach, **Alan J** and **Gorodnichenko**, **Yuriy** (May 2012). "Measuring the Output Responses to Fiscal Policy". *American Economic Journal: Economic Policy* 4.2, pp. 1–27. ISSN: 1945-7731, 1945-774X. DOI: 10.1257/pol.4.2.1.

Auerbach, **Alan J**, **Gorodnichenko**, **Yuriy**, and **Murphy**, **Daniel** (Mar. 2020). "Local Fiscal Multipliers and Fiscal Spillovers in the USA". *IMF Economic Review* 68.1, pp. 195–229. ISSN: 2041-4161, 2041-417X. DOI: 10.1057/s41308-019-00102-3.

Barattieri, **Alessandro**, **Cacciatore**, **Matteo**, and **Traum**, **Nora** (Sept. 2023). "Estimating the Effects of Government Spending Through the Production Network". *NBER Working Paper* 31680. DOI: 10.3386/w31680.

Barnichon, **Regis**, **Debortoli**, **Davide**, and **Matthes**, **Christian** (Jan. 2022). "Understanding the Size of the Government Spending Multiplier: It's in the Sign". *The Review of Economic Studies* 89.1, pp. 87–117. ISSN: 0034-6527, 1467-937X. DOI: 10.1093/restud/rdab029.

Barro, **Robert** and **Redlick**, **Charles J.** (Feb. 2011). "Macroeconomic Effects From Government Purchases and Taxes *". *The Quarterly Journal of Economics* 126.1, pp. 51–102. ISSN: 0033-5533, 1531-4650. DOI: 10.1093/qje/qjq002.

Basu, **Susanto** and **Fernald**, **John** (1997). "Returns to Scale in U.S. Production: Estimates and Implications". *journal of political economy*.

Baxter, **Marianne** and **King**, **Robert G** (June 1993). "Fiscal Policy in General Equilibrium". *American Economic Review* 83.3, pp. 315–334.

Ben Zeev, **Nadav** and **Pappa**, **Evi** (Aug. 2017). "Chronicle of A War Foretold: The Macroeconomic Effects of Anticipated Defence Spending Shocks". *The*

Economic Journal 127.603, pp. 1568–1597. ISSN: 0013-0133, 1468-0297. DOI: 10.1111/ecoj.12349.

Benkard, **C Lanier** (2000). "Learning and Forgetting: The Dynamics of Aircraft Production". *American Economic Review* 90.4.

Bilbiie, **Florin O.** (Feb. 2011). "Nonseparable Preferences, Frisch Labor Supply, and the Consumption Multiplier of Government Spending: One Solution to a Fiscal Policy Puzzle". *Journal of Money, Credit and Banking* 43.1, pp. 221–251. ISSN: 00222879. DOI: 10.1111/j.1538-4616.2010.00372.x.

Bils, **Mark** and **Cho**, **Jang-Ok** (1994). "Cyclical Factor Utilization". *Journal of Monetary Economics* 33, pp. 319–354.

Blanchard, **Olivier** and **Perotti**, **Roberto** (Nov. 2002). "An empirical characterization of the dynamic effects of changes in government spending and taxes on output". *The Quarterly Journal of Economics* 117.4, pp. 1329–1368.

Bond, **Steve** et al. (July 2021). "Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data". *Journal of Monetary Economics* 121, pp. 1–14. ISSN: 03043932. DOI: 10.1016/j.jmoneco.2021.05.004.

Bouakez, **Hafedh**, **Rachedi**, **Omar**, and **Santoro**, **Emiliano** (Jan. 2023). "The Government Spending Multiplier in a Multisector Economy". *American Economic Journal: Macroeconomics* 15.1, pp. 209–239. ISSN: 1945-7707, 1945-7715. DOI: 10.1257/mac.20200213.

Bourgoine, **Philip** and **Collins**, **Kathy** (1982). "An Investigation of Changes in Direct Labor Requirements Resulting from Changes in A-10 Aircraft Production Rate." *Dissertation*.

Briganti, **Edoardo** and **Sellemi**, **Victor** (Mar. 2023). "Why Does GDP Move Before Government Spending? It's all in the Measurement". *UCSD Manuscript*.

Broer, **Tobias** and **Krusell**, **Per** (Apr. 2021). "Fiscal Multipliers: A Heterogenous-Agent Perspective". *Working Paper*.

