When Does Government Spending Matter? It's All in the Measurement

Edoardo Briganti Gillian Brunet * Victor Sellemi

July 3, 2025

Abstract

Since government spending is a long, complex process and not an event that occurs at a single point in time, its timing depends on its measurement. In this paper, we argue that national income accounting (NIPA) measures government spending too late in the process to fully capture its economic effects, resulting in a systematic downward bias in aggregate time series estimates of the fiscal multiplier. This bias occurs because new government purchases initially show up in GDP as inventory investment rather than government purchases. We combine budget and contract data to produce an alternate measure of government spending based on authorizations. This new measure anticipates NIPA G by 3-4 quarters and allows for cleaner, more precise identification of fiscal shocks. We show that our new measure produces time series multiplier estimates of approximately 1 or higher at all time horizons using both linear projection (LP-IV) and structural VAR methods.

(JEL E60, E62)

---

*Contacts, acknowledgments and disclaimer.

Briganti: Bank of Canada (email:`ebriganti@bankofcanada.ca`). Brunet: Smith College (`gbrunet@smith.edu`). Sellemi: University of California San Diego (email: `vsellemi@ucsd.edu`). *Disclaimer*: The views expressed here are those of the authors and do not necessarily reflect the views of the Bank of Canada or other members of its staff. We are particularly thankful to our advisors, Valerie Ramey and Christina Romer, for their help and support. We also acknowledge helpful comments from Karl Boulware, J. Bradford DeLong, Barry Eichengreen, Yuriy Gorodnichenko, James Hamilton, Juan Herreño, Nir Jaimovich, Marc Muendler, Emi Nakamura, Martha Olney, Sarah Quincy, David Romer, Allan Timmermann, Fabian Trottner, Johannes Wieland and anonymous referees. We are further indebted to Richard Kogan and Paul Van de Water for sharing invaluable expertise on the technical details of the federal budget process. We would also like to thank seminar participants at UC Berkeley, UC San Diego, the Federal Reserve Board of Governors, the Bank of Canada, the Banco de España, the University of Colorado Boulder, ISER Osaka University, the UC Davis Macrohistory Workshop, Smith College, the University of Naples Federico II, the Frankfurt School of Finance and the Kiel Institute for the World Economy. All errors are our own.

"*We shall not, however, succeed in spending the £1,250 million this year, and some of the late Government's programme must necessarily roll forward into a future year. [...] I have never yet seen a munitions programme—and I have seen several—which did not lag behind the plans.*"

— Winston Churchill (December 6, 1951)[1]

I. Introduction

Measurement is fundamental to macroeconomic analysis, often more fundamental than we realize. This paper examines how the measurement of government spending influences estimates of the fiscal multiplier. The modern literature on aggregate fiscal multipliers has emphasized the role of anticipation effects, or the idea that responses to fiscal shocks may come before the shocks themselves are observed in the data (e.g. Ramey and Shapiro, 1998; Mertens and Ravn, 2010; Ramey, 2011). Implicitly or explicitly, these papers assume a Ricardian channel: anticipation effects occur because firms and households are forward-looking and act upon news before government spending occurs. In this paper, we propose a new explanation for anticipation effects in the government spending multiplier, grounded in measurement concepts and nuances of national accounting. This new explanation has far-reaching implications for both the estimated size of the fiscal multiplier and also the theoretical mechanisms driving the macroeconomic response to government spending shocks.

The time series literature often uses the National Income Product Accounts (NIPA) to measure government spending. NIPA measures government spending as the flow of government funds out of the Treasury for spending purposes. For some applications—such as national income accounting—this measure is ideal. However, if the goal is to measure the influence of government spending on economic activity, we argue that NIPA measures government spending too late in the process. A significant share of government payments are delayed until firms deliver final goods to the government. Firms are often paid after they have received a contract, hired workers, and purchased materials. These actions initially show up in NIPA as inventory investment by firms with gov-

[1]We thank Edward Nelson and Valerie Ramey for bringing this quote to our attention.

ernment contracts. When the firm delivers the finished goods, NIPA records negative inventory investment and positive government spending. As a result, NIPA may record government spending after its direct effects on the economy have already begun, and sometimes after the direct effects have concluded. Because the lag between when a firm engages in economic activity in response to a government contract and when it is paid by the government can be long and variable, correcting this timing problem requires a different measure of government spending.

We argue that the significant (and variable) time lags between the point in time at which government spending begins affecting economic activity and the point in time at which government spending is measured provide another explanation for anticipation effects in the fiscal multiplier literature. Importantly, our measure neither assumes nor requires a Ricardian mechanism, instead attributing anticipation effects to the mechanics of national income accounting. That said, our approach does not preclude the Ricardian channel posited by the previous literature, either.

Following the aggregate multiplier literature (including Hall, 2009; Barro and Redlick, 2011; Ramey and Zubairy, 2018, among others), we focus on defense spending because changes to defense spending are plausibly exogenous to economic conditions over the short run, unlike non-defense spending. We then compile and combine two separate measures of anticipated defense spending to produce our preferred measure, which we call spending authorizations. We combine two separate measures because each has strengths and weaknesses; using either measure independently produces similar estimates (see Online Appendix G).

The first measure we rely on is called budget authority, which directly measures authorizations of government spending through the Congressional budget process. The advantage of budget authority is that it includes all defense spending, including direct spending by the government (e.g., paying the salaries of military personnel). The drawback of budget authority is that it is inherently annual and thus lower frequency than many macroeconomists would prefer to use. The second measure of spending is military contracts. Military contracts are a narrower measure of defense spending than budget authority, but are available at quarterly frequency. We use the quarterly variation in military contracts to interpolate quarterly variation within the annual aggregates established by budget authority. We call the resulting measure *spending authorizations*.

A major advantage of our measure is its flexibility: it can easily be combined with a wide range of existing approaches to fiscal multiplier measurement, including both linear projections and recursive SVARs (i.e. Cholesky identification). We illustrate how spending authorizations can be used with both of these approaches.

We construct a simple econometric model relating authorizations and NIPA, modeling NIPA G as a delayed moving average of spending authorizations. Our model shows that multipliers estimated using NIPA G will have a systematic downward bias. Our empirical results confirm the model's predictions. The bias in multiplier estimates is particularly acute at shorter time horizons, exactly as predicted by our model. We estimate fiscal multipliers above one at short time horizons (under two years). From two to four years after the shock, we estimate a multiplier of approximately one using linear projections. VAR approaches consistently produce somewhat larger multiplier estimates at all time horizons. These results are robust to excluding the outbreak of the Korean War. Including WWII in the sample reduces our multiplier estimates at all time horizons (below one when using linear projections); this is consistent with prior literature examining the short-run macroeconomic impact of WWII (Fishback and Cullen, 2013; Brunet, 2024).

In addition to correcting for systematic bias in multiplier estimation due to the timing mismatch, our results also reveal that multiplier estimates are extremely sensitive to how we define the fiscal multiplier. The fiscal multiplier per dollar spent (i.e. per dollar of NIPA) is often very different from the fiscal multiplier per dollar authorized. To relate these quantities, we introduce a new decomposition of the fiscal multiplier, separating the multiplier on private spending ($GDP - G$) from the response of government spending to spending authorizations. This decomposition allows us to produce a "heuristic multiplier," corresponding to how people typically think about the fiscal multiplier. The heuristic multiplier counts every dollar of authorized spending as spent and then adds the multiplier on private spending, corresponding to how people typically conceptualize the multiplier. When estimated using our baseline sample, the heuristic multiplier is around 1.2 upon impact and for the first year after impact, after which it gradually declines to the vicinity of one.

In addition to the new multiplier estimates and new distinctions between multiplier definitions, we also establish new stylized facts about the response of certain GDP components to shocks to government spending. Specifically, we document the inventories channel, which is the central mechanism for the timing mismatch between GDP and NIPA G. We show that government purchases via private contractors affect inventories before they affect NIPA G. Furthermore, the scale of inventory response is large enough to explain our large short-run multipliers. On the basis of this evidence, we argue that firms' response to government spending shocks—increasing productions and accumulating inventories—is the primary channel through which government spending shocks affect short-run economic activity. Our results do not disprove the traditional Ricardian channel in which government spending primarily influences short-run economic activity through household expectations, but they do suggest that the inventory channel—driven by firms rather than households—dominates.

The paper is organized as follows: Section 2 discusses the government spending process and alternative measures of government spending and then 3 presents a stylized model showing how different measures relate. Section 4 describes the data and explains how we construct our new measure of government spending, spending authorizations, using historical data on budget authority and military contracts. Section 5 examines the aggregate effects of spending authorizations in the economy, it provides estimates of the fiscal multiplier, and highlights the important role of inventories in the context of anticipation effects of government spending. Section 6 concludes.

II. Government Spending as a Process

Quantitative analysis requires economists to make simplifying assumptions about the process of government spending. However, we must remember the assumptions used to construct our data, no matter how ubiquitous.[2] The commitment to spend, the allocation of contracts, and the transfer of funds for government purchases are collapsed into a single moment—and a single measure, as each

[2]Other economists have carefully considered these issues in related contexts. One example is Leduc and Wilson (2013), who distinguish between the allocation of funding—which follows closely after commitment to spend—and the actual receipt of funds.

dollar spent must appear precisely once as government spending in the NIPA (the most standard measure of government purchases in the literature). In practice, it can take months or years from the moment the government commits to spending to the moment when funds are transferred.

While conceptually simplifying the process of government spending is helpful in many settings, it is a hindrance when we are concerned with precisely identifying the timing and process by which spending affects economic activity, and when our goal is to identify and estimate the fiscal multiplier. Recall that the multiplier is defined as the ratio of a change in output to a change in government spending. Intuitively, timing simplifications in NIPA G mean that changes in government spending are measured as changes in final payments, which can occur several months or years after the government initially authorized the spending.

II.1. Measures of Spending and their Correspondence to the Budget Process

Below we describe three measures of government spending and discuss the timing of each in the context of the federal budget process. The first two measures, budget authority and contract awards, correspond to events early in the budget process. The NIPA measure of government spending corresponds to events much later in the budget process.

Budget Authority. Before the federal government actually spends money, Congress and the executive branch partake in a formal budgeting process. As part of this process, Congress establishes *budget authority* for each account in the federal budget. Specifically, budget authority is the legal authority for a government entity (e.g., the Department of Defense) to make contracts to spend money on a specific program or purpose.

Why can budget authority be interpreted as a measure of anticipated spending? Before the government can increase spending, it must authorize spending increases. Budget authority directly measures those authorizations.

Over sufficiently long time periods, cumulative budget authority converges with cumulative spending, since the commitment to spend leads and predicts the eventual transfer of funds. Budget authority tends to lead budget outlays (the disbursement of funds from the Treasury) by about a year on average, and shocks to budget authority are generally sharper than shocks to outlays. Over time, however, \$1 of budget authority translates fairly directly into \$1 of spending in NIPA.[3]

Establishing budget authority is the first step in the government spending process. While budget authority may itself be anticipated by observers following budget negotiations, budget authority reflects the point in the process when uncertainty about spending authorizations for the coming year is resolved.

Contracts. Once budget authority has been established (typically through the passage of appropriations bills), the relevant agencies have the legal authority to "obligate funds" (i.e., to commit to spending). These commitments can take the form of both direct spending (e.g., paying salaries for the Armed Forces or production in facilities both owned and operated by the federal government) and indirect spending (e.g., contracts with private firms). Contracts with private firms are much more common in the defense sector than in other areas of federal spending: since 1980, defense spending has always accounted for less than 30 percent of the federal budget; yet on average, 80 percent of federal contracts since 1980 have been military contracts.

Legally, budget authority may be obligated (i.e., new contracts may be made) for up to two years after its authorization. In practice, changes in contracts tend to be contemporaneous with changes in budget authority (i.e., within the same fiscal year). The top panel of Figure 1 shows real per capita Budget Authority and military contracts by fiscal year since 1947. The bottom panel of Figure 1 shows that the two series move synchronously, with the lead-lag correlation between budget authority and contracts peaking at .914, at lag 0. This indicates that funds are awarded as contracts by federal agencies largely within a year of being authorized by Congress, which allows us to combine higher frequency contract data with lower-frequency movements in Budget Authority in section 4.3.

While it is possible that firms may anticipate individual contract awards, we show in Appendix

[3]There are exceptions, primarily involving things like funding for revolving loan accounts. However, these exceptions are almost exclusively concentrated in non-defense spending.

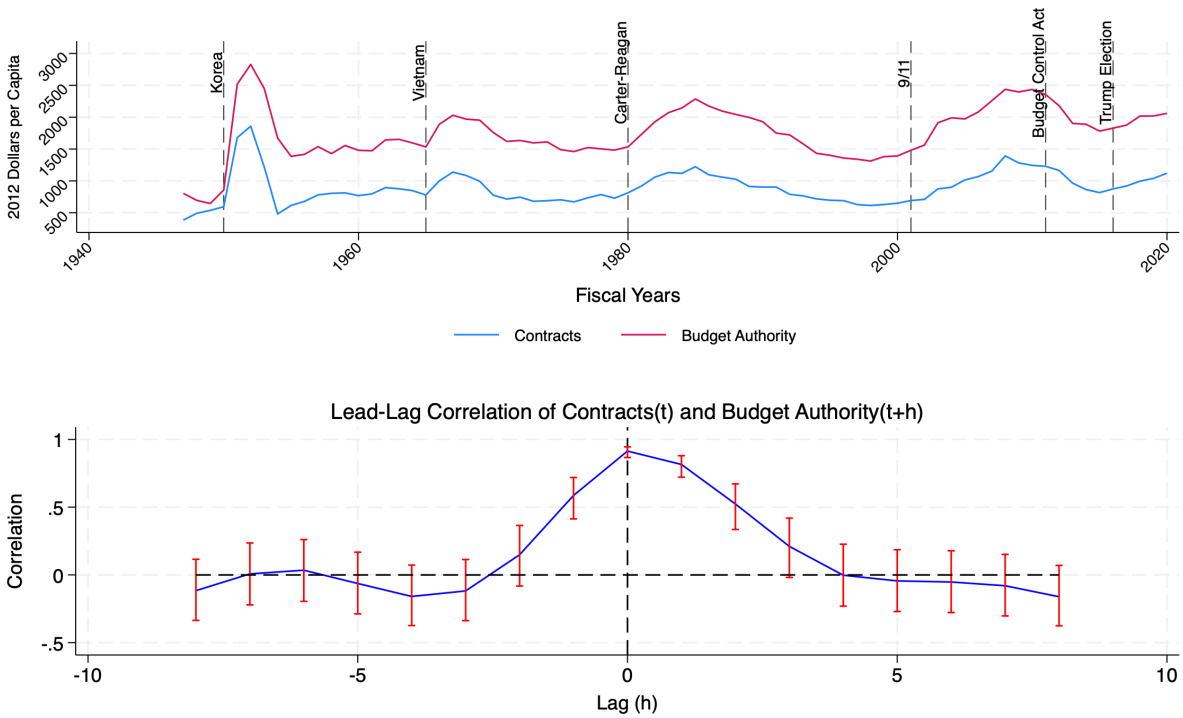


Figure 1: Budget Authority and Military Contracts Move Synchronously

Notes: We aggregate contract data to fiscal year frequency by averaging quarterly annualized variables. The fiscal year starts on July 1st before 1977 and on October 1st starting in 1977. We use the GDP price deflator to adjust for inflation, with 2012 as the base year. We measure lead-lag correlation as: $\text{Corr}(\text{Contracts}_t, \text{BA}_{t+h})$, where variables are in real per capita values. We omit WWII from the figure because the scale of the shock is so large that it becomes difficult to see variation in other time periods. In Appendix B.1, we show that the lead-lag correlation is robust to the inclusion of WWII, with a contemporaneous correlation value of .909 in the extended sample. In Appendix B.2, we also show that the correlation between budget authority and contracts peaks at zero if we include WWII in the series of contracts and Budget Authority, or calculate lead-lag correlations using first differences.

