Title
Understanding the Effects of Fiscal Policy: Measurement, Mechanisms, and Lessons from History

Author
Brunet, Gillian

Publication Date
2017

Peer reviewed|Thesis/dissertation

Understanding the Effects of Fiscal Policy:
Measurement, Mechanisms, and Lessons from History

by

Gillian Michele Brunet

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Christina D. Romer, Chair
Professor Barry Eichengreen
Professor Yuriy Gorodnichenko
Professor Noam Yuchtman

Summer 2017

Understanding the Effects of Fiscal Policy:
Measurement, Mechanisms, and Lessons from History

Abstract

Understanding the Effects of Fiscal Policy:
Measurement, Mechanisms, and Lessons from History

by

Gillian Michele Brunet

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Christina D. Romer, Chair

A key question in macroeconomics is the government's ability to stimulate economic activity through expansionary fiscal policy. How much economic activity results when the government increases spending by one dollar, and how does the economic and institutional context affect the answer to that question? This dissertation uses a variety of empirical techniques to explore aspects of this question using historical data on U.S. military spending.

In chapter one I use state-level variation in war production spending to measure the fiscal multiplier during World War II, and examine how features of the wartime economy influenced the size of the fiscal multiplier. Chapter two focuses on how the measurement of government spending influences the estimated size of the multiplier. I introduce a new time series measure of aggregate defense spending. In chapter three I return to World War II, but this time examine the effects of wartime military spending on the post-war economy, establishing causal evidence for its role in driving the immediate post-war boom.

In chapter one I use war production spending to quantify the idiosyncratic factors affecting estimates of the fiscal multiplier during World War II. World War II is often viewed as a quintessential example of government spending stimulating the economy, and is interesting both because it was such a significant economic event and because it strongly influences estimates of the multiplier whenever it is included in the sample. Newly digitized war supply contract data allow me to construct state-level panel data on U.S. spending for 1940–45 and examine state-level outcomes. Using state-level variation I estimate a relative multiplier of 0.25 to 0.3, depending on the estimation approach. This implies an aggregate multiplier of roughly 0.3 to 0.4 given wartime economic conditions. I find small employment effects: an additional job-year is associated with \$165,000–\$255,000 of spending (in 2015 dollars), also depending on the estimation approach. I also find evidence that the effects of stimulus were systematically larger in states that had lower employment levels pre-war. To explain why the stimulative effects of war spending were so small, I look for guidance from the historical narrative. I show that unique features of the wartime economy significantly reduced the stimulative impact of wartime spending. Conversion from civilian manufacturing to war

production reduced the initial stimulus from war production. At least 75 percent of the income generated by war spending went into increased saving and income taxes, implying that the add-on effects from increased consumption were minimal in the short run.

Chapter two focuses on how the measurement of government spending influences the estimated fiscal multiplier. Economists have previously focused on measuring shocks to expectations rather than the measurement of government spending itself. My approach is driven by the observation that government spending is a long and complex process. Specifically, I introduce an alternative measure of government spending, called *budget authority*, which uses authorizations to measure the government's commitment to spend. Budget authority is established annually as part of the congressional budget process, and is readily available from 1976 onward. I use historical budget publications to construct defense budget authority for 1938 to 1975, extending the available data backwards by several crucial decades.

Using annualized data (for purposes of comparison) to estimate the aggregate fiscal multiplier using shocks to defense spending, budget authority produces similar point estimates to the traditional NIPA measure, but much more precisely estimated. Budget authority is conceptually different from the best-known measure of shocks to anticipated defense spending, Ramey's narrative measure, particularly in how it measures shocks to expectations and how it treats uncertainty. Budget authority implies an aggregate fiscal multiplier of 0.8, while Ramey's narrative measure implies a much smaller fiscal multiplier, around 0.1. Budget authority shows consumption responses to spending more clearly than other available measures, and also picks up strong investment responses over a one-year time horizon. Ramey's narrative measure shows significant investment responses over all time horizons up to three years. While shocks to all three measures predict strong responses in total government spending, it appears that both budget authority and Ramey's measure understate the response of government spending due to timing differences between those measures and NIPA. The definition of spending mostly closely aligned to national accounting is subtly different from the definition that is most relevant for measuring the stimulative effect of government spending. Thus using the NIPA definition of spending creates a downward bias in measuring the fiscal multiplier. A fourth measure of spending, budget outlays, allows me to estimate a lower bound for the magnitude of this bias. When this bias is corrected, budget authority implies an aggregate fiscal multiplier of 1.3 to 1.4, and potentially as large as 1.4 to 1.6.

Chapter three examines the influence of World War II spending in the U.S. on household consumption and savings in the immediate post-war years (1946–1949). Chapter three uses geographic variation in war spending to measure the effects of World War II spending on household consumption and savings behavior after the war ended and rationing was relaxed. I find that compared to households in locations receiving less war spending, similar households in locations which received more war spending were significantly more likely to purchase both cars and houses in the immediate post-war years. These households also had higher liquid asset holdings and, conversely, higher total debt. With the exception of debt, all of these effects were stronger for households headed by an individual age 45–64, which was the age cohort most likely to have worked in war production.

We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.
– T.S. Eliot, “Little Gidding”

Contents

List of Figures

List of Tables

Acknowledgments

I did not make—and could not have made—this journey on my own.

First, I am grateful to Christina Romer and David Romer for tremendous guidance and support. Their insight, intellectual discipline, and knack for asking essential questions have challenged me and helped me grow, while their kindness, humor, and good sense set an example I can only hope to emulate. I am also deeply indebted to J. Bradford DeLong, Barry Eichengreen, Yuriy Gorodnichenko, Martha Olney, and Noam Yuchtman for their guidance. I am a far better economist thanks to all of them.

I am equally thankful for the classmates who have been wonderful companions throughout graduate school. From thoughtful research critiques to steadfast support to delightful adventures, they have enriched my life personally and professionally. Their friendship, intellect, and humor have been invaluable on this journey, from math camp through the job market and beyond. I would particularly like to acknowledge Carola Binder, Fenella Carpena, Aluma Dembo, Amy Filipek, Sandile Hlatshwayo, Erik Johnson, and Dmitri Koustas.

My friends Rachel Turtledove and Jason Katayanagi welcomed me into their family and shared so much time with me during my years in the East Bay. It has been a privilege to meet their daughters Cordelia and Phoebe and watch them grow from sleepy newborns into small people. I will miss all four of them more than I can say.

I am also grateful to people far outside of the Bay Area. My gratitude for their support is by no means diminished by geographic distance.

My parents, Anne Pautler and Jim Brunet, have given me their unconditional love and support over the past 30 years. No parents could be more dedicated, and their love has been a foundation for my life. I am also grateful for the love of my extended family on all sides—my aunts, uncles, and cousins—and for my family-by-choice. From the friends I've known since childhood (most particularly Caroline Brumfield and Olga Radovsky) to my many Smithies (too numerous to name, but you know who you are) to my D.C. friends, I am lucky to have so many wonderful, caring people in my life.

I am also indebted to the federal budget tax group at the Center on Budget and Policy Priorities, especially Richard Kogan. They taught me to understand and appreciate the policy process and gave me a healthy respect for institutional history and context. Above all, their tireless work continues to inspire me. I would see the applications of economics far less clearly without the benefit of my time among them.

Finally, I am grateful to the mentors who shaped me before I ever contemplated becoming an economist, especially Rosemary Valaire, Julie Whittaker, and Ruth Haas. In their different ways, their patience, kindness, and wisdom served as an anchor for me at various points in my life. Together, they taught me the focus, discipline, and dedication without which I would never have grown into a researcher or completed a dissertation. I owe all of them more than I can express, and I think of each of them often and with great fondness.

Introduction

The fiscal multiplier measures the effect of government spending on output in the short run. It is difficult to measure for several reasons.

First, government spending is often intentionally countercyclical, with the result that naïve attempts to measure the multiplier will generally be downward biased. To avoid this endogeneity problem, economists often focus on military spending because it is much less likely to be influenced by short-run economic fluctuations than other types of spending. Following this practice, I focus on measuring the multiplier on military spending.