Brunet, **Gillian** (May 2022). "When Does Government Spending Matter? Evidence from a New Measure of Spending". *Working Paper*, p. 41.

Burnside, **Craig**, **Eichenbaum**, **Martin**, and **Fisher**, **Jonas** (Mar. 2004). "Fiscal shocks and their consequences". *Journal of Economic Theory* 115.1, pp. 89–117. ISSN: 00220531. DOI: 10.1016/S0022-0531(03)00252-7.

Chang, **Yongsung**, **Gomes**, **Joao F**, and **Schorfheide**, **Frank** (2002). "Learning-by-Doing as a Propagation Mechanism". *THE AMERICAN ECONOMIC REVIEW* 92.5.

Christiano, **Lawrence**, **Eichenbaum**, **Martin**, and **Evans**, **Charles** (Feb. 2005). "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy". *Journal of Political Economy* 13.1, pp. 1–45.

Christiansen, **Lone Engbo** and **Goudie**, **Bryan** (2007). "Essays on Productivity, Technology, and Economic Fluctuations". *UC San Diego, Dissertation*.

Coibion, **Olivier**, **Gorodnichenko**, **Yuriy**, and **Weber**, **Michael** (Dec. 2020). "Fiscal Policy and Households' inflation Expectations: Evidence from a Randomized Control Trial." *Woriking Paper*.

Cox, **Lydia** et al. (Sept. 2022). "BIg G". *Working Paper*.

D'Alessandro, **Antonello**, **Fella**, **Giulio**, and **Melosi**, **Leonardo** (Aug. 2019). "Fiscal Stimulus with Learning-by-Doing". *International Economic Review* 60.3, pp. 1413–1432. ISSN: 0020-6598, 1468-2354. DOI: 10.1111/iere.12391.

Darr, **Eric D.**, **Argote**, **Linda**, and **Epple**, **Dennis** (Nov. 1995). "The Acquisition, Transfer, and Depreciation of Knowledge in Service Organizations: Productivity in Franchises". *Management Science* 41.11, pp. 1750–1762. ISSN: 0025-1909, 1526-5501. DOI: 10.1287/mnsc.41.11.1750.

Devereux, **Michael B.**, **Head**, **Allen C.**, and **Lapham**, **Beverly J.** (May 1996). "Monopolistic Competition, Increasing Returns, and the Effects of Government Spending". *Journal of Money, Credit and Banking* 28.2, p. 233. ISSN: 00222879. DOI: 10.2307/2078025.

Dupor, **Bill** and **Guerrero**, **Rodrigo** (Dec. 2017). "Local and aggregate fiscal policy multipliers". *Journal of Monetary Economics* 92, pp. 16–30. ISSN: 03043932. DOI: 10.1016/j.jmoneco.2017.07.007.

Eichenbaum, **Martin** and **Fisher**, **Jonas** (2005). "Fiscal Policy in the Aftermath of 9/11". *Journal of Money, Credit, and Banking* 37.1, pp. 1–22. ISSN: 1538-4616. DOI: 10.1353/mcb.2005.0005.

Fatas, **Antonio** and **Mihov**, **Ilian** (2001). "The Effects of Fiscal Policy on Consumption and Employment: Theory and Evidence". Available at SSRN 267281 (2001), p. 39.

Fernald, **John** (Apr. 2012). "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity". *Federal Reserve Bank of San Francisco, Working Paper Series*, pp. 01–28. DOI: 10.24148/wp2012-19.

Ferraz, **Claudio**, **Finan**, **Frederico**, and **Szerman**, **Dimitri** (May 2015). "Procuring Firm Growth: The Effects of Government Purchases on Firm Dynamics". *NBER Working Paper*. DOI: 10.3386/w21219.

Fisher, **Jonas** and **Peters**, **Ryan** (May 2010). "Using Stock Returns to Identify Government Spending Shocks". *The Economic Journal* 120.544, pp. 414–436. ISSN: 0013-0133, 1468-0297. DOI: 10.1111/j.1468-0297.2010.02355.x.

Gabriel, **Ricardo Duque** (2022). "The Credit Channel of Public Procurement". *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.4270202.

Galí, **Jordi**, **López-Salido**, **J David**, and **Vallés**, **Javier** (Mar. 2007). "Understanding the Effects of Government Spending on Consumption". *Journal of the European Economic Association* 5.1, pp. 227–270.