E that there is limited evidence for this in the data. Rather, the bulk of direct economic activity relating to a contract—hiring workers, purchasing raw materials, subcontracting to other firms for parts, and the actual production process—occurs once the contract is in place.

NIPA "G". The most common measure of government purchases used by economists is from the National Income and Product Accounts (NIPA), produced by the Bureau of Economic Analysis (BEA). NIPA uses several methods to construct defense spending.[4] For some types of military purchases (e.g. aircraft, missiles, vehicles, and petroleum products), the timing of NIPA reflects the

[4]For a detailed discussion of the BEA's methods and how they are applied, see Appendix A.8 and BEA (*Government Transactions* (2005)), pp. II: 33-45.

delivery of goods to the military. Goods cannot be delivered until after they have been produced, which introduces long and variable time lags depending on the length of the production process for each individual good. In the time between production and delivery, these government purchases are recorded in NIPA as an inventory investment by supplying firms. When products are delivered, NIPA shifts their value from inventories (i.e. reduction in firms' inventory investment) to government purchases (i.e. increase in delivered goods). As BEA itself explains:

> "The largest timing difference is for national defense gross investment for relatively long-term production items, such as aircraft and missiles, for which the work in progress is considered as part of business inventories until the item is completed and delivered to the Government." (BEA *Government Transaction Methodology Paper*, p. II-11)

For other military purchases, the timing of NIPA reflects the disbursement of funds from the Treasury (i.e. when contractors are paid). The timing of these purchases matches up exactly with budget outlays, which are measured in the same way. While the details depend on the contract, final payments are typically not made until after goods are delivered to the government. Again, NIPA G often measures government spending long after production occurs. This is a feature of NIPA's accounting methodology: NIPA G measures government purchases as transactions occur (i.e. when purchased goods become government property), rather than measuring the the value of goods purchased by the government as goods are produced. During the intermediate time interval between when a good is produced and when it is sold (payment received and property transfer), the value of the good is recorded as inventory investment by the producing firm.[5]

There is significant variation in contract duration. The time from contract award to contract completion is substantial, on average, especially for larger, more complex contracts (e.g., expensive goods like fighter planes and aircraft carriers). In WWII, when contracts were kept intentionally short (to avoid renegotiating contract terms due to inflation), the median dollar-weighted contract duration was 18 months (Brunet, 2024). In the modern era, however, it is quite common

[5] Note that the value of production still contributes to GDP at the moment it occurs, but is initially accounted for in inventories until it is sold to the government, at which point it becomes a part of NIPA G.

for large military contracts to span 5 years or longer. Using data since 2000, Cox et al. (2024) find a dollar-weighted median contract duration of 42 months. Moreover, shifting the timing of NIPA cannot reproduce either budget authority or contract of authorizations. Intuitively, military spending authorizations can change very sharply at a single point in time, but contract lengths vary significantly, leading to long and variable lags between authorizations and payments. Because of this, shocks to authorizations not only show up later in NIPA, but they are also spread over more time periods (resulting in lower variance). In other words, NIPA can be understood as a delayed moving average of budget authority and contracts. In practice, this means that shocks to military spending are more pronounced in budget authority and military contracts than NIPA G. This has significant ramifications for time-series estimation of the fiscal multiplier, as we will show in sections 3 and 5.4.

Finally, NIPA differs from all budget measures of government spending in that it includes depreciation (consumption of fixed capital). Depreciation is not "spending" in a budgetary sense, but is included in NIPA accounting. Thus, while \$1 of budget authority directly translates into \$1 of NIPA, total NIPA will always exceed total budget authority because depreciation is included in budget authority.

II.2. The Government Spending Process and Economic Activity

Government spending is a slow, complex process. Economic activity directly associated with government spending begins relatively early in the process—when materials are purchased and goods are produced—rather than when goods are delivered to the government and payments are disbursed.

The timing of government spending is highly sensitive to which measure of spending researchers use. In practice, the time elapsed between when a contract is signed and when the contract is completed (and final payments are made to firms) may be years. Moreover, the time elapsed may vary significantly from one contract to another, depending on the scale of the project, the complexity of the goods involved, and production schedules and delays. Thus the lag between spending authorizations (as measured by budget authority or contract awards) and government purchases (as

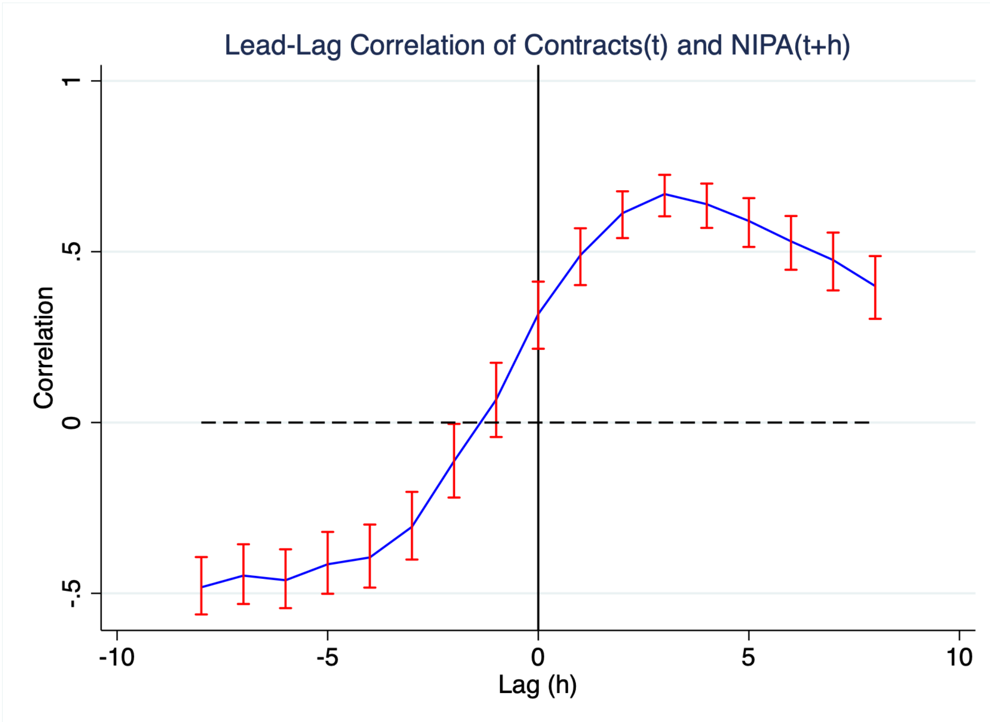


Figure 2: Military Contracts Lead NIPA Defense Procurement Spending

Notes: Sample goes from 1947:1 to 2019:4. The price deflator is the GDP price deflator. Variables are in quarterly year-over-year changes of real per capita values. Confidence intervals are 95%. Defense procurement spending is constructed following Cox et al. (2024): add up (i) NIPA intermediate goods and services purchased, (ii) government gross investment in equipment, structures and software.

measured by NIPA or budget outlays) is both long and variable.

Figure 2 shows the lead-lag correlation map between military contracts and NIPA defense procurement. Notice that the correlation between the two series reaches its maximum when NIPA defense procurement is delayed by three quarters, relative to military contracts.[6] Moreover, the maximum correlation is only 0.65, reflecting the reality that shocks to authorizations (measured by either budget authority or contracts) are spread across multiple periods in NIPA. In Appendix B.3 we also show that Budget Authority leads NIPA defense spending by one year, when measured at fiscal year frequency.

[6] All correlations are calculated using quarterly data and real year-over-year changes.

II.3. Inventories Explain the Timing Mismatch

If authorizations—as measured by budget authority, contract awards, or spending authorizations—consistently lead NIPA by an average of 3 quarters, how can authorizations influence GDP before authorized spending is reflected in NIPA G? This may seem puzzling or even impossible, given that NIPA G is a direct component of GDP.

The answer is in inventories. Specifically, NIPA incorporates accumulation of private inventories, reflecting the the work-in-progress of military contractors before the items are completed and delivered to the government. In other words, the economic activity related to the production of military goods initially enters GDP as a component of inventories, rather than as government purchases.

When a firm receives a contract and begins work—buying materials, hiring and paying workers, beginning production—all of that activity is recorded as inventory investment by the firm receiving the contract. When the firm completes work and delivers final goods to the government—months or years later—NIPA records the government purchase of the finished goods (exactly as described in section 2.1) and negative inventory investment (canceling out the inventories accumulated over previous quarters). The variation in the length of the spending process accounts for both the lag (averaging 3 quarters) and low peak correlation between NIPA and military contracts shown in section 2.2. Since the actual economic activity stemming from government purchases occurs after contracts are awarded (when goods are produced), rather than when completed goods are delivered, government purchases affect economic activity well before they show up in NIPA G.

Earlier generations of economists understood this process, likely because the construction of NIPA was relatively recent and thus much discussed. For example, Hickman (1955, p. 10) described the timing of the economic response to military mobilization in the Korean War:

> It is apparent that a defense mobilization will provide a stimulus to economic expansion if government expenditures increase the aggregate demand for goods and services. However, the expansion need not await the actual growth of government expenditures.

> In the first place, some government expenditures for defense will lag behind the placement of orders. For some time, the increase in production that follows orders will be reflected in private inventory investment rather than in government expenditures.

Time series identification of the aggregate fiscal multiplier depends on the timing of government spending. In section 3, we will use a stylized model to show that measuring government spending too late in the process—that is, using NIPA G to identify government spending shocks—leads to systematic underestimation of the fiscal multiplier.

III. Stylized Model of Measurement Delays in Government Spending

In this section we develop a stylized model of delayed measurement of government spending. Our goal is to build intuition on the differences in timing between various measures of government spending and to illustrate the importance of these differences in the identification of fiscal shocks and the fiscal multiplier. We show that if economic agents react to structural shocks to budget authority and contract awards rather than sufficiently delayed payments-on-delivery, then we can no longer identify true fiscal shocks or the multiplier from payments data alone. In later sections, we will show empirically that this model—-although highly simplified— captures key features of the data.

The model. Suppose that the government spending process begins with authorizations (denoted by B_t), which follow an AR(1) process with white-noise shocks:

$$B_t = \rho \cdot B_{t-1} + \eta_t, \quad \eta_t \sim_{iid} WN(0, \sigma_\eta^2). \tag{1}$$

To capture the delayed accounting of production-in-process, suppose that NIPA G records payments for a share λ of authorized funds with a partial one-period delay, as follows:

$$G_t = (1 - \lambda) \cdot B_t + \lambda \cdot B_{t-1}, \quad \lambda \in [0, 1]. \tag{2}$$

In this setup, values of λ close to zero imply minimal delay in payments relative to authorizations, and $G_t \approx B_t$. Conversely, as λ approaches one, NIPA G becomes approximately a one-period lagged version of authorizations (i.e., $G_t \approx B_{t-1}$). Assuming an AR(1) process for authorizations (Equation (1)) is standard in macroeconomic modeling, while the delayed moving average representation of NIPA G relative to authorization (Equation (2)) is consistent with the *time-to-spend* concept in Leeper, Walker, and Yang (2010) and Ramey (2021), in the context of infrastructure investments.

Moreover, we assume that GDP is driven by an autoregressive term and contemporaneous changes in budget authority reflect increases in inventories coming from production-in-progress. Formally, we can write:

$$Y_t = \phi \cdot Y_{t-1} + \gamma \cdot B_t + \varepsilon_t, \quad \phi \in (0,1),\ \gamma > 0,\ \varepsilon_t \sim_{\text{iid}} WN(0, \sigma_\varepsilon^2), \tag{3}$$

where Y_t is GDP and ε_t is a white noise structural shock to output.

Implications for Variance of Fiscal Shocks. Suppose now that the delay in equation (2) exists, but we proceed anyway with a misspecified recursive VAR model that orders government spending first and approximates B_t with G_t. Combining equations (1) and (2), we can rewrite NIPA G as an ARMA(1,1) process:[7]

$$G_t = \rho \cdot G_{t-1} + \underbrace{(1-\lambda) \cdot \eta_t + \lambda \cdot \eta_{t-1}}_{:=\xi_t}, \tag{4}$$

where the shocks identified directly from lags of NIPA G, ξ_t, represent a weighted average of current and past shocks to budget authority. As a result, the transformed shocks ξ_t underestimate the true

[7] From (1), we have $B_t = (1-\rho L)^{-1} \cdot \eta_t$. From (2), we have $G_t = ((1-\lambda) + \lambda \cdot L) \cdot B_t$. By combining the two expressions and re-arranging, we obtain (4).

variance of underlying structural shocks to government spending:[8]

$$\mathbb{V}ar(\xi_t) \leq \mathbb{V}ar(\eta_t) \quad \forall\, \lambda \in [0, 1],$$

with equality only when $\lambda \in \{0, 1\}$. This result formalizes the idea that shocks to military spending are more pronounced in budget authority and military contracts than in NIPA G. Moreover, this implies that Cholesky shocks to NIPA G should be predictable using shocks identified from more timely measures of government spending (e.g., budget authority or contracts). We verify this empirically in Section 4.4.

Measurement Delays and Identification of Fiscal Shocks. We show in the following proposition that reduced-form structural shocks η_t can only be identified directly from G_t when the share of delayed payments λ is sufficiently small.

Proposition 1 (Identification of Government Spending Shocks from NIPA)**.** *Under equations* (1) *and* (2)*, the moving average representation of the NIPA G process is invertible if and only if* $\lambda < 0.5$*. In other words,* η_t *cannot be identified directly from* G_t *when* $\lambda \geq 0.5$*.*

Proof.

$$\begin{aligned} G_t &= (1-\lambda)\cdot\left(1+\frac{\lambda}{1-\lambda}\cdot L\right)\cdot B_t \quad \text{(Re-arrange terms in Equation (2))} \\ &= (1-\lambda)\cdot\left(1+\frac{\lambda}{1-\lambda}\cdot L\right)\cdot(1-\rho\cdot L)^{-1}\cdot\eta_t \quad \text{(Use Equation (1))} \\ \rightarrow \eta_t &= \frac{1}{1-\lambda}\cdot\left(1+\frac{\lambda}{1-\lambda}\cdot L\right)^{-1}\cdot(1-\rho\cdot L)\cdot G_t. \end{aligned}$$

The term $\left(1+\frac{\lambda}{1-\lambda}\cdot L\right)$ can be inverted, provided that $\lambda < 1-\lambda$, that is, $\lambda < \frac{1}{2}$. ■

In this simple setup, the above result implies that standard Cholesky identification of government spending shocks from NIPA data (i.e., ordering NIPA G first in a VAR à la Blanchard and

[8] In Appendix D, we show empirically that the variance of cyclical fluctuations in NIPA military purchases is lower than the variance of cyclical fluctuations in both budget authority and military contracts.

Perotti, 2002) is only valid for values of $\lambda < 0.5$. When $\lambda \geq 0.5$, VAR methods are still valid, but only for correctly timed measures of government spending such as budget authority and contracts. This argument closely parallels Leeper, Walker, and Yang's (2013) discussion of fiscal foresight.