Second, evidence suggests that the size of the fiscal multiplier varies significantly depending upon the macroeconomic context. For instance, it appears that the multiplier may be significantly larger when the macroeconomy is weak (and there is more "slack" in the economy) than when the economy is strong. Additionally, spending shocks are often accompanied by tax changes which can also have macroeconomic effects. In other instances fiscal shocks are accompanied by shocks to the price of oil or other essential commodities, or by formal or informal changes in institutions. All of these factors can influence estimates of the fiscal multiplier, particularly if they are not carefully accounted for. Because the size of the fiscal multiplier depends on context, there is no single "true" multiplier. Given this, it is extremely important to be aware of the economic context in which the multiplier is measured.

Finally, estimates of the fiscal multiplier are highly sensitive to assumptions about the timing of spending, because multipliers are most often measured using variation in the timing of both spending and output growth. The "correct" estimate of the fiscal multiplier depends on measuring government spending *when it affects economic activity.* In practice this is quite difficult. It is not merely a question of shifting spending backwards or forwards by a lag or lead, because the process of government spending can be quite complex and its length varies significantly depending on the producer and the good being bought.

In this dissertation I use historical data on military spending in the United States to explore how both economic context and the measurement of spending affect the estimated size of the fiscal multiplier.

World War II was the largest fiscal shock in modern American history, dwarfing all other changes to military spending. Whenever World War II is included in the sample, it has an outsized influence on estimates of the fiscal multiplier. The economic context of World War II was also extremely unusual, due to rationing, conversion of civilian industrial capacity to wartime use, and other factors relating to economic mobilization. However, relatively little attention has been given to how this economic context affected the size of the fiscal multiplier.

Chapter one uses detailed, newly-digitized data on World War II production contracts to first measure the fiscal multiplier during World War II and then understand how the size of the fiscal multiplier was influenced by features of the wartime economy. It is the first paper using both time and geographic variation to measure the fiscal multiplier during World War II. My estimate of the fiscal multiplier for this period is quite small: I find a relative multiplier of 0.25, implying an aggregate multiplier of approximately 0.3. I also find that rationing and the conversion of industrial capacity both reduced the size of the multiplier, and that together they account for around 60 percent of the difference between the "normal" relative multiplier for the post-war period and the much smaller relative multiplier I estimate for World War II.

Chapter two turns to the question of how government spending should be measured for the purpose of estimating the aggregate fiscal multiplier. Unlike most consumer spending, government spending is a long and complex process, spanning from an initial authorization for spending and subsequent contract placement until final goods are produced and delivered to the government and funds are dispersed from the Treasury. Because the government buys many different types of goods and services, ranging from office supplies to aircraft carriers, the length of this process varies significantly. In some instances, it can span years. Thus shocks to spending can look very different—concentrated or dispersed, etc.—depending on how government spending is measured. One measure of government spending is not equivalent to taking another measure of government spending and adjusting it with appropriate leads or lags. So, picking the measure of government spending affects the measurement of the fiscal multiplier. Ideally, one should pick the measure of government spending that most closely corresponds to when spending affects economic activity, but this is far easier said than done.

I introduce a new measure of government spending, called *budget authority*, which has been used in the federal budget process for decades but has received relatively little attention from economists. Budget authority measures spending authorizations, which are the government's commitment to spend. By focusing on authorizations, I implicitly assume that economic agents are extremely cautious about acting on news before uncertainty is resolved. Previous literature has focused on whether shocks to government spending may be anticipated, and consequently on measuring shocks to expectations about future spending. Implicit to that approach is the assumption that economic agents ignore uncertainty when optimizing based on their beliefs about the path of future spending. With differing assumptions about responses to uncertainty, these approaches produce different estimates of the fiscal multiplier.

In addition, spending shocks are visible in budget authority before they show up as government spending in the National Income Product Accounts (NIPA). Budget authority produces much more precise estimates of the fiscal multiplier than NIPA, though the point estimates are not significantly different. In addition, budget authority produces much clearer responses in private consumption and investment than previous measures.

Chapter three returns to World War II, this time focusing on the stimulative effects of war spending in the immediate post-war years. If the immediate stimulative effect of wartime spending was diminished by rationing and other features of the wartime economy, a natural question is whether wartime spending—which largely translated into saving in the short run—translated into increased economic activity after the war. I use household-level survey data from the immediate postwar years (covering 1946–1949) to examine the effect of war spending in a locality on household behavior. I find that households in areas which received more war spending were more likely to have purchased both cars and homes in the past year than demographically similar households in areas which received less war spending. This implies that the total stimulative effect of World War II spending was likely somewhat larger than the fiscal multiplier estimated in chapter one would imply.

Together, these findings suggest that both macroeconomic context and the measurement of government spending are first-order issues when estimating the size of the fiscal multiplier, and should not be ignored. They collectively suggest that we should be particularly cautious in interpreting any particular estimate of the multiplier as "the" fiscal multiplier, and never without considering both the macroeconomic context and the details of how government spending is measured.

Chapter 1

Stimulus on the Home Front: the State-Level Effects of WWII Spending

1.1 Introduction

If asked to name a historical episode in the United States in which government spending had a large stimulative effect, many economists would be tempted to point to World War II as a defining experiment. Certainly, the war effort caused a huge exogenous shock to government spending, and per capita GDP grew very quickly during the war. In the popular imagination, World War II appears to be a quintessential example of a large fiscal stimulus causing a large multiplier effect.

But this widely-held view of World War II is not supported by the evidence of the empirical literature on World War II. Fishback & Cullen (2013) find that total war production spending had no influence on growth in county-level retail sales between 1939 and 1948. Higgs (1992) does not directly address the economic effects of wartime spending, but challenges the notion of wartime prosperity based on how much of the wartime growth in both output and employment was directly attributable to war production and consequently did not contribute to civilian economic wellbeing.

In this paper I provide new evidence on the effects of government spending in World War II and confirm that the multiplier was indeed very small during this period. I then seek to explain why those effects were so small. I find that converting significant amounts of industrial capacity to war production significantly reduced the initial stimulative effect of wartime spending. I further show that majority of the income effects from war production appear to have gone into saving, implying that the add-on effect from consumption responses was weak at best, and arguably negligible.

I use newly digitized data on war supply contracts with private firms to measure the stimulative effects of World War II spending on the U.S. economy. This contract data was published by the Civilian Production Administration (the successor to the War Production Board) in 1946. The total value of contracts in the sample is \$183 billion (in nominal dollars), or more than half of total war spending.[1] I use these data to construct a state-level panel of U.S. defense spending.

Due to the many unusual features of the wartime economy that are difficult to account for using a time series approach, this time period is perhaps particularly ill-suited to aggregate analysis. Large federal tax increases were instituted to help finance the war, and interest rates were slightly reduced and then kept low in spite of expansionary fiscal policy and rising inflation, for the same purpose. It is also difficult to argue that less-quantifiable aspects of the war such as patriotism had no economic effects, even if identifying those effects would be difficult. But with panel data, one can estimate the effects of government spending using within-panel variation, controlling for time fixed effects (time trend effects, with differencing) and state fixed effects (state-specific underlying trends, with differencing), making the argument for identification much stronger. These fixed effects make it possible to difference out the effects like patriotism, rationing, and the aggregate effects of wartime tax increases, which should affect all states equally.

I show that war production spending had small effects on both employment and income during World War II. These findings are broadly similar to those found in the time series literature and in Fishback & Cullen's (2013) cross-sectional comparison of pre- and post-war county-level economic outcomes. My findings imply that it was the massive scale of the war effort that enabled the U.S. economy to rally after the Great Depression, not that war production was a particularly effective form of stimulus. While wartime economic growth was exceptionally large, so was the stimulus: the U.S. spent \$340 billion on national defense over 1940–1945, or \$5.7 trillion in contemporary dollars.

The stimulative effects of war production spending on the wartime economy were quite modest. Each additional job-year is associated with the equivalent of \$165,000–\$255,000 of spending in 2015 dollars, depending on the estimation approach, a cost which is high though around the upper end of the range in the literature. Using personal income data (the closest equivalent to GDP available at the state level for the early 1940s), I estimate an open economy relative multiplier of 0.25 to 0.31 depending on the estimation approach. This implies an aggregate multiplier of roughly 0.3 to 0.4 given wartime economic conditions. For comparison, Nakamura & Steinsson (2014) estimate a relative multiplier of 1.4 on defense spending for 1969–2006 using similar methodology.