Galí, **Jordi** and **Monacelli**, **Tommaso** (Sept. 2008). "Optimal monetary and fiscal policy in a currency union". *Journal of International Economics* 76.1, pp. 116–132. ISSN: 00221996. DOI: 10.1016/j.jinteco.2008.02.007.

Giavazzi, **Francesco** and **McMahon**, **Michael** (2012). "The Household Effects of Government Spending". *Fiscal policy after the financial crisis*. University of Chicago Press, pp. 103–141.

Ginsburg, **David** (Jan. 1952). "Price Stabilization, 1950-1952: Retrospect and Prospect". *University of Pennsylvania Law Review* 100.4, p. 514. ISSN: 00419907. DOI: 10.2307/3310172.

Gordon, **Robert** and **Krenn**, **Robert** (Sept. 2010). "The End of the Great Depression 1939-41: Policy Contributions and Fiscal Multipliers", w16380. DOI: 10.3386/w16380.

Gugler, **Klaus**, **Weichselbaumer**, **Michael**, and **Zulehner**, **Christine** (Feb. 2020). "Employment behavior and the economic crisis: Evidence from winners and runners-up in procurement auctions". *Journal of Public Economics* 182, p. 104112. ISSN: 00472727. DOI: 10.1016/j.jpubeco.2019.104112.

Gulledge, **Thomas R.** and **Womer**, **Norman K.** (1986). *The Economics of Made-to-Order Production: Theory with Applications Related to the Airframe Industry*. Ed. by **M. Beckmann** and **W. Krelle**. Vol. 261. Lecture Notes in Economics and Mathematical Systems. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-540-16055-7 978-3-642-46566-6. DOI: 10.1007/978-3-642-46566-6.

Hall, **Robert** (Nov. 2009). "By How Much Does GDP Rise if the Government Buys More Output?" *NBER Working Paper* Working Paper 15496.

Hamilton, **James D.** (Dec. 2018). "Why You Should Never Use the Hodrick-Prescott Filter". *The Review of Economics and Statistics* 100.5, pp. 831–843. ISSN: 0034-6535, 1530-9142. DOI: 10.1162/rest_a_00706.

Hickman, **Bert G.** (Jan. 1955). *The Korean War and United States Economic Activity, 1950-1952*. hick55-1. National Bureau of Economic Research.

Ilzetzki, **Ethan** (Jan. 2023). "Learning by Necessity: Government Demand, Capacity Constraints, and Productivity Growth". *Working Paper*.

Jaimovich, **Nir** and **Rebelo**, **Sergio** (Aug. 2009). "Can News about the Future Drive the Business Cycle?" *American Economic Review* 99.4, pp. 1097–1118. ISSN: 0002-8282. DOI: 10.1257/aer.99.4.1097.

Jørgensen, **Peter L.** and **Ravn**, **Søren H.** (Jan. 2022). "The inflation response to government spending shocks: A fiscal price puzzle?" *European Economic Review* 141, p. 103982. ISSN: 00142921. DOI: 10.1016/j.euroecorev.2021.103982.

Krugman, **Paul** (Nov. 1979). "Increasing Returns, Monopolistic Competion, and International Trade". *Journal of International Economics* 9.4, pp. 469–479.

Lee, **Munseob** (2017). “Government Purchases, Firm Growth and Industry Dynamics.pdf”. *Working Paper*.

Leeper, **Eric**, **Traum**, **Nora**, and **Walker**, **Todd** (Aug. 2017). “Clearing Up the Fiscal Multiplier Morass”. *American Economic Review* 107.8, pp. 2409–2454. ISSN: 0002-8282. DOI: 10.1257/aer.20111196.

Lucas, **Robert E** (1988). “On the Mechanics of Economic Development”. *Journal of Monetary Economics* 22.1, pp. 3–42.

McGrattan, **Ellen R.** and **Ohanian**, **Lee** (May 2010). “DOES NEOCLASSICAL THEORY ACCOUNT FOR THE EFFECTS OF BIG FISCAL SHOCKS? EVIDENCE FROM WORLD WAR II*: effects of fiscal shocks in wwii”. *International Economic Review* 51.2, pp. 509–532. ISSN: 00206598, 14682354. DOI: 10.1111/j.1468-2354.2010.00591.x.

McKay, **Alisdair** and **Wieland**, **Johannes F.** (2021). “Lumpy Durable Consumption Demand and the Limited Ammunition of Monetary Policy”. *Econometrica* 89.6, pp. 2717–2749. ISSN: 0012-9682. DOI: 10.3982/ECTA18821.