Moreover, we find that large values of λ are empirically consistent. In particular, we estimate a value of 0.821 for λ via OLS on equation (2) using real per capita values of NIPA defense spending and budget authority. Table 1 provides supporting evidence that our stylized model is a realistic description of government spending data across multiple sample periods (i.e., with and without including the Korean War), regardless of whether OLS coefficients are constrained to sum to one. We find that more than 80% of time-series variation in NIPA G is explained by variation in lags of budget authority. Our empirical results indicate that the measurement delay in NIPA defense spending is significant enough that we cannot recover true fiscal shocks from lags of NIPA alone. We provide further evidence on the relationship between NIPA and spending authorizations in Section 4.4.

Table 1: Share of Budget Authority Recorded with Delay in Spending

Dependent Variable: NIPA Defense Spending (G_t)								
	Constrained Regression (1)				*Unconstrained Regression* (2)			
	Coefficient	*pvalue*	*Coefficient*	*pvalue*	*Coefficient*	*pvalue*	*Coefficient*	*pvalue*
Budget Authority $(1-\lambda)$	0.180	0.019	0.260	0.035	0.067	0.35	0.083	0.432
	(0.074)		(0.133)		(0.071)		(0.123)	
Lagged Budget Authority (λ)	0.820	0.00	0.740	0.00	0.742	0.00	0.635	0.00
	(0.074)		(0.133)		(0.069)		(0.124)	
Sample	1947-2019		1951-2019		1947-2019		1951-2019	
T	72		68		72		68	
R^2	-		-		84.59%		81.23%	

Notes: Columns (1) present results for the constrained regression, while Columns (2) report the unconstrained estimates. Standard errors are reported in parentheses. The dataset covers fiscal years 1947 to 2019. Data for NIPA defense spending is available starting in 1947. All values are in real per capita terms. We use the GDP price deflator (2012=100) to deflate nominal values. Quarterly NIPA data is aggregated at fiscal year frequency.

Implications for the Fiscal Multiplier. To understand the implications of measurement delays on the fiscal multiplier in our stylized model, we write the impulse response of output at time $t+h$

to a fiscal shock at time t implied by Equation (3), as follows:

$$\frac{\partial Y_{t+h}}{\partial \eta_t} = \gamma \cdot \frac{\rho^{h+1} - \phi^{h+1}}{\rho - \phi}$$

Similarly, given equation (4), the impulse response function of NIPA G at time $t + h$ to a fiscal shock at time t can be written as follows:

$$\frac{\partial G_{t+h}}{\partial \eta_t} = \begin{cases} 1 - \lambda & \text{if } h = 0, \\ \rho^{h-1} \cdot (\lambda + \rho \cdot (1 - \lambda)) & \text{if } h > 0. \end{cases}$$

We then sum the impulse response functions at different horizons h to obtain the cumulative impulse response. Next, we define the fiscal multiplier as the ratio of the cumulative impulse response of output and the cumulative impulse response of government spending.[9] Intuitively, this definition represents "*the effect of $1 of government spending on GDP*." Formally, the multiplier at horizon H is equal to:

$$\mathcal{M}(H) := \frac{\sum_{h=0}^{H} \frac{\partial Y_{t+h}}{\partial \eta_t}}{\sum_{h=0}^{H} \frac{\partial G_{t+h}}{\partial \eta_t}} = \begin{cases} \dfrac{\gamma}{1 - \lambda} & \text{if } H = 0, \\ \dfrac{\gamma \cdot \frac{f(\rho) - f(\phi)}{\rho - \phi}}{1 - \lambda + (\lambda + \rho(1 - \lambda)) \frac{1 - \rho^H}{1 - \rho}} \quad \text{where } f(x) := \frac{x(1 - x^{H+1})}{1 - x} & \text{if } H > 0, \\ \frac{\gamma}{1 - \phi} & \text{if } H \to \infty. \end{cases} \tag{5}$$

Expression (5) provides key insights on the role of government spending delays via λ on the multiplier's shape across different horizons. When delays in NIPA G relative to budget authority are significant (i.e., λ close to 1), then the impact multiplier $\mathcal{M}(0)$ is mechanically inflated, and eventually diverges to infinity as a result of GDP responding before the effect of shocks to budget authority actually materialize in NIPA G. For horizons greater than zero, increasing the share of

[9] We adopt this definition for the remainder of the paper, consistent with Ramey (2016), Stock and Watson (2018), and Plagborg-Møller and Wolf (2021).

payment delays also inflates multipliers, but by a finite amount for all horizons greater than zero. The only exception is the infinite horizon multiplier, which is not affected by delayed payments at all. Moreover, as long as $\lambda > \phi$, then the multiplier is a decreasing function over the horizon H, moving from its peak at horizon 0 down to its long-run value. We will show in Section 5 that this shape is consistent with empirical estimates of the fiscal multiplier.

Suppose now that we chose not to distinguish between B_t and G_t, and consider the following misspecified model of output:

$$Y_t = \tilde{\phi} \cdot Y_{t-1} + \tilde{\gamma} \cdot G_t + u_t, \tag{6}$$

where u_t is an error term orthogonal to lagged output and government spending, and $\tilde{\phi}$ and $\tilde{\gamma}$ are misspecified parameters.[10] In this model, we define impulse responses in terms of smoothed shocks ξ_t as opposed to the true structural shocks to budget authority η_t.[11] In particular, we write:

$$\frac{\partial G_{t+h}}{\partial \xi_t} = \rho^h.$$

We can also write the misspecified multiplier as follows:

$$\tilde{\mathcal{M}}(H) := \frac{\sum_{h=0}^{H} \dfrac{\partial Y_{t+h}}{\partial \xi_t}}{\sum_{h=0}^{H} \dfrac{\partial G_{t+h}}{\partial \xi_t}} = \begin{cases} \tilde{\gamma} & \text{if } H = 0, \\ \dfrac{\tilde{\gamma} \cdot \frac{f(\rho) - f(\tilde{\phi})}{\rho - \tilde{\phi}}}{\frac{1-\rho^H}{1-\rho}} \quad \text{where } f(x) := \frac{x(1-x^{H+1})}{1-x} & \text{if } H > 0, \\ \frac{\tilde{\gamma}}{1-\tilde{\phi}} & \text{if } H \to \infty. \end{cases} \tag{7}$$

The theorem below formalizes the comparison between the true and misspecified multipliers. In particular, we show that the misspecified multiplier defined in terms of NIPA is consistently lower than the multiplier obtained using authorizations:

[10] Note that these parameters can be written as functions of the true underlying parameters of the model (i.e., $\lambda, \phi, \gamma, \rho, \sigma_b$, and σ_y).

[11] In Equation (4), note that we can write the misspecified persistence parameter as $\tilde{\rho} = \rho + \frac{\lambda(1-\lambda)(1-\rho^2)}{1-2\lambda(1-\lambda)(1-\rho)} > \rho$. Higher persistence in NIPA inflates the cumulative impulse response of NIPA to smoothed shocks, and in turn dampens the resulting misspecified fiscal multiplier.

Theorem 1. *Let the data be generated by Equations* (1)*,* (2)*, and* (3)*. If $\lambda > \frac{1}{2}$, i.e., there are significant measurement delays in NIPA, the (large sample) fiscal multiplier obtained using NIPA is biased downward:*

$$\tilde{\mathcal{M}}(H) < \mathcal{M}(H) \quad \forall H \geq 0.$$

Proof. See Online Appendix C.

While the formal proof is primarily algebraic and deferred to the Appendix, we highlight pieces which provide insight into the origin of the downward bias, emphasizing why researchers should be cautious when using delayed measures of government spending to estimate multipliers.

To illustrate this, consider the relationship between NIPA and authorizations in Equation (2), rewritten in terms of B_t:

$$B_t = \frac{1}{1-\lambda} \cdot G_t - \frac{\lambda}{1-\lambda} \cdot B_{t-1}.$$

Substitute this expression into the output equation (Equation (3)):

$$Y_t = \phi \cdot \underbrace{Y_{t-1}}_{f(\eta_{t-1}, \eta_{t-2}, \dots)} + \gamma \cdot \frac{1}{1-\lambda} \cdot \underbrace{G_t}_{g(\eta_{t-1}, \eta_{t-2}, \dots)} - \gamma \cdot \frac{\lambda}{1-\lambda} \cdot \underbrace{B_{t-1}}_{h(\eta_{t-1}, \eta_{t-2}, \dots)} + \varepsilon_t.$$

As shown, lagged authorizations, B_{t-1}, behave as an omitted variable in the mis-specified output equation (6). Using a control function approach, we can orthogonalize lagged authorizations and derive the mis-specified model:

$$Y_t = \underbrace{\left(\phi - \gamma \frac{\lambda}{1-\lambda} \theta_1\right)}_{:=\tilde{\phi}} \cdot Y_{t-1} + \underbrace{\gamma \cdot \frac{1-\lambda\theta_2}{1-\lambda}}_{:=\tilde{\gamma}} \cdot G_t + \underbrace{\varepsilon_t - \gamma \frac{\lambda}{1-\lambda} e_t}_{:=u_t},$$

where θ_1 and θ_2 are the (large sample) coefficients from regressing B_{t-1} on Y_{t-1} and G_t, respectively, and e_t is a white noise error orthogonal to Y_{t-1} and G_t.

The bias in the mis-specified parameters vanishes as $\lambda \to 0$ and $G_t \approx B_t$. The source of the bias lies in the measurement delay: past shocks to authorizations, η_{t-1}, affect output at time $t-1$, but the fraction λ of these shocks is recorded in NIPA with delay.

We visualize the above result in Figure 3. The left panel shows the impulse response functions of output, authorizations, and NIPA government spending to a unit shock to authorizations, i.e., $d\eta_t = 1$, obtained from the stylized model when $\lambda = 0.82$ (Table 1). Notice that authorizations jump on impact to one, and output jumps to $\gamma = 0.71$. This immediate response captures the NIPA practice of using private inventories to record contractors' work-in-progress. Only a minimal fraction of authorizations, $1 - \lambda = 0.18$, is recorded immediately in NIPA government spending. After one period, NIPA government spending records the fraction λ of the shock, while the previously response persists through the process: $\rho(1-\lambda)$. One period after the initial shock, the total response of NIPA to the shock is $\lambda + \rho(1-\lambda)$. Thereafter, NIPA government spending decays geometrically. NIPA smooths shocks relative to authorizations and also responds with a noticeable delay.

The right panel of the same figure shows the multipliers calculated through each spending measure in our stylized model. The orange line represents the true fiscal multiplier, $\mathcal{M}(H)$, which can be consistently estimated using authorizations. Conversely, the blue line represents the misspecified multiplier, $\tilde{\mathcal{M}}(H)$, obtained using NIPA.

In this simulation, the impact multiplier, $\gamma/(1-\lambda) = 3.96$, represents the maximum value. After impact, the multiplier decreases, converging to its long-run value: $\gamma/(1-\phi) = 0.95$. In Section 5, we will demonstrate that fiscal multipliers estimated using timely measures of government spending indeed exhibit this shape.

IV. Data

Comparing multiple measures of government spending requires consistent time series for each measure, ideally going back far enough to include WWII and the Korean War. Obtaining consistent historical data on NIPA was relatively straightforward.[12] In contrast, constructing consistent historical time series for budget authority and military contracts required significant effort.

Section 4.1 describes our data sources for budget authority, while section 4.2 details our sources

[12] Officially, both GDP and NIPA "G" are measured annually prior to 1947. To extend our data back to 1939, we use quarterly data from Ramey (2011). The original data source is a 1954 supplement to the *Survey of Current Business*, described in Appendix I.C of Ramey (2011), p. 44.

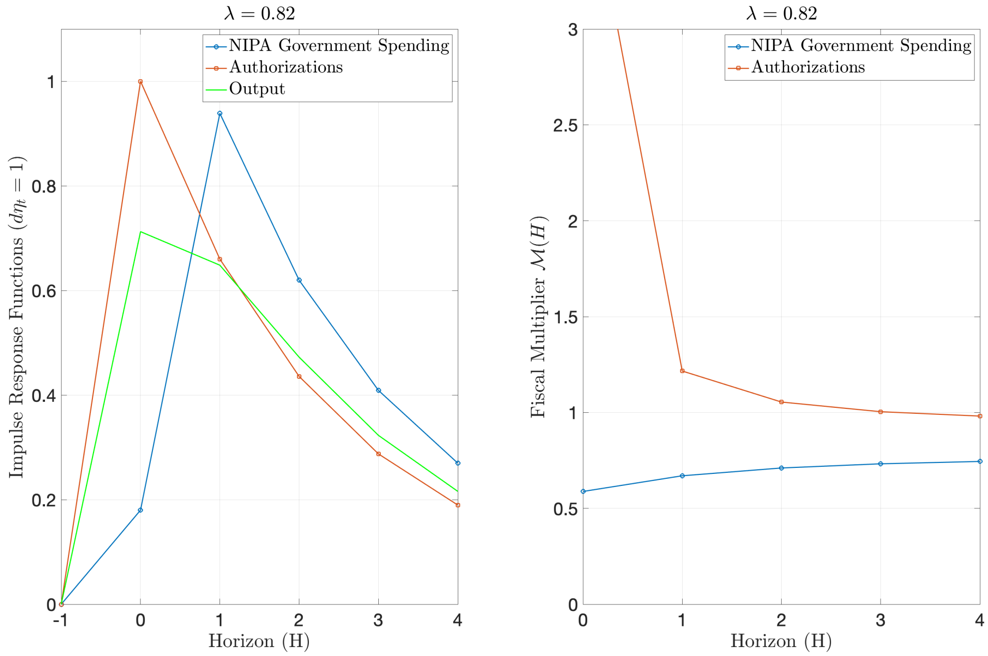


Figure 3: Fiscal Multiplier from Stylized Model

Notes: We simulate data from Equations (1), (2), and (3), and calibrate model parameters based on fiscal year changes in real per capita values of budget authority, NIPA defense spending, and output: $\rho = 0.66$, $\sigma_b = 0.30$, $\phi = 0.25$, $\sigma_y = 0.65$. We set $\lambda = 0.82$, as shown in Table 1, and $\gamma = 0.7125$ to match our baseline estimate of the four-year multiplier.

for military contracts. Section 4.3 explains how we combine budget authority and contracts to form a single quarterly measure of spending authorizations. Section 4.4 compares spending authorizations with NIPA, including both lead-lag correlations and Granger causality tests. These descriptions are necessarily brief. Further details are provided in Appendix A.

IV.1. Budget Authority

While the federal budget process has changed over time, some aspects of it (e.g., appropriations, the President's budget) have remained the same since the 1920s. At the federal level, budgeting has always been done on an annual basis. Budget authority is granted for the federal fiscal year; money does not need to be spent by the end of the year, but it must be obligated (legally committed) by the end of the year or it is returned to the Treasury's general fund.

Budget authority is similar to but broader than appropriations. $1 of appropriations directly creates $1 of budget authority, but budget authority also includes government spending not funded through the annual appropriations process (i.e., "mandatory" or "entitlement" spending on programs like Social Security). In addition, in the (relatively rare) instances when money is appropriated but never obligated (committed) or spent before the appropriation expires, budget authority will show a (negative) adjustment for the unspent funds, formally called a rescission. Unlike appropriations, budget authority incorporates rescissions, which can sometimes be important (e.g., the billions of dollars in canceled military contracts at the end of WWII are recorded as a reduction in budget authority, but would not be accounted for in appropriations).