Recent contributions to the fiscal multiplier literature have focused on heterogeneity in the multiplier stemming from the state of the economy (see Auerbach & Gorodnichenko, 2012; Ramey & Zubairy, 2016). I exploit the fact that there was significant variation across space in economic conditions on the eve of World War II to see if there is evidence of such state-dependence in the effects of government spending within this period.

I show that war production spending led to significantly larger growth in employment and personal income in states whose total employment rates (including agriculture) were

[1] These contracts do not include payroll costs for the Armed Forces or other direct employees of the War and Navy Departments, nor do they include spending on facilities, electricity, or food/food processing. The vast majority of war spending not accounted for in the production contract data falls into these categories. The data also omit war production at government-owned facilities.

below the median level in 1939, compared to states with above-median employment levels. The cost per job-year is $186,000 (in 2015$) in states with below-median initial employment rates and $308,000 in states with above-median initial employment rates. The open economy relative multipliers are 0.36 and 0.17, respectively. While smaller by an order of magnitude, these estimates are consistent with the findings of Auerbach & Gorodnichenko.

To understand why fiscal stimulus from war production had such small stimulative effects, I turn to features of the wartime economy. I show that these features of the wartime macroeconomy significantly reduced the stimulative effects of war production spending. The textbook Keynesian multiplier is $1/(1-mpc)$. I find that the large-scale conversion of industrial capacity for the war effort reduced the initial impulse from war production spending, resulting in a numerator below 1. Furthermore, most of the income increases attributable to war production are accounted for by the response of saving to war production. Thus, the wartime savings boom caused the *mpc* to fall considerably during the war years, increasing the denominator relative to more normal periods. Together, these two effects produced a fiscal multiplier on wartime spending well below 1.

Because the scale of war production was so large—private-sector war production by itself equalled roughly a year of U.S. GDP at the time—large-scale crowd-out of non-war industries was inevitable. This was particularly true within manufacturing, as many existing manufacturing firms converted from producing consumer goods to producing supplies for the war, whether uniforms, radar, or tanks. Because of this conversion, non-war industrial production fell significantly even as total production soared. I show that war spending had systematically smaller effects on the economy in states where conversion accounted for a larger share of spending.

Private saving skyrocketed during World War II. The wartime savings boom can be attributed both to patriotic efforts to support and finance the war and to rationing of durable goods, which constrained consumers' choices. While saving increased across the U.S., I show that saving increased by more in states that received more war spending. Specifically, if war production spending increased by 10% of income in a given state (relative to other states), bank deposit holdings in that state grew by 1.1% of income relative to trend, and purchases of series E war bonds grew by another 0.3%. These increases may seem small, but a state's income grew by only 2.5% relative to trend when war production spending increased by 10% of income, so these savings responses are quite large given the small size of the income response. Data limitations make it difficult to measure the comprehensive state-level response of private saving to war spending across all savings vehicles, but the relative growth in deposits and E-bonds alone account for nearly 60% of the relative increase in income attributable to war spending. In addition, because federal income taxes are progressive, relative income growth is partially offset by relative increases in income taxes. Differential increases in federal income tax payments account for another 17% of the relative increase in income attributable to war spending. Together, bank deposits, E-bonds, and federal income taxes account for 75% of the relative increase in income attributable to war spending. This suggests that the add-on effects of stimulus through consumption responses were minimal during the war.

While World War II is interesting in and of itself, it is also highly relevant to how we understand the broader literature on fiscal stimulus. In recent years it has become common to measure the effectiveness of fiscal stimulus using shocks to military spending. Many other types of government spending are directly or indirectly countercyclical, meaning that spending often increases in response to weakness in the economy, leading to a downward bias (and causality problem) when estimating the fiscal multiplier. Shocks to military spending are often preferred since they are driven by geopolitical events and not the short-run performance of the domestic economy.

When World War II is included in these estimates, wartime observations are highly leveraged due to the massive scale of wartime spending shocks. Currently, there is no consensus in the literature about whether World War II should be included when estimating the fiscal multiplier.

As mentioned above, a sizable body of literature uses shocks to military spending to measure the fiscal multiplier. A number of these papers exclude World War II entirely, often due to data limitations. These include Ramey & Shapiro (1998), Fisher & Peters (2010), Auerbach & Gorodnichenko (2012; 2013), and Nakamura & Steinsson (2014). Other papers estimate the fiscal multiplier for periods both including and excluding World War II, notably Hall (2009), Barro & Redlick (2011), Ramey (2011), and Ramey & Zubairy (forthcoming). Estimates of the fiscal multiplier are somewhat smaller when World War II is included in the sample, with the exception of Ramey (2011). (For a discussion of how World War II affects the estimates in these papers, see Appendix A.) Moreover, because the variance in spending is so large during the war, estimated standard errors are always much smaller when World War II is included in the sample. However, just because estimates including World War II are in some sense more precise does not mean that they provide more accurate information about the size of the fiscal multiplier under normal conditions. In this paper I will argue that it is not appropriate to include World War II when estimating a general fiscal multiplier because the underlying economic environment is so unique and influences the results so strongly.

While there has been recent work on microeconomic experiments stemming from World War II (including Goldin & Olivetti, 2013; Goldin, 1991; Collins, 2001; and Fetter, 2016), economists seem to have largely avoided working on the U.S. macroeconomy during World War II for some time, largely because the complexities of the wartime economy made it too difficult to disentangle using aggregate data. Higgs (1992) argues broadly that wartime living standards and economic conditions were considerably less rosy in fact than in popular perception, but focuses more on consumption, rationing, and household purchasing power than on the macroeconomic effects of war spending.

Fishback & Cullen (2013) examine the economic effects of war production spending on retail sales using a county-level cross-sectional comparison of pre- and post-war economic outcomes. They find no effect of war spending on the change in county retail sales, even with a wide variety of controls. At some level this is not surprising. Spillovers are almost certainly much larger at the county level, so the relative output effects of a spending shock should be considerably smaller than at the state level. Moreover, Fishback & Cullen cannot exploit timing variation in spending. Given these differences in methodology, their findings are broadly consistent with my finding of a small (but non-zero) fiscal multiplier for the World War II period.

This paper is structured as follows. Section 1.2 provides a brief overview of the historical context of World War II and U.S. war production. Section 1.3 describes the data, while Section 1.4 discusses both identifying assumptions and interpretation of estimates from panel data. Section 1.5 presents baseline specifications and results, and discusses several robustness checks. The evidence for state-dependence in the multiplier is presented in section 1.6. Sections 1.7 and 1.8 address conversion and the influence of the wartime savings boom, respectively. Section 1.9 presents an accounting exercise to understand how the mechanisms and estimates presented in the previous sections fit together. Section 1.10 concludes.

1.2 Historical Background

Most historians date the start of World War II to September 1939, when Germany invaded Poland, although Japan had been at war with China since 1937. Even after the Nazi invasion of Poland, it was some months before either the U.S. government or public opinion viewed the war as a serious conflict. The U.S. defense program began in earnest over May and June of 1940, starting the so-called "defense period" during which civilian and military officials began actively preparing for conflict without any certainty of war. But public opinion was still against American intervention. The defense program grew gradually through the fall of 1941, but there was still significant uncertainty as to whether the U.S. would enter the war and, if so, what its role would be. This was the case even within the higher ranks of the government and the military, complicating planning efforts. The United States did not enter the war until December 1941, after the Japanese bombing of Pearl Harbor.

World War II was fought to unconditional surrender in both Europe and Asia. The war in Europe ended with German surrender on May 8, 1945. The war in Asia ended on August 15, 1945, when Japan surrendered after the United States dropped atomic bombs on Hiroshima and Nagasaki.