Monacelli, **Tommaso** and **Perotti**, **Roberto** (Dec. 2008). *Fiscal Policy, Wealth Effects, and Markups*. Tech. rep. w14584. Cambridge, MA: National Bureau of Economic Research, w14584. DOI: 10.3386/w14584.

Mountford, **Andrew** and **Uhlig**, **Harald** (Sept. 2009). “What are the effects of fiscal policy shocks?” *Journal of Applied Econometrics* 24.6, pp. 960–992. ISSN: 08837252. DOI: 10.1002/jae.1079.

Nakamura, **Emi** and **Steinsson**, **Jón** (Mar. 2014). “Fiscal Stimulus in a Monetary Union: Evidence from US Regions”. *American Economic Review* 104.3, pp. 753–792. ISSN: 0002-8282. DOI: 10.1257/aer.104.3.753.

Nekarda, **Christopher** and **Ramey**, **Valerie** (Jan. 2011). “Industry Evidence on the Effects of Government Spending”. *American Economic Journal: Macroeconomics* 3.1, pp. 36–59. ISSN: 1945-7707, 1945-7715. DOI: 10.1257/mac.3.1.36.

— (Dec. 2020). “The Cyclical Behavior of the Price-Cost Markup”. *Journal of Money, Credit and Banking* 52.S2, pp. 319–353. ISSN: 0022-2879, 1538-4616. DOI: 10.1111/jmcb.12755.

Perotti, **Roberto** (Jan. 2007). “In Search of the Transmission Mechanism of Fiscal Policy [with Comments and Discussion]”. *NBER Macroeconomics Annual* 22, pp. 169–249. ISSN: 0889-3365, 1537-2642. DOI: 10.1086/ma.22.25554966.

— (May 2014). “Defense Government Spending Is Contractionary, Civilian Government Spending Is Expansionary”. *NBER Working Paper*, w20179. DOI: 10.3386/w20179.

Pinardon-Touati, **Noemie** (Feb. 2022). “The Crowding Out Effect of Local Government Debt: Micro- and Macro-Estimates”.

Ramey, **Valerie** (June 1989). “Inventories as Factors of Production and Economic Fluctuations”. *American Economic Review* 79.3, pp. 338–354.

Ramey, **Valerie** (Apr. 1991). "Nonconvex Costs and the Behavior of Inventories". *Journal of Political Economy* 99.2, pp. 306–334. ISSN: 0022-3808, 1537-534X. DOI: 10.1086/261752.

— (Feb. 2011). "Identifying Government Spending Shocks: It's all in the Timing*". *The Quarterly Journal of Economics* 126.1, pp. 1–50. ISSN: 0033-5533, 1531-4650. DOI: 10.1093/qje/qjq008.

— (2016). "Macroeconomic Shocks and Their Propagation". *Handbook of macroeconomics*. Vol. 2. Elsevier, pp. 71–162.

— (July 2020). *The Macroeconomic Consequences of Infrastructure Investment*. Tech. rep. w27625. Cambridge, MA: National Bureau of Economic Research, w27625. DOI: 10.3386/w27625.

Ramey, **Valerie** and **Shapiro**, **Matthew** (1998). "Costly capital reallocation and the effects of government spending". *Carnegie-Rochester Conference on Public Policy* 48.1998, pp. 145–194.

Ramey, **Valerie** and **Zubairy**, **Sarah** (Mar. 2018). "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data". *journal of political economy*, p. 52.

Romer, **Paul M.** (Oct. 1986). "Increasing Returns and Long-Run Growth". *Journal of Political Economy* 94.5, pp. 1002–1037. ISSN: 0022-3808, 1537-534X. DOI: 10.1086/261420.

Smith, **Larry** (1976). "An Investigation og Changes in Direct Labor Requirements resulting from Changes in Airframe Production Rates". *Dissertation*.

Wright, **T. P.** (Feb. 1936). "Factors Affecting the Cost of Airplanes". *Journal of the Aeronautical Sciences* 3.4, pp. 122–128. ISSN: 1936-9956. DOI: 10.2514/8.155.

Young, **Alwyn** (1991). "Learning by Doing and the Dynamic Effects of International Trade". *The Quarterly Journal of Economics* 106.2, pp. 369–405.