Annual budget authority totals are available from the Office of Management and Budget (OMB) for 1976 onward, disaggregated by category of spending. We construct a historical series of defense budget authority for 1938-1975 using the *Budgets of the United States*, which have been issued annually since 1923.[13] The advantage of budget authority is that it is a broad measure of defense spending (including direct spending and not just military contracts), while its primary drawback is its annual frequency.

IV.2. Military Contracts

We also construct a new series of quarterly military contracts from 1940 onward. These military contracts (also referred to as "military contracts") measure the dollar value of contracts awarded to private firms by the Department of Defense. Contract values are assigned to the contract award date (typically a specific month or quarter). Thus contracts anticipate production and government expenditures as measured by NIPA. The advantage of defense contract data is its higher frequency: military contracts are available at least quarterly, and for some periods even monthly.

To construct a complete quarterly time series of military contracts, we combine military contract data from many different sources, each covering a different time period. A brief overview of our data sources is presented below. Further details are provided in Appendix A.

[13] We begin our series in 1938 because prior to the New Deal budget totals were regarded as upper bounds rather than as targets, so budgetary information for earlier years cannot be interpreted in the same way. See Schick (1990) for the institutional history of this change.

(i) 1940:1-1945:4. The Civilian Production Administration (successor to the War Production Board, or WPB) published a complete listing of WWII contracts in 1946. The data encompass all war contracts of $50,000 or more awarded between June 1940 and September 1945, excluding only food products and electricity generation. This data is described in detail in Brunet (2024).

We aggregate military contracts to quarters using contract award date and annualize these values. We refer to this series as WPB, referring to the data source. Unfortunately, the contract data are not complete after 1945:3 (the end of WWII).

(ii) 1945:4-1950:4. We do not have data on military contracts for 1945:4-1950:4. The WWII data does not report (complete) contract awards after 1945:3, and the next complete data source (described below) does not begin until 1951:1. Due to lack of better options, we fill the gap in military contract data from 1945:4 to 1946:4 using a simple linear interpolation.

Because macroeconomic data collection expanded significantly in 1947, we can do better from 1947 onwards. We estimate military contracts over 1947:1 to 1950:4 using two variables highly correlated with military contracts: (i) average weekly hours of production workers in aircraft manufacturing and (ii) NIPA defense procurement spending. Average weekly hours in aircraft manufacturing are available monthly from the BLS (discontinued databases). Aircraft manufacturing hours strongly comove with military contracts since aircraft and their components represent a large fraction of military purchases, and vice versa (see Appendix A for further details). We construct NIPA defense procurement following Cox et al. (2024): we sum NIPA defense intermediate goods and services purchased with NIPA defense gross investments on structure, equipment and software.

We use our BCD military contract data for 1951:1 to 1980:4 (described below) to estimate the following equation via OLS:

$$\text{military contracts}_t = \kappa + \beta \cdot (\text{Avg.Hours.Aircraft})_t + \sum_{h=0}^{4} \psi_h \cdot \text{NIPA}_{t+h} + \varepsilon_t \tag{8}$$

The in-sample R^2 is 64.8%.

We then extrapolate military contracts for 1947:1 to 1950:4 using the estimated OLS coefficients, along with observed values of aircraft production hours and NIPA defense procurement. Contracts extrapolated using this method increase sharply with the outbreak of the Korean war in 1950. The timing of the shock is driven by the aircraft production data and is consistent with the narrative of the Survey of Current Business from August 1950. Further details are reported in the Online Appendix A.4. Because we use the BCD data for these estimates, we refer to these data as "*BCD Extrapolated*."

(iii) 1951:1 - 1982:4. We take military contract data for 1951:1 through 1982:4 from Ramey (1989). The original data sources are *Business Condition Digest*, and *Business Cyclical Developments* before 1961 ("BCD").[14] BCD reports seasonally adjusted military contracts monthly from January 1951 until November 1988.[15]

(iv) 1983:1-2003:3. We use the annual Official Federal Procurement Summary Report ("FPSR"), produced by the Directorate for Information Operations and Reports (DIOR), for 1983:1 to 2003:3. FPSR contains data on the value of military contract awards by fiscal year. It also features bar charts illustrating quarterly values of total federal procurement contracts (combined defense and non-defense). On average, 80% of total federal contracts were military contracts. We interpolate the fiscal-year series of military procurement contracts using the quarterly values of all federal contracts.

(v) 2003:4 onward. Starting from the fourth quarter of 2003, all daily federal procurement transactions are reported in the Next Generation of FPDS ("FPDS-NG"). We aggregate contracts by fiscal year and then interpolate the series using variation in newly awarded military contracts aggregated by quarters using methodology analogous to what we employed for the FPSR data. We

[14] Military contracts became part of the set of instruments known as the Hall-Ramey instruments. See Hall (1990) and Ramey (1989).

[15] The series was discontinued in 1988 and then migrated to the Survey of Current Business (SCB) in 1990. The SCB data is available from January 1990 through September 1995, but systematically omits the fourth quarter of each year. Because of this, SCB is less reliable. To maximize data quality, we use the BCD data through the end of 1982, switching to the FPSR data when it becomes available. Using BCD data through 1988 produces similar but less precise results.

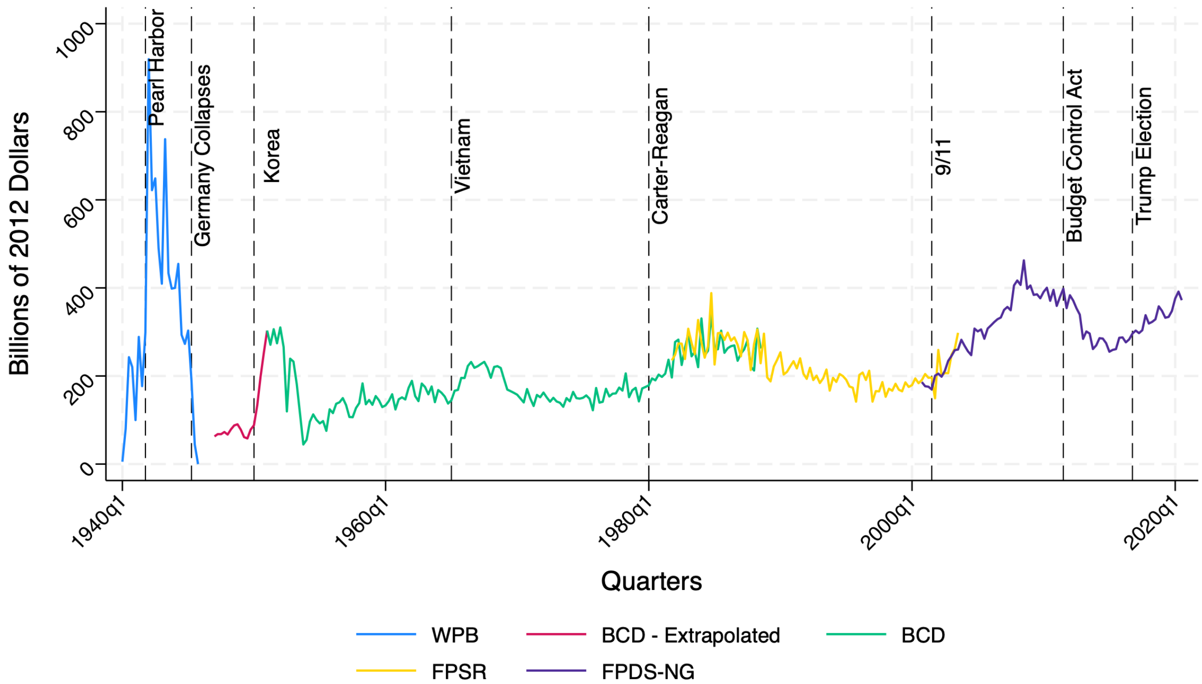


Figure 4: Measures of Military Contracts

Notes: The price deflator is the GDP price deflator. Price deflator before 1947:1 comes from Ramey and Zubairy (2018), adjusted to be consistent with 2012 as base year. We seasonally adjust FPSR and FPDS-NG data using X13 ARIMA-SEATS program from the Census Bureau. Data is annualized.

take this approach rather than aggregating contracts at quarterly frequency because contracts are sometimes canceled in the next quarter, exaggerating quarterly variation. Aggregating by fiscal year and then interpolating helps avoid this noise.

Assembling the Time Series: Military Contracts 1940 to 2019. Figure 4 displays real military contracts in billions of 2012 dollars, using the data sources detailed above. Quarterly dollar amounts are annualized. Vertical lines depict Ramey and Shapiro (1998)'s war dates, augmented with the Pearl Harbor attack in the fourth quarter of 1941, the collapse of Germany, the 9/11, the Budget Control Act of 2011, and the first election of President Trump.

When two measures of military contracts are available for the same quarter, as in the 1980s and early 2000s, both measures are displayed. The discrepancies between measures are extremely small, reflecting consistency across data sources (and relative ease of measuring military contracts).

Because the measures are so similar, the choice of which source we rely on when multiple sources are available does not meaningfully affect our results. We always choose the measure we perceive to be more accurate and complete. In the remainder of the paper, we will refer to this new series as *military contracts*, or sometimes simply *contracts*.

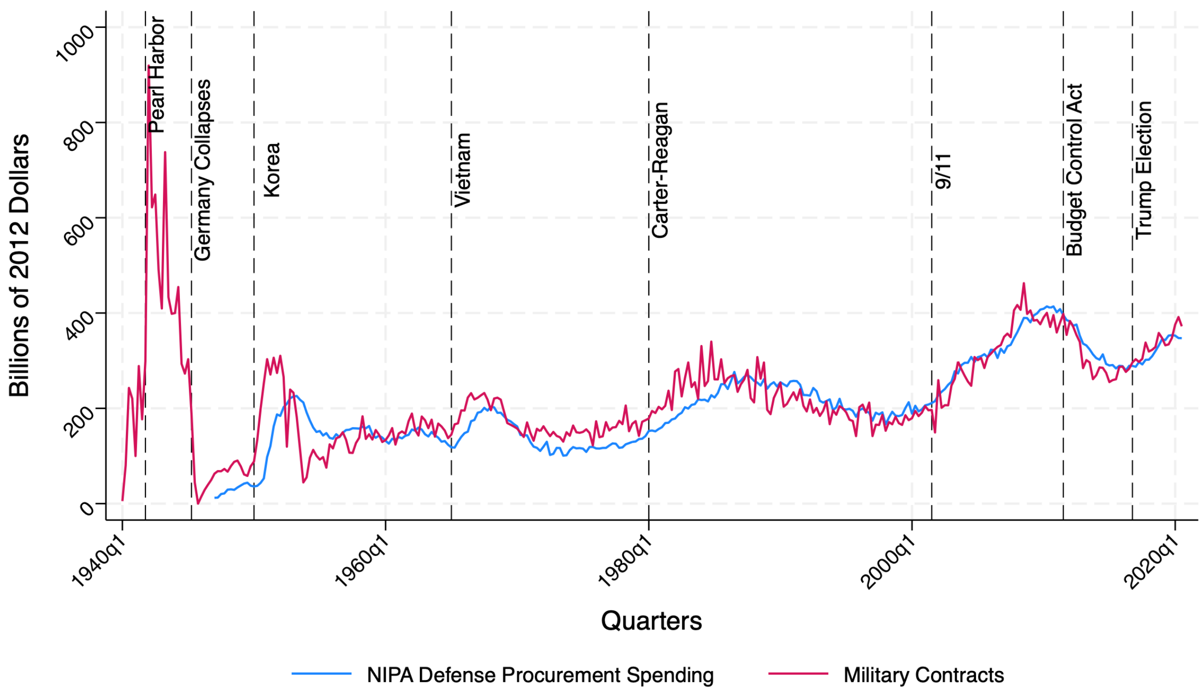


Figure 5: Measures of Military Contracts

Notes: The price deflator is the GDP price deflator. Price deflator before 1947:1 comes from Ramey and Zubairy (2018), adjusted to be consistent with 2012 as base year. We linearly interpolate data for calendar year 1946.

Figure 5 displays our newly constructed series of military contracts (red line) alongside NIPA defense procurement spending (blue line), constructed following Cox et al. (2024). The latter series begins in 1947 (when quarterly NIPA was introduced). In Appendix A.7 we show that our contract data, aggregated by fiscal year, closely tracks the data used by Dupor and Guerrero (2017).[16]

[16] In turn, the series used by Dupor and Guerrero (2017) is closely related to the series used by Nakamura and Steinsson (2014). An advantage of contract data is its consistency within the literature, even when compiled by different scholars.

IV.3. Spending Authorizations: Interpolating Budget Authority with Military Contracts

Both budget authority and contract data have distinct advantages and disadvantages. Budget authority is the most complete measure of spending authorizations—it includes all authorized spending and accounts for rescissions when previous authorizations are rescinded–but its use for fiscal multiplier measurement is limited by its annual frequency. In contrast, military contracts are a less complete measure of defense spending—they omit direct spending (e.g., personnel costs) and production at government facilities (e.g., Navy Yards both owned and operated by the government)—but are available at a much higher frequency. On an annual basis, the two measures are coincident and very highly correlated, as shown in Figure 1.

We combine budget authority and military contracts into a single new measure, which we call *spending authorizations*. We construct spending authorizations by interpolating quarterly military contract data into annual defense budget authority. Specifically, we use the quarterly variation in military contracts within a fiscal year to assign quarterly values of budget authority so that average budget authority over the fiscal year matches our annual budget authority measure.[17]

Figure 6 shows defense spending authorizations and budget authority, both in per capita 2012 dollars. The top panel shows the WWII and Korean War, while the bottom panel shows 1947 to 2020. We show these periods separately because the scale of spending authorizations during WWII is so large that it becomes very difficult to see post-1950 variation when displaying the entire series with one scale.

IV.4. Comparing Spending Authorizations and NIPA

The upper panel of Figure 7 shows defense spending authorizations and NIPA defense spending from 1947 to 2016. Consistent with the predictions of our stylized model, note that spending authorizations increase both sooner and more sharply than NIPA following significant shocks to military

[17] Until 1976, the federal fiscal year was July 1 until June 30 (Q3 to Q2). Following a "transition quarter" in the summer of 1976, the federal fiscal year shifted and began October 1, ending September 30 of the following year (Q4 to Q3). Happily, defense budget authority for the transition quarter (Q3 1976) is reported in the historical budget tables, allowing us to construct a continuous quarterly measure with no gaps. See OMB Historical Table 5.1 ("Budget Authority by Function and Subfunction"), currently reported here.