During the defense period—roughly June 1940 to December 1941—the U.S. began war production. Compared to the 1930s, spending during the defense period was very large, though it pales in comparison to spending after U.S. entrance into the war. Many defense contracts during this period were for the development of arms and munitions for mass production. These were demonstration orders for small quantities, and the goal was producing detailed, tested designs for production. While these efforts proved essential when large-scale production began in earnest, they did not require large-scale production capacity. During this period the U.S. also began producing war matériel for the United Kingdom, the Soviet Union, and France.

Even for those familiar with the history, it can be easy to forget that until Pearl Harbor, American entrance into the war was not a foregone conclusion. Throughout 1940 and most of 1941 it was far from clear—even to many of the planners in Washington—that the U.S.

would engage in "a shooting war" abroad rather than just supplying its future allies with arms and mounting a preemptive defense at home.

Consequently it was nearly impossible for planners (1) to obtain accurate estimates of production requirements from a military leadership with no concrete sense of the kind of war it would be fighting, and (2) to convince Congress and the successive agencies responsible for war planning to both make the demands upon stocks of raw materials and place production orders of raw materials on the scale that would become necessary for all-out war.

When war came after Pearl Harbor, it was impossible to meet war needs only through expansions of production capacity. Building new facilities takes considerable time, and while appropriations were made and contracts approved during 1940 and 1941, the scale was very small relative to the eventual scale of production needed at the height of the war. Conversion of pre-existing manufacturing facilities was necessary for ramping up production as quickly as needed in early 1942.

The economic consequences of conversion are discussed later in this paper, but it is useful to describe what it entailed as context for understanding the wartime economic environment. Conversion was in many circumstances an all-or-nothing proposition:

> [T]here could be no such thing as partial conversion in the automobile industry. Charles E. Wilson, president of General Motors, told [Donald Nelson], as other motor magnates were to tell [him] later, that you do not "partly convert" a production line. You do it all the way or not at all. Mr. Wilson said: "When you convert one of our factories, you move everything out and start with blank space. Out of a long row of intricate machines on the production line a certain percentage may be used in the manufacture of a war product, depending on what the war product is. But the production line will necessarily consist mainly of new, special-purpose machines along with any of the old machines that can be rebuilt for the new manufacturing process."[2]

For many military goods (as opposed to civilian goods like blankets and gloves needed for the war), firms could not choose a specific balance between war and non-war production. Most firms producing arms and munitions faced a binary choice: guns or butter.

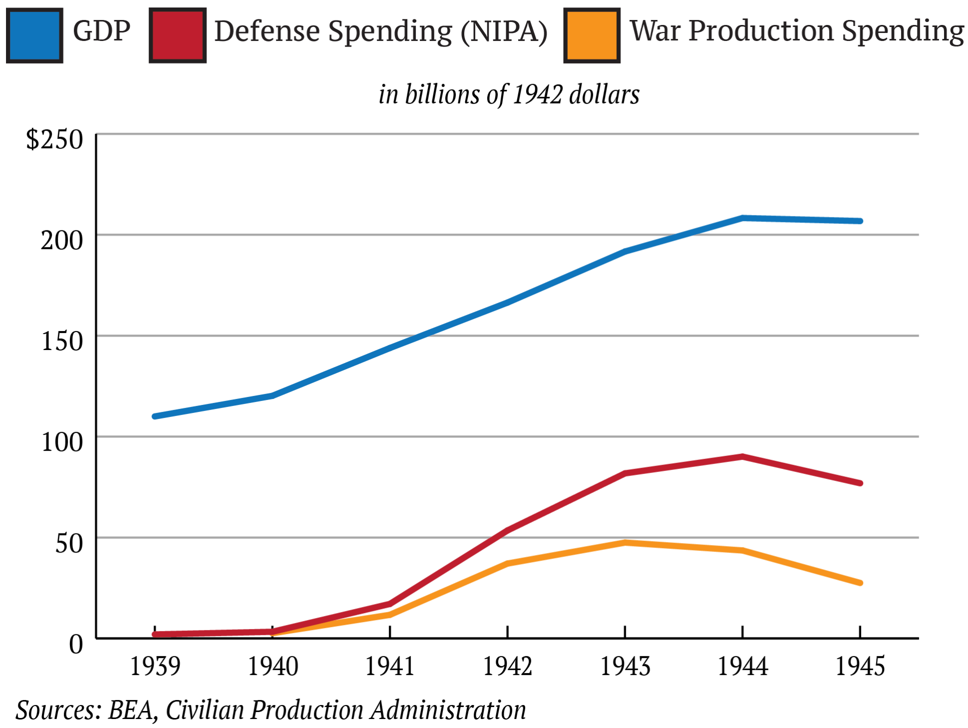
In aggregate terms, the U.S. spent $340 billion on national defense over 1940–1945, or $5.7 trillion in contemporary dollars[3]—a massive increase over the $8 billion total U.S. defense spending over the entirety of the 1930s.[4] Nominal GDP grew from $103 billion in 1940 to $228 billion in 1945. Real GDP grew by more than 75% over the same period, at an annual average rate of 11.8%. (See Figure 1.1.) Total non-farm employment grew by 8 million over the same period.

[2]Nelson, p. 218.

[3]Bureau of Economic Analysis, tables 1.1.5 and 1.1.6.

[4]*Budget of the U.S. Government for the Fiscal Year Ending June 30, 1940*, table 2, p. VII.

Figure 1.1: GDP Growth and War Spending During WWII (United States)



1.3 Data

I use a panel approach to study the macroeconomic effects of World War II spending. This structure allows me to use time fixed effects, which provide first-order controls for wartime events (e.g. patriotism affecting productivity around major events in the war), and makes it easier to include controls for other factors affecting the economy. I am also able to control for state fixed effects, which implicitly become state-specific time trends since I use difference specifications.

Data Construction

I build a panel of defense spending using listings of military contracts with private firms. My data comes from an ex-post summary tabulation of war supply contracts published by the Civilian Production Administration (the successor to the War Production Board). Contracts are listed by the location of main production (establishment level).[5] See Figure 1.2 for an example of contract listings from the data. The data includes all contracts for $50,000 or more awarded between June 1940 and September 1945. Contracts include airplanes and victory ships, but also intermediate goods (e.g. propellers and gun fittings), raw materials

[5]Many larger firms appear in contract listings for numerous locations, though in some cases it is unclear whether there are multiple firms of the same name or one firm with multiple facilities.

Figure 1.2: Example of Civilian Production Administration Contract Data

NAME OF MANUFACTURER / LOCATION OF MANUFACTURE / PRODUCT AND CONTRACT NUMBER		AGENCY	FOOTNOTE	VALUE THOUSANDS DOLLARS	AWARD DATE MO.	AWARD DATE YR.	COMPLETION DATE MO.	COMPLETION DATE YR.
INSECT POWDER	TPS77605L	T		130	1	45	3	45
DDT INSECTICIDE	28021 QM29451	A		471	1	45	4	45
INSECTICIDE	28021 QM26736	A		170	1	45	1	45
DDT INSECT POWDER	28021 QM30839	A		93	2	45	9	45
INSECTICIDE POWDER	28021 QM33715	A		282	3	45	9	45
INSECTICIDE	28021 QM33751	A		314	3	45	7	45
LIQUID INSECTICIDE	28021 QM41648	A		135	6	45	12	45
INSECTICIDE POWDER	28021 QM40418	A		90	6	45	12	45
DDT INSECT POWDER	28021 QM38508	A		72	7	45	12	45
				4852				
MCCORMICK B B SONS								
JACKSONVILLE FLA								
PRECUT BARRACKS	8123ENG 45	A		244	6	44	10	44
				244				
MCCORMICK BAXTER CREBOTING CO								
PORTLAND OREG								
CREOSOTING WOOD PILES	406XSY12045	N		1367	6	44	4	45
LUMBER	406XSX13307	N		163	6	44	9	44
LUMBER	406X8X13489	N		55	8	44	10	44
LUMBER	406XSX13727	N		141	8	44	10	44
LUMBER	406XSX13719	N		141	8	44	11	44
WOOD PILING	130X8X21070	N		150	7	45	12	45
				2017				
MCCORMICK BAXTER CREOSOTING CO								
STOCKTON CAL								
RAILROAD CROSS TIES	35052ENG 477	A		71	8	44	11	44
RAILROAD CROSSTIES	35052ENG 526	A		73	9	44	12	44
				144				
MCCORMICK BROS CO								
ALBANY IND								
METAL SLEEVES	ORD 7592	N	3	1950	11	44	7	45
METAL SLEEVES	ORD 9378	N	3	1322	7	45	1	46
				3272				
MCCOY COUCH FURNITURE MFG CO								
BENTON ARK								
LOCKERS	586 QM 272	A		302	2	43	8	43
LOCKERS	586 QM 387	A		116	4	43	8	43
				418				