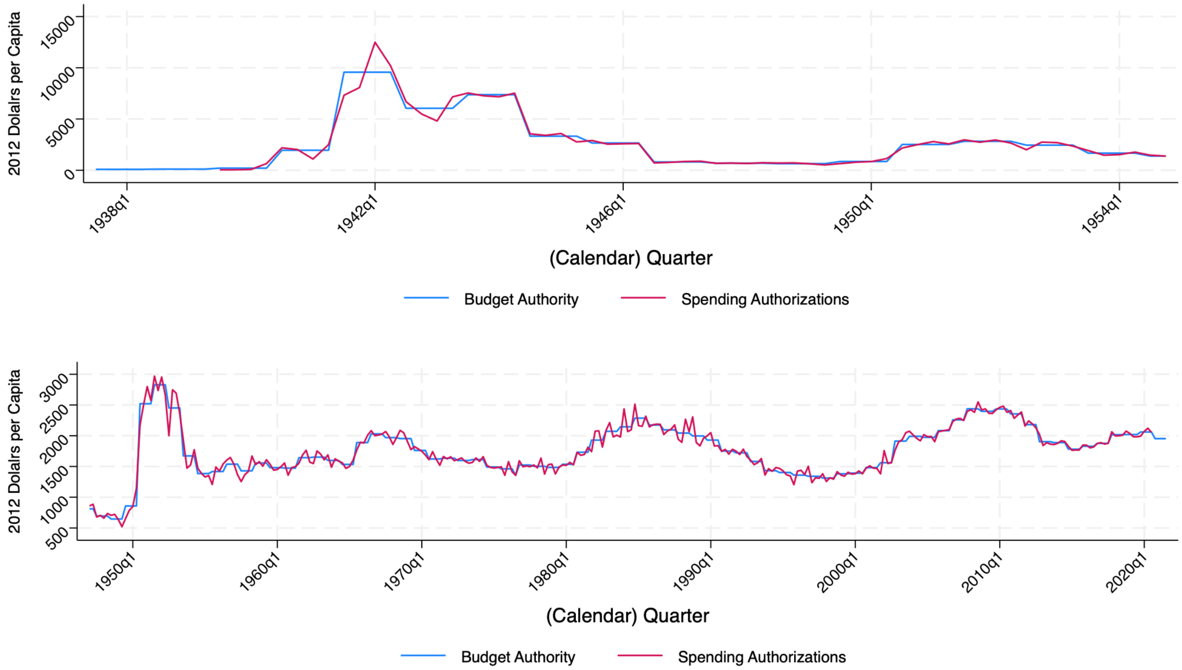


Figure 6: Spending Authorization and Budget Authority (with and without WWII)

spending. NIPA defense spending typically exceeds defense spending authorizations because NIPA includes consumption of fixed capital (depreciation), which is not included in budgetary measures of spending. Despite this discrepancy, every dollar of spending authorizations eventually translates to a dollar of NIPA.

The lower panel of Figure 7 shows the lag-lead correlation map of year-over-year changes in defense spending authorizations and NIPA defense spending. Changes in spending authorizations lead changes in NIPA by three quarters on average. However, note that the correlation remains high (above 0.5) throughout *t+1* to *t+6*. This is consistent with NIPA being a delayed moving average of past spending authorizations, as in our stylized model from section 3.

While sections 2, 3, and 4 provide extensive intuition for the timing differences between spending measures, it is helpful to formally show that defense spending authorizations Granger cause NIPA defense spending, but not vice versa. The upper panel of Table 2 shows Granger causality tests of defense spending authorizations versus NIPA defense spending. Our main sample begins in the first quarter of 1947, when official quarterly macroeconomic data first becomes available.

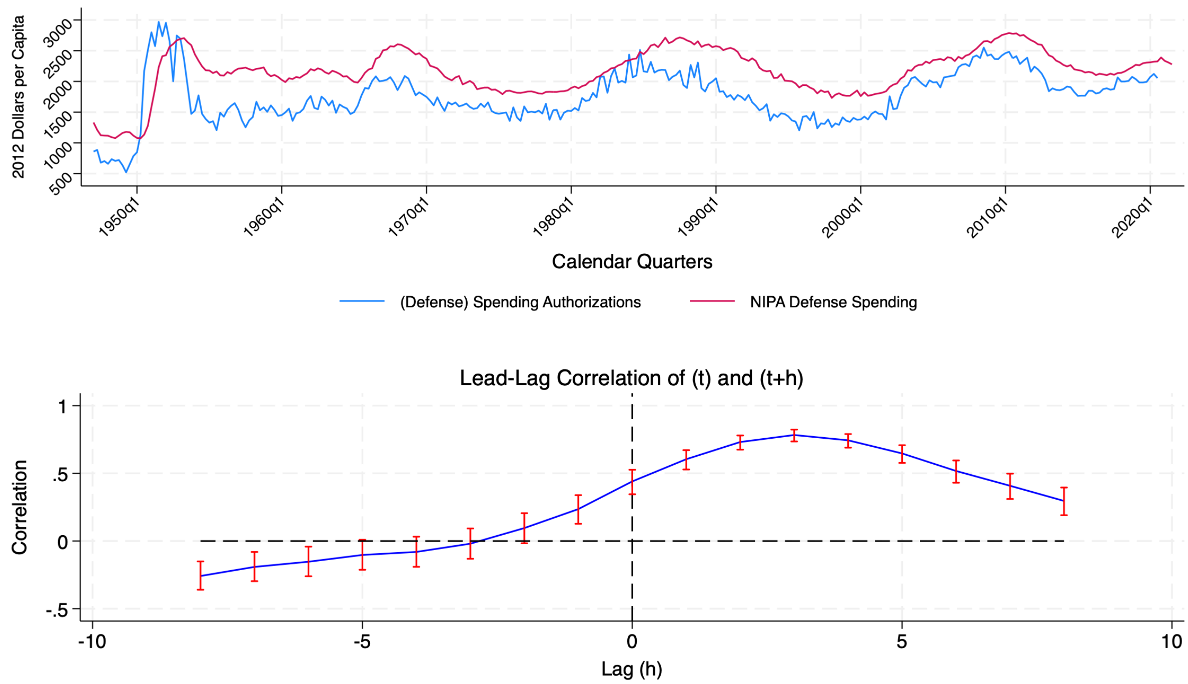


Figure 7: Comparing Defense Spending Authorizations and NIPA Defense Spending

We also present results starting from the first quarter of 1951, for two reasons. First, quarterly variation in spending authorizations comes from military contract data, and our contract data for 1946 to 1950 is significantly lower quality than our contract data for later years. Second, the third quarter of 1950 is highly leveraged due to the large spending shock following the outbreak of the Korean War (see Perotti, 2014; Ramey, 2016; Dupor and Guerrero, 2017). Starting our sample in 1951 allows us to ensure that neither of these factors is driving our results. We also show an extended sample, going back to the first quarter of 1940 and including WWII. We extend the sample backwards to 1940 using unofficial national account data from Ramey (2011). Whichever sample period we use, defense spending authorizations Granger Cause NIPA defense spending but not vice versa.

Appendix B.5 reports further Granger causality tests using several other measures of government spending, including budget authority, military contracts, and budget outlays. Our findings are consistent with the conceptual arguments presented in section 2: budget authority and contracts both Granger cause NIPA defense spending, but not vice versa. NIPA defense spending is indeed

Table 2: Granger Causality Tests

Predicted	Predictor	Sample	F Statistic	pvalue
Spending Authorizations vs. NIPA Defense Spending				
NIPA Defense Spending	Spending Authorizations	1947:1 - 2019:4	13.916	0
NIPA Defense Spending	Spending Authorizations	1951:1 - 2019:4	12.145	0
NIPA Government Spending	Spending Authorizations	1940:1 - 2019:4 7	15.806	0
Spending Authorizations	NIPA Defense Spending	1947:1 - 2019:4	1.678	.104
Spending Authorizations	NIPA Defense Spending	1951:1 - 2019:4	.852	.558
Spending Authorizations	NIPA Government Spending	1940:1 - 2019:4	2.059	.04
Cholesky Shocks to NIPA Government Spending vs. Cholesky Shocks to Spending Authorizations				
Predicted	*Predictor*	*Sample*	*F Statistic*	*pvalue*
NIPA Government Spending	Spending Authorizations	1947:1 - 2019:4	8.688	0
NIPA Government Spending	Spending Authorizations	1951:1 - 2019:4	8.782	0
Spending Authorizations	NIPA Government Spending	1947:1 - 2019:4	1.182	.311
Spending Authorizations	NIPA Government Spending	1951:1 - 2019:4	1.186	.308

Notes: Granger causality tests are conducted using Stata's `vargranger` post-estimation command from the `var` command, which estimates a bivariate VAR model for the predicted variable and the predictor with `p` lags. The VAR configuration employs the `dfk` and `small` options of the `var` command to adjust the F statistics for small-sample bias. The testing procedure involves (i) running an OLS regression of the predicted variable on `p` lags of itself along with `p` lags of the predictor, and (ii) conducting a Wald test to assess the null hypothesis that the `p` lags of the predictor are jointly non-significant. The lag `p` is set to *eight*. Variables are in real per capita values. Price deflator is the GDP price deflator (2012=100).

delayed relative to all of our measures of authorizations.

The lower panel of Table 2 shows that the timing advantage of spending authorizations extends beyond NIPA to standard measures of spending shocks derived from NIPA. Following the spirit of Blanchard and Perotti (2002), we construct Cholesky shocks by ordering NIPA government spending first in a recursive recursive VAR which also includes GDP, the 3-month T-Bill rate, and Romer and Romer (2010)'s narrative series of exogenous tax changes.[18] We then use spending authorizations to identify fiscal shocks, treating spending authorizations as an internal instrument in a recursive recursive VAR. Once again, the measure of fiscal shocks derived from spending authorizations Granger causes the measure derived from NIPA: applying sophisticated VAR techniques does not alter the fundamental timing difference between the underlying measures of defense spending.

[18] GDP and G are in logs of real per capita values. The Romer and Romer tax shocks are in real per capita values. Deflator is the GDP price deflator.

These results are consistent with the stylized model from section 3: shocks to NIPA represent a moving average of current and past shocks to budget authority and contracts. One measure predicts the other, but not vice versa.

V. New Results from a New Measure

A major advantage of spending authorizations is that it is compatible with a wide variety of approaches and methodologies. We show how our measure can be incorporated into two of the most popular current methodologies, local projections and structural VARs.

Spending authorizations are essentially an internal instrument for NIPA government spending (in the sense of Plagborg-Møller and Wolf, 2021). They meet the criteria for identifying structural shocks laid out in Ramey (2016). They must be (i) uncorrelated with contemporaneous and lagged values of endogenous variables, (ii) uncorrelated with other shocks and (iii) unanticipated. Ordering spending authorizations first in a structural VAR meets condition (i) by construction. The plausible exogeneity of shocks to military spending fulfills condition (ii). Finally, condition (iii) is also more plausible at the quarterly frequency because we now observe spending shocks when they first occur, rather than with a delay of 3-4 quarters. For both approaches we use the same baseline sample period, described below.

Sample. Our baseline sample spans from 1947:1 to 2007:4. While spending authorizations extend back to 1940:1, official quarterly data begin in 1947, making it a natural starting point, and the economic environment prevailing during WWII was extremely unusual due to rationing, price controls, and other economic interventions. In the twenty-first century, short-run changes in aggregate government spending have been dominated by non-defense spending, largely due to the rise of large countercyclical stimulus packages (in 2001, 2008-09, and 2020-21). This stands in stark contrast to the twentieth century, where short-run fluctuations in aggregate government spending were primarily driven by changes in military spending. Given this structural break and given that the Romer and Romer tax shock series ends in 2007, we end our baseline sample in 2007. Appendix

G presents results using other samples, all of which are qualitatively similar.

The literature has debated whether it is appropriate to include the Korean War in the sample when estimating the fiscal multiplier (see Perotti, 2014; Dupor and Guerrero, 2017). We include the Korean War in our baseline sample because, unlike WWII, the U.S. economy never transitioned to a full scale war economy (with price controls, rationing, and other non-market mechanisms; see Hickman, 1955). The outbreak of the Korean War thus provides meaningful variation that is comparable to other shocks in our data (though highly leveraged). As a robustness check, we also show results using a sample starting after the outbreak of the Korean War.

V.1. Local Projections

We estimate the impulse response functions of different components of GDP in response to a shock to spending authorizations using what Ramey (2016) calls the Hall-Barro-Redlick approach to local projections, after Jordà (2005). We estimate the impulse responses of a variety of macroeconomic variables to spending authorizations at time horizon h. Both spending authorizations (SA) and the outcome variable (Y) are normalized by the initial level of GDP. When the outcome is GDP, the impulse response can be interpreted as the GDP multiplier.

Following best practices in the literature, we include four quarters of lagged values for a standard set of controls: GDP, spending authorizations, NIPA government spending, the three-month Treasury Bill rate, exogenous tax shocks from Romer and Romer (2010). When the outcome variable is not GDP, we also include lagged values of the outcome. All control variables measured in dollars (that is, everything but the 3-month Treasury Bill rate) are normalized by the lagged initial level of GDP. All dollar variables are measured in real (2012) dollars.

Our estimating equation is thus:[19]

$$\frac{Y_{t+h} - Y_{t-1}}{\text{GDP}_{t-1}} = \beta^h \cdot \frac{\text{SA}_t - \text{SA}_{t-1}}{\text{GDP}_{t-1}} + (\text{Lagged Controls}) + \varepsilon_{t+h}, \tag{9}$$

[19]Formally, the control lags are the following:
$\sum_{j=1}^{4} \left(\beta_j^h \cdot \frac{\text{SA}_{t-j} - \text{SA}_{t-1-j}}{\text{GDP}_{t-1-j}} + \gamma_j^h \cdot \frac{\text{G}_{t-j} - \text{G}_{t-1-j}}{\text{GDP}_{t-1-j}} + \phi_j^h \cdot \frac{\text{GDP}_{t-j} - \text{GDP}_{t-1-j}}{\text{GDP}_{t-1-j}} + \rho_j^h \cdot \frac{\text{Y}_{t-j} - \text{Y}_{t-1-j}}{\text{GDP}_{t-1-j}} + \pi_j^h \cdot \text{TB3}_{t-j} + \tau_j^h \cdot \frac{\text{Tax}_{t-j}^{\text{R\&R10}}}{\text{GDP}_{t-1-j}} \right).$

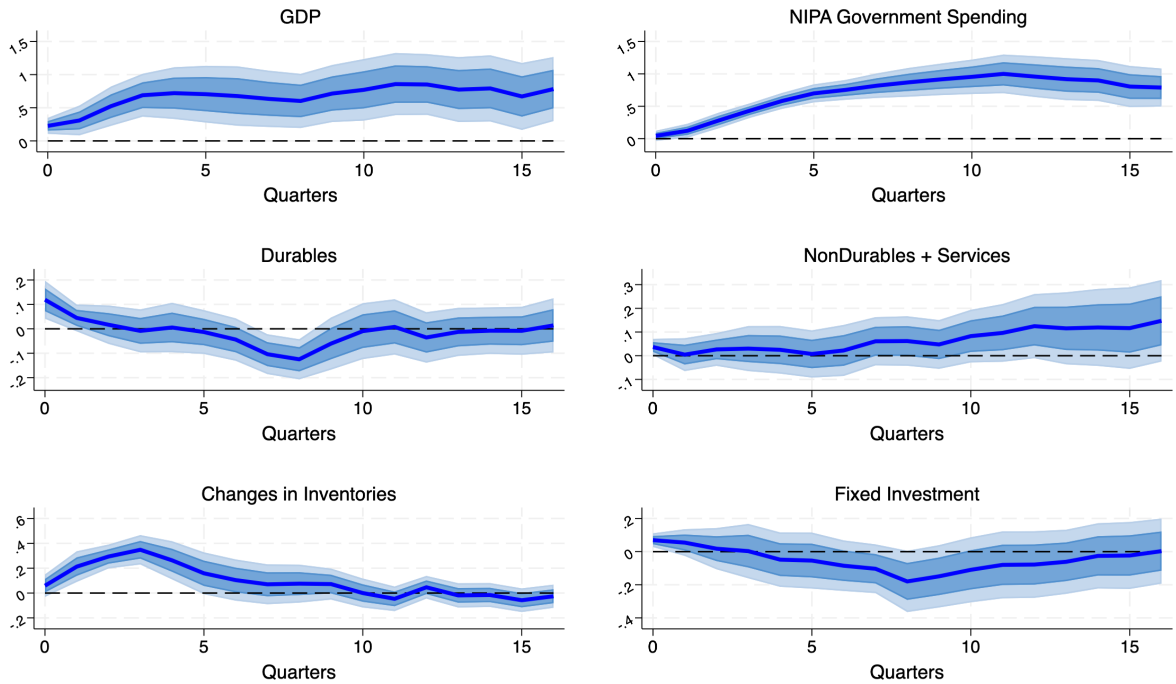


Figure 8: LP Responses of GDP Components to Spending Authorizations

Notes: Responses are normalized by the peak response of the IRF of G. Confidence bands represent 68% and 90%. Standard errors are heteroskedasticity-robust (Montiel Olea and Plagborg-Møller, 2021). Sample: 1947:1 to 2007:4.

where the OLS estimate of β^h represents the impulse response function at horizon h.