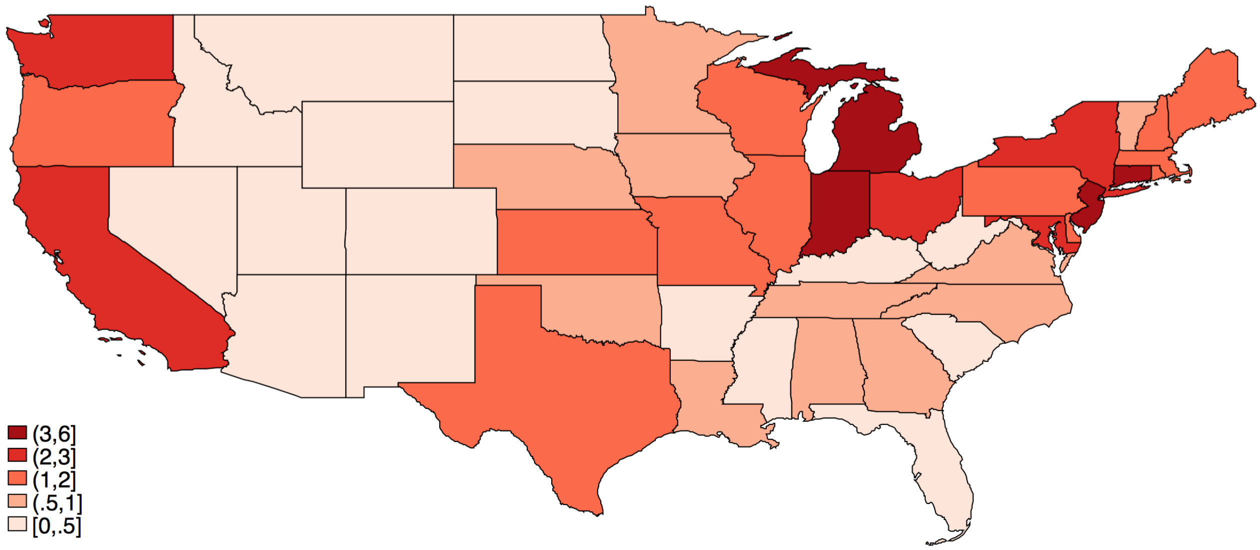
(e.g. aluminum and leather), and smaller items (e.g. mattresses, gloves, insect repellent, and toilet paper). In total, the data include over 190,000 contracts, worth $183 billion—a huge sum given that U.S. GDP was only $103 billion in 1940. Most contracts were relatively small, though numerous contracts are recorded for many firm/location pairs. More than 90% of contracts were for less than $1 million, fewer than 200 contracts were for more than $100 million. The largest contract was for Boeing to build B-17 bomber planes in Seattle for $669 million.

Figure 1.3 shows the distribution of war supply contract spending by state. It is clear that war production spending was concentrated in the industrial northeast, in the midwest, and along the Pacific coast. Appendix B provides an in-depth discussion of how the contract data was cleaned and prepared.

While in many ways very similar to the modern data on military contracts used by Nakamura & Steinsson (2014),[6] my World War II contract data has one distinct advantage: precise timing. Instead of an approximate date associated with a paperwork filing for each contract (an issue discussed in their paper), each contract listing in the World War II data

[6]Nakamura & Steinsson's contract data begins in 1966.

Figure 1.3: Total Per Capita War Supply Spending (dollars in thousands)



includes the month and year of both contract award and contract completion. The median contract length in my sample is 7 months, with a range of 1 to 83 months.[7] 95% of contracts lasted 18 or fewer months.

To construct a panel of war supply spending, I first take the total value of each contract and distribute it uniformly over all of the months in which the contract was open (inclusive of the award and completion months). I then sum spending by state and quarter and/or year.[8] Once again, state-level data are the ideal unit of observation because spillovers are likely to become significantly larger as the geographic unit of observation shrinks.

I focus on two outcome measures: employment and personal income. Employment data was collected and published monthly at the state level in the Current Employment Statistics (CES), administered by the Bureau of Labor Statistics (BLS). BLS had been collecting this data and slowly improving its quality for several decades by the start of World War II,[9] and the historical narrative indicates that BLS was considered a particularly trustworthy data source by policymakers at the time. While not as sophisticated as modern data collection,

[7]Many of the longest contracts are not particularly large, and were for things like gun forgings, ball bearings, and storage space rentals that were simply needed in predictable quantities over long periods of time. The large long-lasting contracts are generally for shipbuilding, and in one case the B-35 bomber, an airplane prototype that was never put into regular production. For some of the largest firms, such as General Motors, contracts are listed by the responsible division within the firm.

[8]For a discussion of different possible assumptions about the timing of spending and how they affect my results, see Appendix B.2.

[9]Published, (theoretically) consistent state-level employment data series start in 1939. However, the CES program began in 1915, and BLS began entering data-sharing agreements with state agencies in 1916. State-level employment fluctuations are systematically reported in BLS publications going back to at least the early 1930s.

employment seems to be the highest quality state-level outcome data for the 1940s.[10] Because employment data is noisy and because using monthly data may exacerbate timing problems in spending data, I will generally use quarterly employment data (employment totals averaged across each quarter).

My other outcome variable is personal income, which was estimated annually at the state level by the Bureau of Economic Analysis (BEA). This is the measure closest to GDP available for the 1940s. At the national level, changes in personal income track changes to output very closely over the relevant period. However, personal income differs from GDP in several ways, most crucially for this paper in that personal income does not include profits retained and reinvested by firms. This is unfortunate, but unavoidable.

Subcontracting

One concern with using contract data to get the locations of war production is that there might be systematic measurement error caused by subcontracting. Several types of subcontracting could theoretically cause these measurement errors.

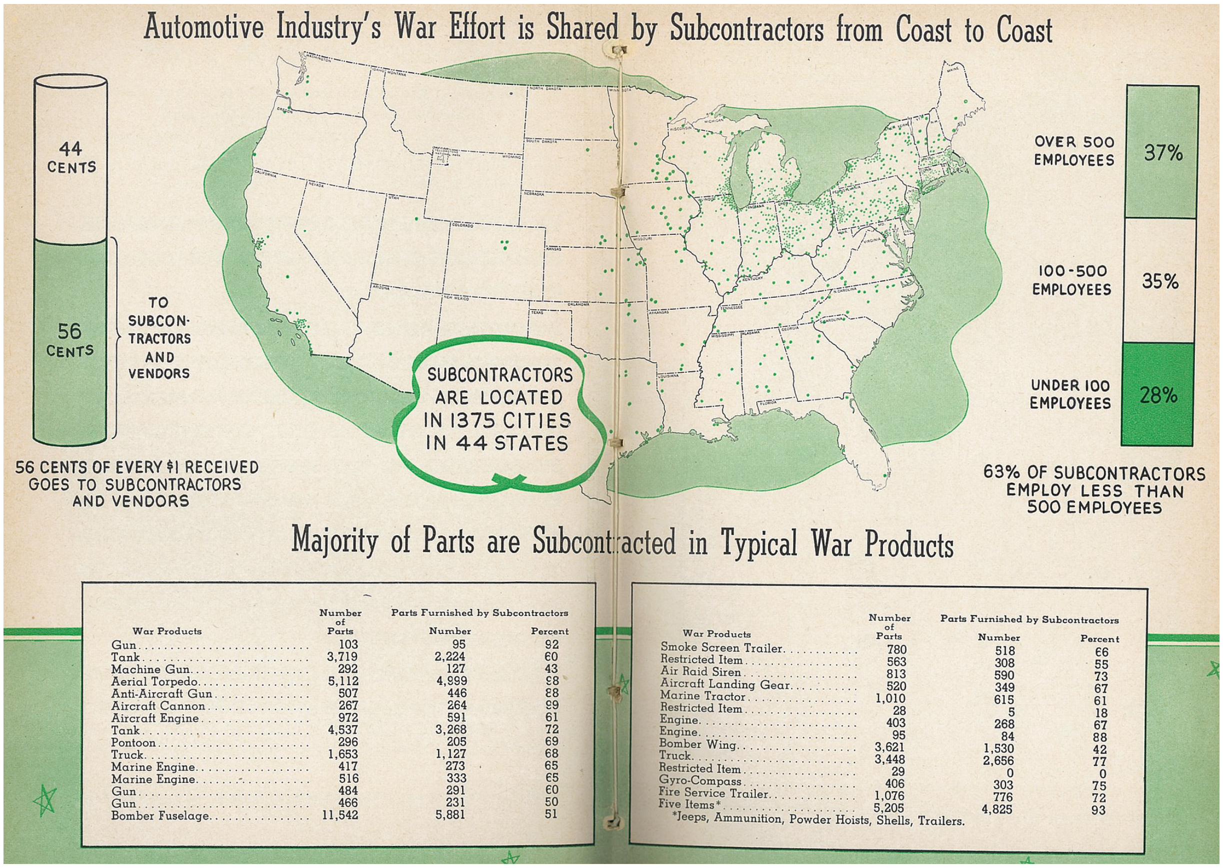
First, a primary contractor might subcontract part or all of their production to another firm, so that the subcontractor produces (some portion of) the product. The World War II contract data accounts for this kind of subcontracting: contracts placed indirectly through other firms (rather than directly by military procurement agencies) are included with a footnote indicating that they were placed in this way. Of the more than 191,000 contracts in the data, roughly 12,000 fall into this category. Many of these contracts appear to be for fuel, cloth, and hardware.

Second, when producing more complex goods such as tanks, guns, and ships, firms often purchase intermediate inputs from other firms and outsource the manufacture of simpler components. While this could potentially cause systematic measurement errors in contract data, in practice these supply chains do not appear to cause systematic errors in state-level data for two reasons.

Subcontracting was channeled directly through primary contractors for ease of administration. It seems that many firms relied on preexisting supply chains, which tended to be geographically proximate to the industries to which they were connected. Figure 1.4 shows a map of wartime subcontracters for the converted auto industry, taken from a 1943 industry trade publication. It is remarkable how strongly the geographic distribution of subcontractors mirrors that of the automobile industry as a whole. For that reason, this kind of subcontracting is unlikely to introduce systematic measurement error in the contract data.

[10] While some employment data was also recorded at the city level at relatively high frequency, the set of cities for which data was collected during this period was considerably smaller than the number of states. Moreover, the relevant city-level data was generally confined to city boundaries, rather than metropolitan areas. There is also reason to believe that spillovers are larger with finer geographic units, which would confound the analysis. Therefore I confine my analysis to state-level outcomes.

Figure 1.4: Subcontracting in the Converted Automobile Industry, 1943



Source: Automobile Facts and Figures *for 1943*

Moreover, where new supply chain relationships were established, policies of the War Production Board strongly incentivized prime contractors to subcontract with local firms. As War Production Board Chairman Donald Nelson explained in his 1946 memoir:

> We were able also to induce large manufacturers to spread the work to smaller firms... letting [large firms] overload somewhat on orders and then refusing permission to expand their facilities when the facilities of small plants were available in the vicinity of the larger plants.[11]

While fortuitous for my purposes, the logic behind this policy was sound given the goals of war production planners. Given the number of small firms with preexisting industrial capacity, utilizing those small firms was essential to American efforts. This was especially true given the complexity of many essential war products: smaller firms usually produced

[11]Nelson, p. 275.

the simpler components, freeing the large firms to specialize in the more complex components that required more extensive equipment and knowledge. The Herculean administrative task of directly managing so many small contracts (and the related materials allocations) would have quickly overwhelmed the already strained administrative capacities of the War Production Board and military procurement agencies. Subcontracting was the clear solution to this problem.

Why did the War Production Board chose to incentivize *local* subcontracting? Planners were concerned with minimizing the strain on shipping capacity, even within the U.S. While shortages of freight capacity were not as acute as shortages of critical raw materials, they were linked in the sense that expansions of shipping capacity required the same raw materials that were so critical for war production. Thus expansions of domestic shipping capacity were carefully calibrated. Local subcontracting meant less shipping within the production process, freeing up raw materials.

The biggest concern about locations in the data is flows of raw materials. There is good reason to believe that the geographic distribution of raw materials production looked very different from the geographic distribution of war production. This suggests that contract data may systematically misrepresent economic activity associated with war production because the production of raw materials are not accounted for.

The procurement agencies did buy some raw materials directly, and these contracts are reflected in the data. However, contracts for materials account for only a small fraction of the materials used for the war effort. Data on manufacturers' shipments shows \$50 billion in "iron and steel and their products" and \$9 billion in textile-mill products for war use.[12] The contract data contain \$3 billion for iron and steel products[13] and nearly \$6 billion for textiles.[14] Thus the contract data account for roughly 6% of war-related manufacturers' shipments for iron and steel and their products, and roughly 63% of war-related textile shipments. This suggests that economic activity relating to war production is systematically undercounted in places where metals were produced and over-counted in the places where final goods were produced.

But while this would seem to be a problem conceptually, there is no correlation between either a state's share of employment in mining or the change in a state's share of employment in mining and the residuals of the baseline regressions relating war spending to employment and income, as one might expect to see if measurement error from the production of raw

[12]Industrial College of the Armed Forces textbook from 1947, p. 289. The earliest version of the SIC, which was in use during WWII, separated metal products into ferrous (iron-containing) and non-ferrous metals and their products rather than separating primary and fabricated metal industries. For this reason, data on primary metals is not available.

[13]This includes all products with descriptions including the terms "iron" or "steel," plus selected products that appear in listings for either Bethlehem Steel or Carnegie Steel that are almost certainly made from those materials, such as forgings, armor, shapes (the molds used to make steel plates or sheets), axles, and rails. This may still be an undercount, but it is clear that the vast majority of iron and steel and their products used for war production were purchased directly by firms and are not accounted for in the contract data.

[14]Textiles are defined as products whose descriptions include any of the terms "cloth," "wool," "cotton," and "duck" (a type of cotton fabric frequently appearing in the data).

materials were systematically biasing the results. It may be that differencing and state fixed effects crudely correct for this measurement error, or that the scale of the effect is too small to pick up. The evidence is unclear, but if measurement error from raw materials' production is systematically biasing the results, the scale of that bias does not appear to be large.

1.4 Identification & Interpretation with Panel Data

Identification

The historical record clearly indicates that both the timing and amount of total war spending was determined by factors unrelated to U.S. economic conditions. Beyond the obvious facts of U.S. entry into the war after Pearl Harbor and the cessation of war after the unconditional surrender of the Axis powers in Europe and the Japanese Empire in Asia, the historical narrative is quite clear about the factors driving the timing of war production specifically. (For a discussion of those factors, see Appendix B.)

If variation in the aggregate timing of spending is taken as exogenous, the identifying assumptions are that the distribution of spending across states (1) was not driven by differences in economic conditions, and (2) was independent of other drivers of variation in government spending shocks. At first glance, these may appear to be very strong assumptions, but an examination of the institutional setting, narrative evidence, and raw correlations between total state-level spending and pre-war economic conditions all suggest that it is largely appropriate.

First, war production contracts were made directly by military procurement agencies. Congress and elected officials had minimal influence on contract placements, shutting down usual political economy channels for the distribution of spending. A placebo test discussed in Section 1.5 shows that state-level war production spending was *not* predictive of changes in government employment, as one would expect if war production spending were strongly correlated with other shocks to government spending.