Figure 8 plots the impulse responses (β^h) of six outcomes in response to a shock in spending authorizations. For comparability across outcomes, all values are normalized by the peak response of G. The top panel shows the impulse responses of GDP (left) and NIPA government purchases (right). Both GDP and NIPA government purchases rise quickly—immediately, in the case of GDP—and remain elevated for four years after the initial shock. This is consistent with the model presented in section 3: GDP responds immediately to spending authorizations—significantly faster than it responds to NIPA G. NIPA G peaks 7 to 11 quarters after the shock to spending authorizations.

The middle panel shows the impulse response of consumption, as broken down into durables (left) and non-durables plus services (right). Durables consumption rises significantly in the immediate aftermath of the shock before falling to zero.[20] In contrast, non-durables and services consumption initially remains unchanged, before slowly rising more than a year after the initial shock—though statistically the increase in non-durables and services consumption is only marginally significant.

The bottom panel shows the impulse responses of inventories (left) and fixed investment (right). Inventories rise immediately and stay elevated throughout the first year after the shock, after which point there is no systematic effect on inventories. This reflects the mechanism discussed at greater length in section 5.3. Fixed investment initially responds positively to the spending shock, but then turns negative. Samples excluding the Korean War yield similar results, although the positive response of investment is more pronounced and the negative response is never significant.[21]

V.2. Structural VARs

Because the *only* meaningful difference between spending authorizations and NIPA is timing, ordering spending authorizations first in a structural VAR solves the timing problem in NIPA. This is not true of other anticipated spending measures, as discussed above in section 4.4. The Granger causality tests reported in the lower panel of Table 2 confirm that shocks to spending authorizations cannot be predicted by quarterly macroeconomic variables and Cholesky-identified shocks to NIPA G.

Following the extensive SVAR literature, we include a number of control variables (ordered later in the VAR), all expressed in natural logs of real per capita values.[22] The controls include total NIPA government purchases, the three-month Treasury Bill rate, and exogenous tax shocks from Romer and Romer (2010).

[20] The negative response of durables from quarters 6 to 9 may reflect the outbreak of the Korean War: households anticipated shortages of goods, as they had during WWII, and bought in advance of anticipated shortages (see Hickman, 1955; Ramey, 2016; Binder and Brunet, 2021). In samples beginning after 1950, the dip in durables consumption in quarters 6 to 9 is not statistically significant. Notably, however, the initial increase in durables consumption also lacks statistical significance when the outbreak of the Korean War is excluded (see Online Appendix F).

[21] The weaker response of fixed investment when including the outbreak of the Korean War may reflect regulations limiting investment during the Korean War (see Perotti, 2014).

[22] Because inventories can take on negative values, we do not use the log transformation on inventories, but simply include real per capita values.

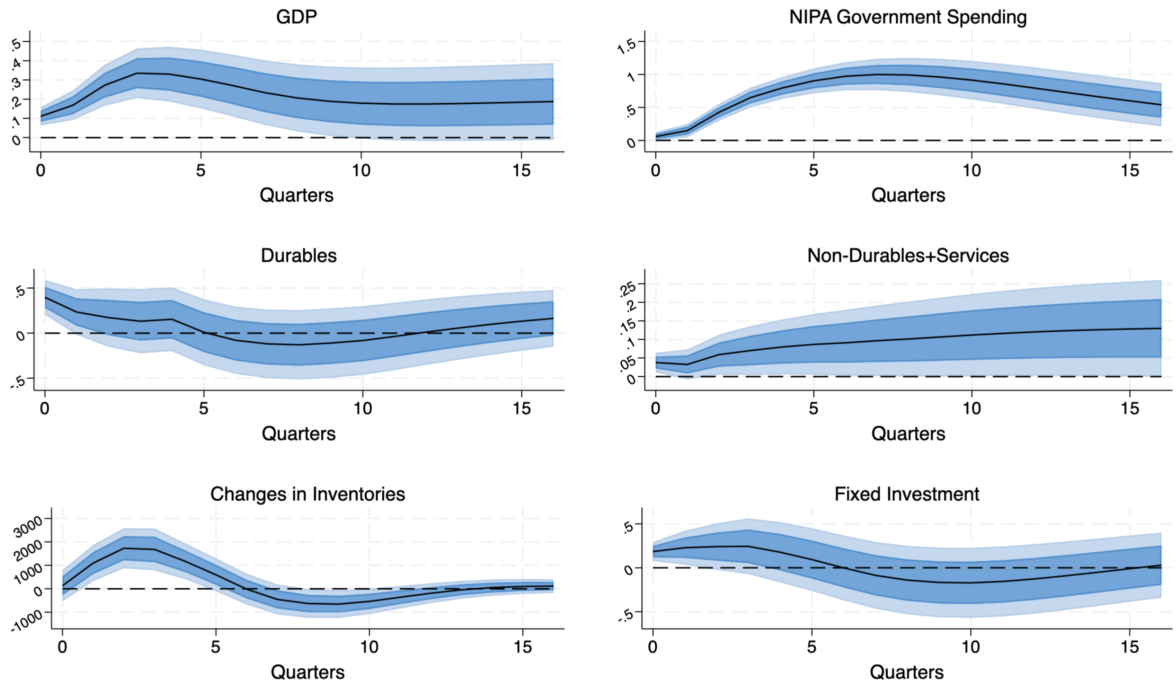


Figure 9: VAR Responses of GDP Components to Spending Authorizations

Notes: Responses are normalized by the peak response of the IRF of G. Confidence bands represent 68% and 90%. Sample: 1947:1 to 2007:4.

Figure 9 plots the impulse responses of the same six outcomes shown in Figure 8, this time estimated with a recursive VAR. Spending authorizations are ordered first in the SVAR. The top panel shows the impulse responses of GDP (left) and NIPA government purchases (right). Once again, GDP responds immediately while NIPA G responds more slowly. Both remain elevated for four years after the initial shock. Once again, NIPA G peaks more than a year after the initial shock.

The middle panel again shows the impulse response of consumption, as broken down into durables (left) and non-durables plus services (right). Durables again respond strongly—and positively—on impact, and then quickly become statistically insignificant. The response of non-durables plus services, which accounts for 83% of total consumption on average, has greater statistical significance in the VAR impulse response estimates, though the shape and magnitude of the response is quite similar using both linear projections and VARs.

The bottom panel shows the responses of inventories and fixed investment. Inventories exhibit a strong, significant, and positive response to a fiscal shock. Notably, the response is faster than that of the NIPA measure of government spending, becoming large and significant as early as the first quarter after the shock. The role of inventories is discussed at greater length in section 5.3. Finally, the response of fixed investment is again positive on impact but again quickly becomes statistically insignificant.

V.3. Evidence for the Inventory Mechanism

As discussed in section 2.3, inventories explain the timing mismatch between NIPA G and GDP: new military production initially shows up in GDP as inventory investment rather than in NIPA G. Before we address how using spending authorizations affects fiscal multiplier estimates, we present evidence that inventories indeed rise—and quickly—in response to a shock to spending authorizations. As shown in Figures 8 and 9, inventories begin rising immediately after the shock, peaking three quarters after the shock at approximately 40% of the peak response of government spending, which occurs many quarters later.

We confirm that increase in inventories following a shock to military spending is driven by military production. Military production is heavily concentrated in specific manufacturing subsectors (Ramey and Shapiro, 1998; Perotti, 2007; Nekarda and Ramey, 2011; Fisher and Peters, 2010; Cox et al., 2024). In Appendix I, we show that the inventory response to spending authorizations is concentrated in the manufacturing subsectors—such as "other transportation equipment" (SIC 3364) and "computer and electronic products" (SIC 334)—with the highest exposure to military contracts. We further show that these sectors are not more responsive to business cycle conditions (i.e. monetary policy shocks), so this result is not driven by higher volatility in these sectors.

The scale of inventory response is also large enough to explain the multiplier estimates presented below in section 5.4, particularly the private spending multiplier. For instance, in the Vietnam War buildup, inventories increased by about 0.75% of GDP, while military contracts rose by approximately 1.7% of GDP. Thus the scale of the inventory response is large enough for inventories to play the important role we claim, but not so large as to be implausible.[23]

V.4. Fiscal Multipliers

V.4.1. LP-IV multipliers

We directly estimate fiscal multipliers using local projections instrumental variables (LP-IV) following Ramey (2016), Stock and Watson (2018), and Plagborg-Møller and Wolf (2021). Specifically, we estimate the following equation:

$$\sum_{h=0}^{H} \frac{\text{GDP}_{t+h} - \text{GDP}_{t-1}}{\text{GDP}_{t-1}} = \mathcal{M}_H \cdot \underbrace{\sum_{h=0}^{H} \frac{\text{G}_{t+h} - \text{G}_{t-1}}{\text{GDP}_{t-1}}}_{\text{Instrument with } Z_t} + \text{lags} + \nu_t, \tag{10}$$

where lags include four lags of the variables from Equation (9). The cumulative change in G is instrumented with:

$$Z_t := \frac{\text{SA}_t - \text{SA}_{t-1}}{\text{GDP}_{t-1}}.$$

The 2SLS estimate of $\mathcal{M}_H$ provides a direct estimate of the cumulative fiscal multiplier at horizon H in response to a shock to spending authorizations. Notably, the point estimates obtained using this method are numerically equivalent to the ratio of the cumulative impulse response function of GDP (top-left panel of Figure 8) to the cumulative impulse response function of G (top-right panel of Figure 8).

Using spending authorizations as an instrument for government spending allows us to account for contemporaneous changes in the non-defense components of total NIPA government spending (G). Intuitively, if military build-ups systematically crowd out other non-military components of G, failing to account for this margin of adjustment could bias multiplier estimates downwards. In addition, instrumenting for changes in total government spending preserves the traditional definition

[23] In particular, this surge was driven by government purchases of McDonnell-Douglas F-4 Phantom II aircraft. Sales of these aircraft (starting in 1966:3) amounted to 1.2 billion dollars, or 0.15% of GDP, accounting for nearly 9% of the total rise in contracts during this buildup. This contract only covered the airframe; as other systems such as navigation, communication, weapon, and engine components were procured separately from other contractors (e.g., General Electric).

of the multiplier: $\mathcal{M}_H$ measures the effect of an additional \$1 of NIPA G on GDP.

Table 3 presents multiplier estimates at time horizons from impact to four years after the shock, estimated using various methods. The uppermost panel uses the baseline sample, beginning in 1947 with modern macroeconomic data. The middle panel starts later, omitting the outbreak of the Korean War, as suggested by Dupor and Guerrero (2017). The bottom panel shows estimates beginning in 1940, including WWII as well as the Korean War.

The leftmost columns of Table 3 present these LP-IV multiplier estimates. The Montiel Olea and Pflueger (2013) effective F-statistics are presented to the right of each multiplier estimate.[24] For comparison, the next pair of columns use the same LP-IV approach, but use NIPA military spending as the instrument for total government spending instead of spending authorizations. Effective F statistics are shown to the right of these estimates, as well. These NIPA-derived estimates range from approximately 0.5 to 0.8 for samples starting before 1950 (and are too imprecise to draw conclusions for samples beginning after 1950), consistent with many prior estimates in the aggregate multiplier literature, e.g. Hall (2009).

Figure 10 shows the impulse response of the fiscal multiplier (estimated using the baseline sample) from 1 to 16 quarters following the shock, depicting both 68% and 90% confidence bands. Ignoring for the moment the very large multiplier estimates upon impact—discussed further in section 5.4.3—we estimate 1-year multipliers significantly above unity, around 1.7. The estimated multiplier then descends to 1 and remains there for all longer time horizons.

At all time horizons estimated and for all samples excluding WWII, the multiplier estimated using defense spending authorizations is substantially larger than the multiplier estimated using NIPA defense spending. The two estimates do converge over time, as we should expect given that NIPA defense spending is a delayed moving average of spending authorizations, but even at a time horizon of 4 years there is a substantial gap between the estimates (0.95 vs. 0.61). This gap is

[24] The effective F-statistics are well below the the 5% and 10% weak instrument thresholds (37 and 23) suggested by Montiel Olea and Pflueger (2013) on impact, but they rapidly rise above these thresholds at longer time horizons. This weak instrument problem is a result of the timing mismatch: NIPA G does not respond to spending authorizations for several quarters. Other anticipation measures, such as Ramey and Zubairy's defense news shocks, suffer from the same problem. This issue is discussed at length in section 5.4.3. A detailed comparison between spending authorizations and Ramey's news shocks is presented in online Appendix F.2.

Table 3: Instrumenting for G with Authorizations Corrects for Systematic Downward Bias in Estimates of the Cumulative Fiscal Multiplier

	Baseline Sample: 1947:1 - 2007:4					
	Local Projections (LP-IV)				Recursive VAR	
	Spending Authorizations		NIPA Defense Spending		Spending Authorizations	NIPA Defense Spending
	Multiplier	*Effective F*	*Multiplier*	*Effective F*	*Multiplier*	*Multiplier*
Impact	5.12	0.87	0.80	462.26	9.18	0.84
	(4.92)		(0.21)			
1 Year	1.68	31.11	0.60	107.73	2.88	0.67
	(0.42)		(0.22)			
2 Years	1.10	98.42	0.48	69.48	1.85	0.56
	(0.25)		(0.19)			
3 Years	0.98	66.69	0.55	56.28	1.53	0.62
	(0.19)		(0.18)			
4 years	0.95	47.11	0.61	49.33	1.51	0.75
	(0.19)		(0.19)			
	Sample Excluding Outbreak of Korean War: 1951:1 - 2007:4					
	Spending Authorizations		NIPA Defense Spending		Spending Authorizations	NIPA Defense Spending
	Multiplier	*Effective F*	*Multiplier*	*Effective F*	*Multiplier*	*Multiplier*
Impact	2.12	15.24	1.08	348.73	2.37	1.20
	(0.85)		(0.21)			
1 Year	1.53	19.28	0.70	83.14	2.43	1.45
	(0.61)		(0.34)			
2 Year	0.82	21.61	0.23	50.75	2.21	1.34
	(0.47)		(0.37)			
3 Year	0.64	23.45	0.25	39.57	2.10	1.34
	(0.44)		(0.41)			
4 year	0.75	23.83	0.35	31.23	2.10	1.39
	(0.44)		(0.46)			
	Sample Including WWII and Most Recent Years: 1940:1 - 2019:4					
	Local Projections (LP-IV)				Recursive VAR	
	Spending Authorizations		NIPA Government Spending		Spending Authorizations	NIPA Government Spending
	Multiplier	*Effective F*	*Multiplier*	*Effective F*	*Multiplier*	*Multiplier*
Impact	-4.66	0.01	0.75	-	5.05	1.01
	(66.58)		(0.10)			
1 Year	0.46	16.40	0.60	92.58	1.32	0.83
	(0.12)		(0.05)			
2 Year	0.57	17.46	0.65	61.22	1.11	0.92
	(0.06)		(0.04)			
3 Year	0.63	16.36	0.68	46.97	1.14	0.99
	(0.05)		(0.05)			
4 year	0.66	14.07	0.72	33.39	1.21	1.06
	(0.05)		(0.05)			

Table 3: Robust standard errors in parentheses. Effective F calculated with `weakivtest` on Stata (Montiel Olea and Pflueger (2013). VAR multipliers are constructed by dividing the cumulative IRF of GDP by the cumulative IRF of G; then the value is multiplied by $1/T\sum_{t=1947:1}^{2007:4} Y_t/G_t$ to convert the elasticity multiplier into a fiscal multiplier.