In contract placement, the military's bias was toward large, established firms known for their reliability. Moreover, the military was not much concerned how war production might affect the civilian economy. Civilian production planners expressed frustration over military attitudes that civilians should tighten their belts and be prepared to make any sacrifices the military deemed necessary (with a very low valuation on civilian needs).[15]

While civilian planners at both the War Production Board and War Manpower Commission did urge military procurement agencies to place contracts in locations with more available labor, in practice the effects of these efforts were negligible. A retrospective on the war effort published in 1946 by the Bureau of the Budget describes efforts to distribute contracts to areas with available labor supply in 1942 and 1943:

[15] See Nelson, pp. 109–112.

> Previous efforts to reduce the demand for labor in areas of shortage had met with only limited success. In October 1942 the War Production Board, by its Directive No. 2, instructed the procurement services to consider the adequacy of labor supply in particular areas as a factor in contract placement; this directive was amended and strengthened in the fall of 1943. The impact of this directive on actual procurement practice, however, was not large. Among the reasons were the traditional independence of the procurement services, preoccupation with price and delivery considerations, reluctance to give up customary sources of supply, reluctance to give up a facility which may then be taken over by a competing procurement branch, and the absence of any continuous policing by Army Service Forces headquarters of compliance by contracting officers with Directive 2.[16]

The narrative record indicates that other attempts to place contracts according to local economic conditions were similarly ineffectual, due to a combination of inter-agency power struggles, competing policy objectives, and opposition from both industry and organized labor.

The correlation between total war production spending and pre-war employment levels (combined farm and non-farm employment from 1939) is 0.11, suggesting that spending was not driven by pre-war economic conditions (and in keeping with the narrative evidence above). Unsurprisingly given the prevalence of conversion, total war production spending is more highly correlated with pre-war manufacturing employment: the correlation between a state's per capita war production spending and the fraction of its population employed in manufacturing in 1939 was 0.58.

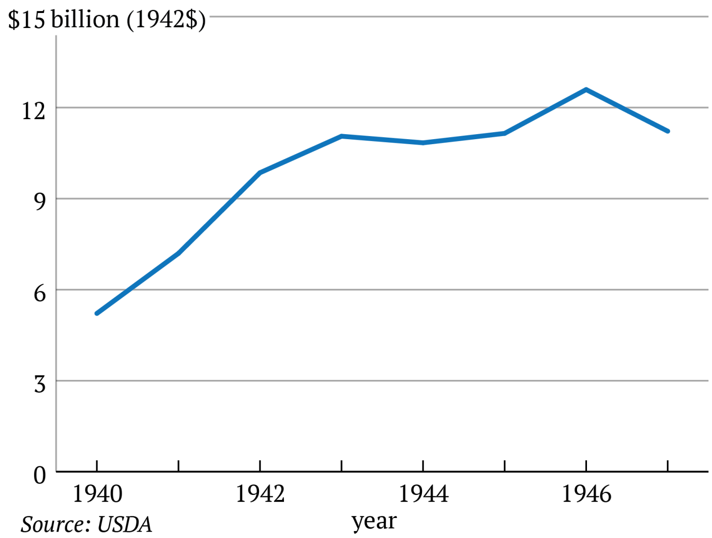
Like manufacturing employment (and likely because of manufacturing employment), pre-war (1939) personal income *levels* are correlated (0.53) with war production spending: the states with the manufacturing capacity that made them suitable for war production tended to be richer. However, when we instead consider pre-war income *growth rates*, this correlation breaks down. The correlation between state personal income growth rates over 1938–1940 and war production spending is 0.33, and for growth rates over 1937–1939 the correlation is -0.16. Since all estimates of the effects of war production spending on the economy use differenced variables, the results should not be biased by these factors.

Controlling for the Farm Sector

One major complication to identification with panel data is that states have varying exposure to agriculture. First, agriculture presents an omitted variable bias problem. Second, because agriculture was essential economic activity for the war effort, wartime constraints on the economy were much looser in the agricultural sector than in other sectors of the economy, suggesting that both underlying trends and the effects of war production spending may have been quite different in areas with significant agricultural activity.

[16]Bureau of the Budget, pp. 432–433.

Figure 1.5: Net Farm Income (United States)

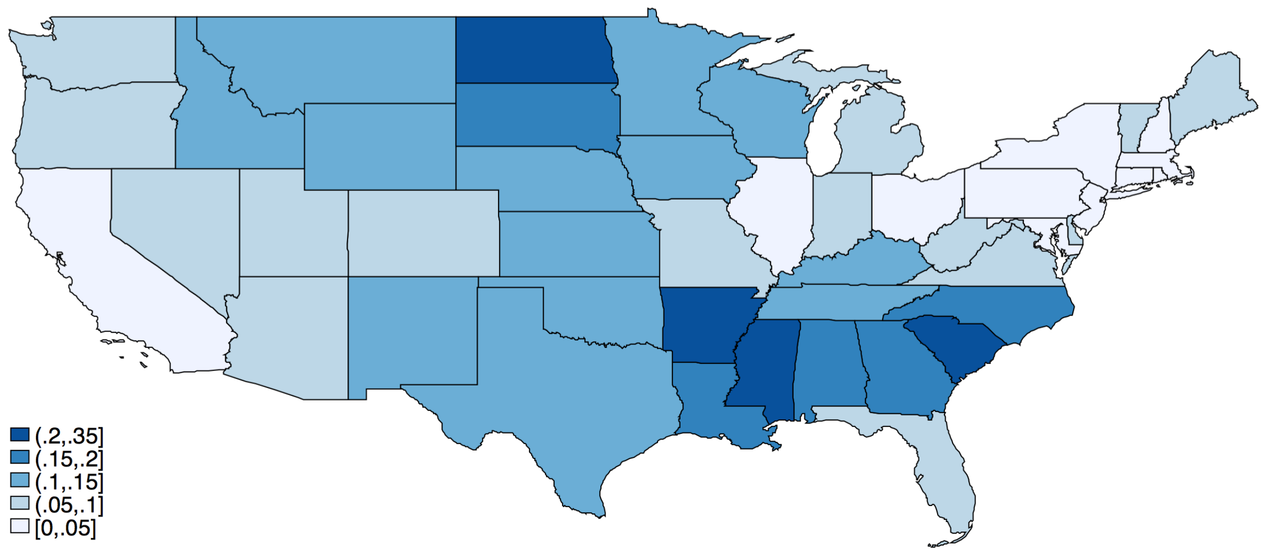


Agriculture presents an omitted variable bias problem because agricultural income grew very quickly at the same time war production was ramping up, making it important to allow for different underlying trends in agricultural states. (See Figure 1.5.) There were several good reasons for the expansion of the agricultural sector during World War II: natural rebuilding of the farm sector after the Dust Bowl, the need to provision the armed forces, food aid to European allies (where war significantly interrupted agriculture in many areas), and the need to provide food for civilians in areas under Allied control, starting with Northern Africa in 1942. The scale of increases was large: real net farm income more than doubled over 1940–43. Net farm income over 1940–45 totalled $57 billion, or roughly one third third of war supply spending over the same period. Especially given that agricultural income was concentrated in less populous states, the scale of agricultural income increases was large enough to be a confounding factor.

The war production contracts do not include food or food processing (often undertaken near agricultural areas), so the increases in farm income are not directly related to my measure of war production. To the extent that increased spending on food was a result of wartime stimulus, food production was still concentrated in certain states and regions (often very different from the states and regions where the production of war materials was concentrated) and systematically shipped across state borders, so it is still a threat to identification.

In addition, wartime rationing affected the farm sector much less intensely than it affected most other sectors of the civilian economy. Understandably, food production was deemed essential for the smooth functioning of the civilian economy, which meant strategic materials such as metals were more readily available for agricultural goods than for other types of goods. In fact, farm implements were one of the few categories of new durable goods for

Figure 1.6: Share of State Population Employed in Agriculture in 1939



which metals were allocated (most metal supplied to civilian durable good industries during the war was for repairs and replacement parts). Thus there is good reason to believe that the underlying economic trends in agricultural and non-agricultural areas were very different during World War II.