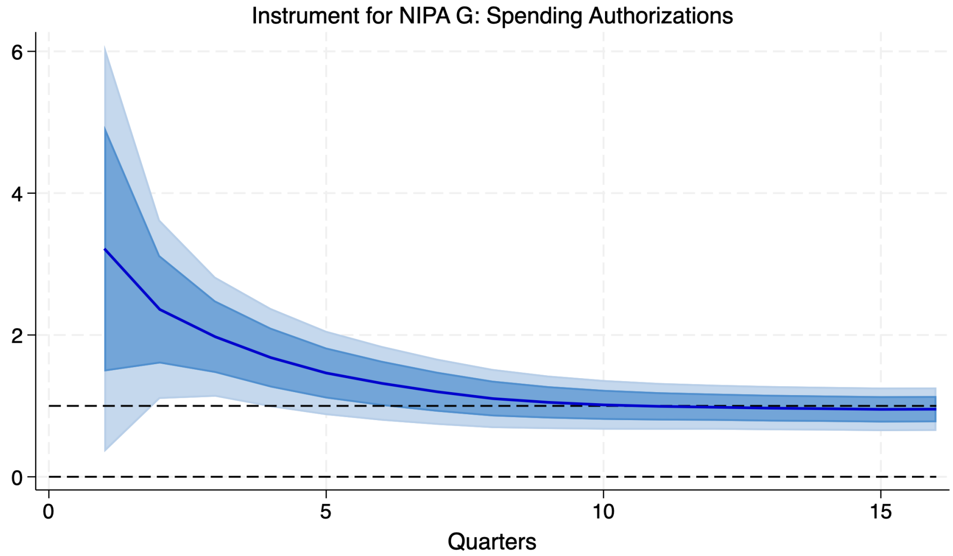


Figure 10: The Cumulative Fiscal Multiplier Reaches 1.7 at the One-Year Horizon and Stabilizes at 1.0 by Year Two

Notes: We report LP-IV estimates using the sample from 1947 to 2007. Confidence bands represent 68% and 90% confidence intervals, calculated with heteroskedasticity-robust standard errors following (Montiel Olea and Plagborg-Møller, 2021).

even larger when the sample begins after the onset of the Korean War. As suggested by our model in section 3, the timing delays inherent to NIPA typically result in a systematic downward bias of aggregate multiplier estimates when NIPA is used to identify government spending shocks.

The literature has long struggled with imprecisely estimated multipliers for samples excluding the onset of the Korean War, largely because there is so much less variation in military spending when WWII and the Korean War are excluded (see Dupor and Guerrero, 2017). While our estimates are much less precise for this sample, we do produce statistically-significant multiplier estimates for this sample at the 1-year time horizon (and indeed, at all time horizons from impact to 7 quarters).

Including WWII in the sample reduces the size of the multiplier estimated using spending authorizations, including relative to multipliers estimated using NIPA. These results are consistent with prior literature, including both estimates specific to WWII (Fishback and Cullen, 2013; Brunet, 2024) and studies estimating multipliers for periods both including and excluding WWII (e.g. Hall, 2009; Ramey and Zubairy, 2018). The scale of fiscal shocks associated with WWII is so large that

WWII will dominate the results whenever it is included in the sample. Two sets of factors explain the WWII-inclusive multiplier estimates shown in the bottom panel of Table 3.

First, multiplier estimates for periods including WWII are typically lower than for periods excluding WWII because of the wartime economic environment. Government control of strategic materials (production inputs) made it extremely difficult for firms to acquire materials or begin production quickly. The labor market was also exceptionally tight, with an unemployment rate below 1% for several years.[25] At the same time, contracts were intentionally short and paid very quickly, resulting in minimal payment delays relative to production delays.

Second, contracts were sharply reduced at the end of WWII, and many contracts were canceled. This coincided with the end of rationing (in the US) and a large increase in private consumption. The change in spending authorizations is large and negative at the end of WWII, even while private consumption rebounded. In contrast, NIPA G declined later and less sharply at the end of WWII (as many contracts in process in 1945 were completed and delivered). This coincidence of timing regarding the end of WWII likely biases NIPA multiplier estimates upwards because of the large increase in private consumption at a time when new military spending was sharply contracting.

V.4.2. VAR Multipliers

Another approach to estimating the fiscal multiplier is to use recursive VAR estimates, dividing the cumulative response of GDP up to horizon H by the cumulative response of government spending. We follow this procedure, using the VAR approach outlined in section 5.2 and alternatively using NIPA defense spending (the main driver of short-term variation in NIPA G) instead of defense spending authorizations. These estimates are presented in the two rightmost columns of table 3. Once again, multipliers derived using spending authorizations are substantially larger than those derived using NIPA. The 1-year multiplier is around 2.9, which falls to 1.5 by year 3. Starting the sample after the initial Korean War shock produces larger multiplier estimates at some time

[25] Unemployment rates for WWII come from the NBER Macrohistory Database. Labor hoarding was fairly common in WWII, as firms could not hire quickly. Brunet (2024) shows that non-manufacturing private employment initially shrank in response to war production in state-level panel data, indicative of a very high degree of crowd out. This environment made it extraordinarily difficult for firms to act quickly on contract awards.

horizons, though it is difficult to draw conclusions due to lower precision. Multiplier estimates for samples including WWII are again smaller than multiplier estimates when WWII is excluded from the sample. As always, the downside of the VAR approach is that we must calculate the multiplier from elasticity estimates, taking a ratio and then adjusting it by noisy sample averages, rather than estimating it directly, reducing precision.

V.4.3. Estimation Challenges, Particularly at Short Time Horizons

Our reported multiplier estimates upon impact (at the time of the shock) and very short time horizons are extremely large but highly imprecise. The low precision at these very short time horizons is due to the timing mismatch between authorizations and NIPA, since both methods use spending authorizations as an instrument for NIPA G.[26] Because it takes several quarters for shocks to spending authorizations to show up in NIPA G, spending authorizations are not a good predictor of NIPA G at very short time horizons, as reflected in the very low F-statistics reported for spending authorizations at the shortest time horizons.

Table 4 presents multiplier estimates by quarter from impact through 2 years post-shock, allowing us to examine these results in greater detail. Whether these estimates are accurate or inflated depends on exactly how we define the fiscal multiplier. If we define the fiscal multiplier as the economic impact per dollar of *government spending* (as measured by NIPA), the very large short-term multipliers reported in sections 5.4.1 and 5.4.2 are correct, if imprecise.

If, however, we define the fiscal multiplier on spending authorizations as the economic impact per dollar *authorized*, then the short-time-horizon multiplier estimates are artificially inflated. At short time horizons, new authorizations outpace new spending (in the sense of dollars leaving the Treasury) by a significant margin. Recall that the fiscal multiplier is obtained by dividing the cumulative impulse response of GDP by the cumulative impulse response of G. Since we use spending authorizations as an instrument for NIPA G, the denominator is spending in sense of dol-

[26] In this paper we always use shocks to military spending—by whatever means they are constructed—as instruments for shocks to total G. If non-defense spending is systematically reduced (crowded out) when defense spending increases, directly estimating the fiscal multiplier on defense spending may bias estimates downwards.

Table 4: Estimated Multipliers Are Very Large at Short Time Horizons,
But Also Highly Sensitive to Precise Multiplier Definition

	Local Projections (LP-IV)				Recursive VAR	
	Spending Authorizations		NIPA Defense Spending		Spending Authorizations	NIPA Defense Sp.
	Multiplier	*Effective F*	*Multiplier*	*Effective F*	*Multiplier*	*Multiplier*
Impact	5.12	0.87	0.80	462.26	9.18	0.84
	(4.92)		(0.21)			
Quarter 1	3.22	2.33	0.75	247.40	6.58	0.72
	(1.74)		(0.23)			
Quarter 2	2.36	6.55	0.81	158.66	4.27	0.76
	(0.76)		(0.23)			
Quarter 3	1.97	15.19	0.73	125.99	3.39	0.71
	(0.51)		(0.22)			
Quarter 4	1.68	31.11	0.60	107.73	2.88	0.67
	(0.42)		(0.22)			
Quarter 5	1.46	54.09	0.50	91.96	2.51	0.62
	(0.36)		(0.21)			
Quarter 6	1.32	77.09	0.46	82.74	2.23	0.59
	(0.32)		(0.20)			
Quarter 7	1.20	94.81	0.46	75.04	2.01	0.57
	(0.28)		(0.20)			
Quarter 8	1.10	98.42	0.48	69.48	1.85	0.56
	(0.25)		(0.19)			

Table 3: Table reports cumulative fiscal multipliers at time horizon given in leftmost column of each row. Robust standard errors in parentheses. Effective F calculated with `weakivtest` *on Stata (Montiel Olea and Pflueger (2013). The VAR multipliers are constructed by dividing the cumulative IRF of GDP by the cumulative IRF of G; then the value is multiplied by* $1/T \sum_{t=1947:1}^{2007:4} Y_t/G_t$ *to convert the elasticity multiplier into a fiscal multiplier.*

lars that have left the Treasury (NIPA G) rather than authorizations. To the extent that cumulative spending authorizations outpace cumulative NIPA G, the multipliers calculated so far will overstate the multiplier on spending authorizations (economic impact per dollar authorized) by the ratio of cumulative spending authorizations to cumulative NIPA G.

To address this issue, we decompose the fiscal multiplier to separate the response of G to authorizations from other components. Define the response of output to authorized spending—that is, the economic impact per dollar authorized—as $\mathcal{M}_H^A$. We further define the response of NIPA G to authorizations as $\mathcal{G}_H^A$.[27] Then the "fiscal multiplier" estimated in Equation (10) above, $\mathcal{M}_H$—that is, the cumulative economic impact per dollar of NIPA at time horizon H—can be decomposed:

$$\mathcal{M}_H = \frac{\mathcal{M}_H^A}{\mathcal{G}_H^A} \tag{11}$$

At sufficiently long time time horizons, the cumulative response of NIPA G to spending authoriza-

[27] For a fuller mathematical treatment of this section, see Online Appendix H.1.

tions converges to one (that is, $lim_{H\rightarrow\infty}\mathcal{G}_H^A \rightarrow 1$), and this distinction is no longer important since $\mathcal{M}_H$ then converges to $\mathcal{M}_H^A$. However, at most relevant time horizons, $\mathcal{G}_H^A$ is substantially below one, making this distinction extraordinarily important.

To illuminate this distinction and directly examine economic impact per dollar authorized, we define *private spending* as $GDP - G$, consistent with the definition in Ramey (2013).[28] Then we calculate the cumulative economic impact on private spending per dollar authorized as $\mathcal{P}_H^A$. We can decompose the total economic impact per dollar authorized into the impact on G and impact on private spending:

$$\mathcal{M}_H^A = \mathcal{G}_H^A + \mathcal{P}_H^A \tag{12}$$

Substituting into Equation (11) above, this gives us:

$$\mathcal{M}_H = 1 + \frac{\mathcal{P}_H^A}{\mathcal{G}_H^A}. \tag{13}$$

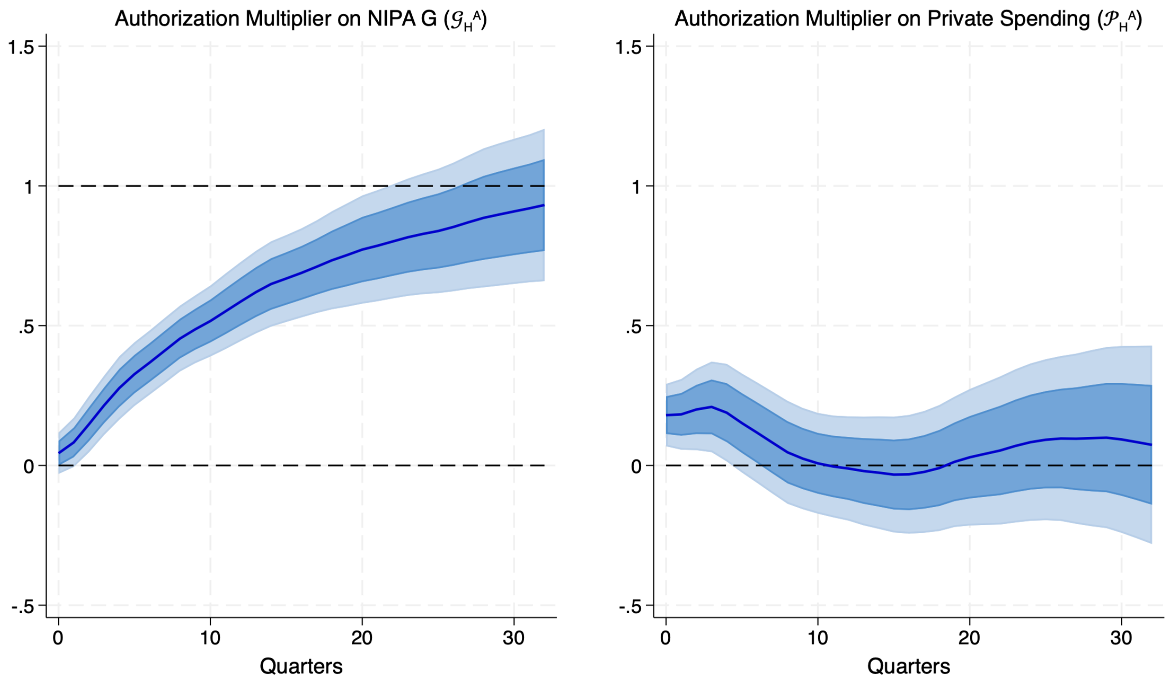
Figure 11 shows the impulse responses of both $\mathcal{G}_H^A$ and $\mathcal{P}_H^A$ to a shock to spending authorizations using our baseline sample. $\mathcal{G}_H^A$ starts just above 0 and slowly converges to 1, taking nearly 8 years before $\mathcal{G}_H^A \rightarrow 1$. This part of the decomposition reflects the speed of the government spending process—that is, how long it takes for newly authorized spending to show up in NIPA G, equivalent to *time-to-spend* in the infrastructure spending literature (Leeper, Walker, and Yang, 2010; Leduc and Wilson, 2013; Ramey, 2021).

For policymakers, the main object of interest is likely $\mathcal{P}_H^A$, or the economic impact on private spending per dollar of authorized spending. Crucially, Equation (13) tells us that as $\mathcal{P}_H^A$ converges to zero, $\mathcal{M}_H$ converges to one. Values of $\mathcal{P}_H^A$ above zero correspond to values of $\mathcal{M}_H$ above one, while values of $\mathcal{P}_H^A$ below zero correspond to values of $\mathcal{M}_H$ below one. Thus estimating $\mathcal{P}_H^A$ sheds light on the multiplier. A value of $\mathcal{P}_H^A$ above zero reflects crowding in of private spending, while a value of $\mathcal{P}_H^A$ below zero reflects crowd out of private spending.

Most policymakers probably want to know the total GDP impact of a fiscal shock per dollar

[28] Notice that *private GDP* is distinct from private spending. Private GDP is equal to private spending (Y-G) plus government purchases of privately produced goods and services, i.e., domestically produced procurement spending.

Figure 11: Components of the Fiscal Multiplier



authorized, ignoring the measurement problems posed by the timing mismatch between authorizations and NIPA. This *heuristic multiplier*, essentially forcing $\mathcal{G}_H^A$ to one even at short time horizons (a dollar authorized will eventually translate into a dollar spent), can be written $1 + \mathcal{P}_H^A$.

Table 5 shows the multiplier per dollar of NIPA, $\mathcal{M}_H$, the private GDP multiplier per dollar authorized, $\mathcal{P}_H^A$, and the heuristic multiplier, $1 + \mathcal{P}_H^A$, at various time horizons. $\mathcal{P}_H^A$ is positive and statistically significant upon impact and for just over a year after impact, after which it falls to zero. $\mathcal{P}_H^A$ peaks 3-4 quarters after impact at 0.21. The multiplier on private spending is initially positive and statistically significant, but becomes indistinguishable from zero just over a year after impact. Point estimates remain weakly positive at most time horizons. Note that the response of private spending corresponds almost exactly to the response of inventories discussed in Section 5.3.

For samples beginning after 1950, we can never reject the hypothesis that $\mathcal{P}_H^A$ is zero at any time horizon, though point estimates remain positive for approximately two years after the shock. For periods excluding WWII, we can conclude that the "true" medium-run $\mathcal{P}_H^A$ is small, but cannot rule out either positive or negative values. When the sample includes WWII, estimates of $\mathcal{P}_H^A$ are strictly

Table 5: Significantly positive values of $\mathcal{P}_H^A$ means crowding-in of private spending
Heuristic Multipliers used to gauge output effect at short horizons

Horizon	$\mathcal{M}_H$	$\mathcal{P}_H^A$	*pvalue*	*Heuristic Multiplier* $(1+\mathcal{P}_H^A)$
Impact	5.122	0.180	0.014	1.180
	(4.917)	(0.067)		(0.067)
Quarter 1	3.217	0.183	0.026	1.183
	(1.737)	(0.076)		(0.076)
Quarter 2	2.360	0.200	0.034	1.200
	(0.764)	(0.088)		(0.088)
Quarter 3	1.975	0.209	0.045	1.209
	(0.511)	(0.098)		(0.098)
Quarter 4	1.681	0.189	0.089	1.189
	(0.422)	(0.106)		(0.106)
Quarter 5	1.464	0.152	0.172	1.152
	(0.359)	(0.107)		(0.107)
Quarter 6	1.318	0.117	0.294	1.117
	(0.318)	(0.109)		(0.109)
Quarter 7	1.199	0.082	0.466	1.082
	(0.282)	(0.110)		(0.110)
Quarter 8	1.104	0.047	0.677	1.047
	(0.252)	(0.111)		(0.111)

Notes: Estimates use data from 1947 to 2007. $\mathcal{M}_H$ *refers to baseline estimates of GDP fiscal multiplier.* $\mathcal{P}_H^A$ *is the private spending authorization multiplier, estimated via LP-IV (see Online Appendix H.1). Heuristic multiplier obtained by adding one to* $\mathcal{P}_H^A$*. Robust standard errors in parentheses.*

negative. This reflects the many constraints of the wartime economic environment, as discussed at the end of Section 5.4.1.

Different readers may have different views on whether "the fiscal multiplier" should mean the economic impact per dollar spent (even when authorizations dramatically outpace spending at short time horizons) or economic impact per dollar authorized. As we have shown, precise estimates vary significantly depending on exactly how we define and understand the multiplier.

What is most striking is that (apart from WWII) we consistently estimate private spending multipliers at or above zero at short time horizons, corresponding to heuristic multipliers (GDP multipliers as typically understood) at or above one. Furthermore, the medium-run multiplier is in the vicinity of one unless influenced by WWII, meaning that government spending does not appear to crowd-out private spending to a significant degree unless there is a total-war economy. This result is robust to estimation approach (linear projection IV or structural VAR). This result is all the more remarkable because in the previous aggregate multiplier literature, estimates produced using NIPA—whether with linear projections or VARs (e.g. Blanchard and Perotti, 2002)—have always been smaller than multiplier estimates obtained through narrative methods that accounted for anticipation effects (e.g. Ramey and Zubairy, 2018). Using spending authorizations—that is, properly accounting for the timing mismatch in NIPA—resolves this puzzle.

V.5. Theoretical Implications

In addition to new evidence on the size of the fiscal multiplier, these results have other important implications for economic theory, particularly the mechanism through which government spending influences economic activity. We show clear evidence that anticipation effects are driven by the behavior of firms, as shown in the immediate response of inventory investment to military spending authorizations. Consumption—particularly durables consumption—does respond in the very short term, but the response of inventories only a quarter or two slower than durables, larger in peak magnitude (see Figure 8), and more sustained: remaining statistically significant until five quarters after the shock rather than just one quarter. The evidence for an inventory channel acting through firm behavior is notable because in almost every model of government spending used to estimate fiscal multipliers, the main mechanism is households' expectations (e.g. expectations of future taxes through income effects and changes in expected labor supply, as in Baxter and King, 1993; Woodford, 2011), especially in sticky price models (see Broer and Krusell, 2021). Our results do not rule out the possibility of strong negative income effects on labor supply, but they do strongly suggest that this channel is probably not the main channel through which government spending influences short-run economic activity.

Rather, the response of inventories to spending authorizations points to the existence of another channel through which government spending influences short-run economic activity. The long and variable lags between spending authorizations and NIPA spending provide a simple non-Ricardian explanation for anticipation effects. All of this is consistent with the recent survey evidence presented by Coibion, Gorodnichenko, and Weber (2021), who conclude that income effects arising from anticipated higher future tax burdens must be minimal. Together, these papers strongly suggest that the Ricardian channel of households responding to expected future tax burdens through expected changes in labor supply is not the primary transmission mechanism through which fiscal policy influences short-run economic activity. Rather, the primary mechanism through which shocks to government spending influence economic activity appears to be through firm behavior, with firms increasing production and accumulating inventories in response to a shock to government spending.

VI. Conclusion

In this paper we argue that the measurement of government spending—particularly its timing—has profound implications for the fiscal multiplier. We argue that anticipation effects reflect the long and variable lags between when new military spending is authorized and when the government makes payments and military spending is recorded in NIPA G. We present a stylized model showing why this timing issue can result in systematic downward bias in estimates of the fiscal multiplier and hinders the identification of fiscal shocks using NIPA.

Because the lag between authorization and payment is long and variable, the bias problem cannot be solved by using leads or lags of NIPA. Nor can defense news measures fully solve the timing problem, as news measures contain other information and so vary from NIPA in other dimensions. To address this measurement problem, we compile historical budget data and extensive data on military contracts. We combine these two measures—budget authority and military contracts—to produce a novel, quarterly measure of military spending, called defense spending authorizations, going back to 1940. Spending authorizations are a simple and flexible solution: they can be used as an instrument for NIPA G in either a linear projection or a VAR framework.

Using spending authorizations as an instrument for NIPA G to correct for systematic bias in the multiplier, we estimate a short-run multiplier at or above one (for time horizons under two years) converging to a medium-run multiplier (for time horizons of two to four years) around one. When

WWII is included in the sample, however, our short- and medium-run multiplier estimates are much lower, between 0.4 and 0.6, reflecting both wartime economic conditions and extraordinarily high statistical leverage of observations from WWII.

Our results also shed light on the mechanism through which government spending affects output in the short run. We show that inventories respond both quickly and strongly to spending authorizations, explaining most of the response of private spending at short time horizons. The strong response of inventories suggests that the primary mechanism through which a military spending shock influences short-run output acts through firms rather than households. This is in stark contrast to much of the literature, which assumes a Ricardian mechanism working through household expectations.

At a deeper level, these findings raise questions about external validity. The literature has long emphasized aggregate multipliers on military spending, not because military spending is a desirable form of fiscal stimulus but because of the extreme reverse causality problems bedeviling time series identification of multipliers on domestic spending. Implicitly (or explicitly, as in the case of Hall, 2009), the time series multiplier literature has assumed that the multiplier on military spending is relevant to understanding the multiplier on various types of domestic spending (particularly those favored in stimulus packages). However, the mechanism shown in this paper may be relatively specific to military spending: around 80 percent of all federal contracts are military contracts, while (outside of major wars) typically less than 20 percent of federal spending is military spending. This raises the likelihood of a complexity that economic theory has heretofore avoided acknowledging: even the short-run effects of government spending may depend on what type of goods (or services) the government buys.

References

Barro, **Robert** and **Redlick**, **Charles J.** (Feb. 2011). Macroeconomic Effects From Government Purchases and Taxes *. *The Quarterly Journal of Economics*. 2011. 126, pp. 51–102.

Baxter, **Marianne** and **King**, **Robert G** (June 1993). Fiscal Policy in General Equilibrium. *American Economic Review*. 1993. 83, pp. 315–334.

Binder, **Carola Conces** and **Brunet**, **Gillian** (Nov. 2021). Inflation expectations and consumption: Evidence from 1951. *Economic Inquiry*. 2021, ecin.13051.

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*. 2002. 117, pp. 1329–1368.

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

Brunet, **Gillian** (Mar. 2024). Stimulus on the Home Front: The State-Level Effects of WWII Spending. *Review of Economics and Statistics*. 2024, pp. 1–46.

Coibion, **Olivier**, **Gorodnichenko**, **Yuriy**, and **Weber**, **Michael** (Feb. 2021). Fiscal Policy and Households' inflation Expectations: Evidence from a Randomized Control Trial. *Woriking Paper*. 2021.

Cox, **Lydia**, **Müller**, **Gernot J.**, **Pastén**, **Ernesto**, **Schoenle**, **Raphael**, and **Weber**, **Michael** (Oct. 2024). Big *G*. *Journal of Political Economy*. 2024. 132, pp. 3260–3297.

Dupor, **Bill** and **Guerrero**, **Rodrigo** (Dec. 2017). Local and aggregate fiscal policy multipliers. *Journal of Monetary Economics*. 2017. 92, pp. 16–30.

Fishback, **Price** and **Cullen**, **Joseph A.** (2013). Second World War spending and local economic activity in US counties, 1939-4458. *Economic History Review*. 2013. 66, pp. 975–992.

Fisher, **Jonas** and **Peters**, **Ryan** (May 2010). Using Stock Returns to Identify Government Spending Shocks. *The Economic Journal*. 2010. 120, pp. 414–436.

Government Transactions (Sept. 2005). Tech. rep. Bureau of Economic Analysis.

Hall, **Robert** (1990). Invariance Properties of Solow's Productivity Residual. *NBER working paper*. 1990. 3034.

Hall, **Robert E.** (2009). By How Much Does GDP Rise If the Government Buys More Output? *Brookings Papers on Economic Activity*. 2009, pp. 183–221.

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

Jordà, **Òscar** (Feb. 2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*. 2005. 95, pp. 161–182.

Leduc, **Sylvain** and **Wilson**, **Daniel** (2013). Roads to Prosperity or Bridges to Nowhere? Theory and Evidence on the Impact of Public Infrastructure Investment. *NBER Macroeconomics Annual*. 2013. 27, pp. 89–142.

Leeper, **Eric**, **Walker**, **Todd**, and **Yang**, **Shu-Chun Susan** (2013). Fiscal Foresight and Information Flows. *Econometrica*. 2013. 81, pp. 1115–1145.

Leeper, **Eric M.**, **Walker**, **Todd B.**, and **Yang**, **Shu-Chun S.** (Nov. 2010). Government investment and fiscal stimulus. *Journal of Monetary Economics*. 2010. 57, pp. 1000–1012.

Mertens, **Karel** and **Ravn**, **Morten O.** (May 2010). Measuring the Impact of Fiscal Policy in the Face of Anticipation: A Structural VAR Approach. *The Economic Journal*. 2010. 120, pp. 393–413.

Montiel Olea, **José Luis** and **Pflueger**, **Carolin** (July 2013). A Robust Test for Weak Instruments. *Journal of Business & Economic Statistics*. 2013. 31, pp. 358–369.

Montiel Olea, **José Luis** and **Plagborg-Møller**, **Mikkel** (2021). Local Projection Inference Is Simpler and More Robust Than You Think. *Econometrica*. 2021. 89, pp. 1789–1823.

Nakamura, **Emi** and **Steinsson**, **Jón** (Mar. 2014). Fiscal Stimulus in a Monetary Union: Evidence from US Regions. *American Economic Review*. 2014. 104, pp. 753–792.

Nekarda, **Christopher** and **Ramey**, **Valerie** (Jan. 2011). Industry Evidence on the Effects of Government Spending. *American Economic Journal: Macroeconomics*. 2011. 3, pp. 36–59.

Perotti, **Roberto** (Jan. 2007). In Search of the Transmission Mechanism of Fiscal Policy [with Comments and Discussion]. *NBER Macroeconomics Annual*. 2007. 22, pp. 169–249.

— (May 2014). Defense Government Spending Is Contractionary, Civilian Government Spending Is Expansionary. *NBER Working Paper*. 2014, w20179.

Plagborg-Møller, **Mikkel** and **Wolf**, **Christian K.** (2021). Local Projections and VARs Estimate the Same Impulse Responses. *Econometrica*. 2021. 89, pp. 955–980.

Ramey, **Valerie** (June 1989). Inventories as Factors of Production and Economic Fluctuations. *American Economic Review*. 1989. 79, pp. 338–354.

— (Feb. 2011). Identifying Government Spending Shocks: It's all in the Timing*. *The Quarterly Journal of Economics*. 2011. 126, pp. 1–50.

— (2013). Government Spending and Private Activity. en. In: *Fiscal Policy after the Financial Crisis, edited by Alberto Aleina and Francesco Giavazzi*. University of Chicago Press, pp. 19–62. url: https://doi.org/10.7208/9780226018584-003.

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

— (2021). The Macroeconomic Consequences of Infrastructure Investment. *Economic Analysis and Infrastructure Investment. Chicago: University of Chicago Press*. 2021. 28215.

Ramey, **Valerie** and **Shapiro**, **Matthew** (1998). Costly capital reallocation and the effects of government spending. *Carnegie-Rochester Conference on Public Policy*. 1998. 48, 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*. 2018. 126, pp. 850–901.

Romer, **Christina D** and **Romer**, **David H** (June 2010). The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks. *American Economic Review*. 2010. 100, pp. 763–801.

Schick, **Allen** (1990). *The capacity to budget*. Washington, D.C. : Lanham, MD: Urban Institute Press ; Distributed by University Press of America. isbn: 9780877664383 9780877664390.

Stock, **James H.** and **Watson**, **Mark W.** (May 2018). Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *The Economic Journal*. 2018. 128, pp. 917–948.

Woodford, **Michael** (Jan. 2011). Simple Analytics of the Government Expenditure Multiplier. *American Economic Journal: Macroeconomics*. 2011. 3, pp. 1–35.