To allow for the differences between the agricultural and non-agricultural sectors, I interact the time fixed effects with the share of the state's population employed exclusively in agriculture in 1939.[17] The variation in 1939 farm employment is shown in Figure 1.6. This allows for a different underlying time trend in the agricultural sector of the economy.[18] I use agricultural employment in 1939 both to avoid endogeneity problems and because the next state-level estimates of agricultural employment were not produced until 1945.

Controlling for the farm sector does not meaningfully change the effects of war spending on employment. This is partly because all employment outcomes are estimated using non-farm employment, and partly because the aggregate size of the agricultural labor force was relatively stable during the war. But personal income includes agricultural income, and accounting for the farm sector significantly alters these results. so allowing separate time fixed effects for farm states is necessary for correctly estimating the effects of war production spending on personal income.

[17] These figures are constructed using estimates from the 1940 *Census of Agriculture*, which measured agricultural employment by state in 1939. Totals for agricultural employment are adjusted downwards to remove workers who also worked off-farm.

[18] An alternative approach is to use a dummy variable that is set to 1 for states with above-median farm employment rates in 1939 and 0 for states with below-median farm employment rates (the median is around 9%). All of the main results reported in this paper are robust to this alternative approach.

Interpreting Estimates from Panel Data

Because a panel data approach with fixed effects identifies the effects of stimulus using within variation, this approach does not allow me to directly estimate the aggregate fiscal multiplier. Instead, I estimate the relative effect of spending shocks across states, or the *open economy relative multiplier*. Nakamura & Steinsson (2014) present a theoretical model relating the open economy relative multiplier to the aggregate multiplier. To understand what my estimates imply about the aggregate fiscal multiplier, it is useful to understand what their model implies for the economic conditions prevailing during World War II.

Their model shows that with sticky prices (surely the right model for World War II given legally mandated limits on wartime price increases), the relationship between the closed economy aggregate multiplier and the open economy relative multiplier is largely a question of monetary policy. When monetary policy responds aggressively to counter fiscal shocks, the aggregate multiplier is small in relation to the open economy relative multiplier. When monetary policy is particularly accommodating (they take the case of a fixed nominal interest rate), the aggregate multiplier should be particularly large. Meanwhile, the open economy relative multiplier should be invariant to monetary policy because interest rates are fixed across a monetary union and thus differenced out across states.[19]

During World War II the first priority of the Federal Reserve System was not price stability or full employment but rather financing the war. While the government did substantially increase taxes during wartime, the vast majority of war finance came from domestic borrowing, largely in the form of war bonds. The Treasury and Federal Reserve played an active role in planning and coordinating war finance. Nominal interest rates were deliberately kept low throughout the war in order to keep costs down—resulting in a policy of low, fixed nominal rates from 1942–1947. With monetary policy diverted from pursuing price stability, the government relied on a mix of large tax increases, production-side rationing, and price controls to limit inflation.

In Nakamura & Steinsson's model, the open economy relative multiplier is moderated by rising prices of locally-produced goods following a localized government spending shock. This reduces consumption of locally-produced goods relative to consumption of goods produced in

[19] This last point does not quite hold for the U.S. in the early 1940s. Discount rates were still set by individual Federal Reserve Banks and not by the Board of Governors. Throughout 1938 and the first half of 1939, discount rates were 1% in New York and 1.5% in all other districts. The Boston Fed dropped its rates to 1% in September 1939, but other reserve banks did not follow suit until early 1942. The last reserve banks (Cleveland and San Francisco) dropped their rates to 1% in April 1942, where all districts' discount rates remained through the end of 1947. If individual reserve banks had been attempting to counterbalance fiscal expansion with the timing of their rate decreases, Minneapolis or Atlanta would have been much more reasonable candidates to decrease their rates first. If fiscal shocks had been a driving force, Chicago, Cleveland, and Philadelphia should have followed policies similar to Boston and New York. They did not. Thus it seems far more likely that the pattern of discount rates over 1939–1942 reflect some mix of idiosyncrasies of individual reserve banks and varying exposure to foreign markets, particularly London. Moreover, the spread between any two states was never more than 50 basis points. So in practice these deviations from uniform interest rates across U.S. states in the before 1942 seem unlikely to greatly affect the analysis.

other states (which do not receive spending shocks, or receive smaller shocks). When relative prices rise in the region receiving the fiscal shock, expectations of future price changes for that region fall relative to expected future price changes in other regions, causing a relative increase in the local long-run interest rate. Limits on price increases were largely implemented at a regional level during World War II, so wartime economic conditions would not have interfered with this adjustment mechanism.

The open economy relative multiplier is larger than the closed economy aggregate multiplier when monetary policy actively responds to (aggregate) fiscal shocks, but smaller when monetary policy is accommodating. However, the mechanism by which accommodative monetary policy should increase the size of the aggregate fiscal multiplier is that the long-run real interest rate falls, strengthening private demand.

But during World War II, households responded to income shocks largely by increasing their savings. Binding limits on consumption–particularly durables consumption—disrupted households' ability to increase their demand. I will show in Section 1.8 that increases in saving account for a large proportion of households' response to war spending shocks, severely dampening their consumption response to fiscal shocks. With this channel limited, the aggregate multiplier should not be as large relative to the open economy relative multiplier.

To conceptualize this more precisely, it is helpful to carefully consider the determinants of the aggregate multiplier in Nakamura & Steinsson's model. The aggregate multiplier is larger under accommodative monetary policy (the constant nominal rate case) than under neutral monetary policy (the constant real rate case) *due to consumption responses.* If households cannot (or will not, for reasons not captured in the model) increase their consumption, this channel will cease to function, so the constant nominal rate case will de facto revert to the neutral monetary policy case. This implies that the relationship between the aggregate multiplier and open economy relative multiplier should be closer to Nakamura & Steinsson's case of neutral monetary policy. In this case, the aggregate multiplier should be about 20% larger than the open economy relative multiplier.

1.5 Baseline Specifications and Results

Employment Responses

I begin by estimating the effects of war production spending on employment:

$$\Delta^k E_{it} = \beta \Delta^k G_{it} + \alpha_i + \gamma_t + F_i * \gamma_t + \epsilon_{it} \tag{1.1}$$

where E is total non-farm employment[20] divided by population, G is real per capita war production spending (in thousands of 1942 dollars), F is the share of the state's population employed in agriculture in 1939. Subscripts are i for state and t for quarter, and Δ^k is the k^{th} difference (e.g. $\Delta^k E_{it} = E_{it} - E_{i,t-k}$).

[20]Excludes personnel in the armed forces, but includes civilians employed by military agencies.

Table 1.1: Baseline Effects of War Production Spending on Total Non-Farm Employment and Personal Income

	Employment		Personal Income	
β	0.268***	0.230***	0.0415	0.248***
	(0.0310)	(0.0318)	(0.0586)	(0.0526)
1939 farmshare x Time FEs?	No	Yes	No	Yes

Standard errors clustered by state. Regressions weighted by states' 1939 populations. Time fixed effects and interaction between time fixed effects and pre-war agricultural employment shares estimated but not shown.
*** p<0.01, ** p<0.05, * p<0.1

As discussed in Section 1.4, interacting time fixed effects with a measure of states' pre-war agricultural intensity allows for different underlying trends in the farm sector, though as Table 1.1 shows this interaction does not greatly alter the estimated employment effects of war production spending.[21]

I set k to 2 years, or 8 quarters. Using long differences reduces attenuation bias because shocks to the variable of interest are more correlated than the errors; it also allows for some imprecision about the exact timing of government spending shocks hitting the economy. Following many papers on the employment effects of government spending, I use a difference specification for employment because it allows a simple calculation of the additional spending associated with an additional job.[22]

Regressions are weighted by states' 1939 populations for two reasons. First, we are interested in the stimulative effect of the average dollar of spending. Because spending is computed in per capita terms, it is necessary to weight by population in order to compute the effect of the average dollar spent. (I weight using 1939 population to avoid endogeneity from wartime population movements.) Second, population weighting is likely to reduce downward bias on the estimates from spillovers between states. This is because spillovers from larger states to smaller states are almost certainly bigger than spillovers from small states to larger states.[23]

Using the baseline specification with $k = 8$ quarters, standard errors clustered by state, and the regression weighted by states' 1939 populations, I find $\beta = 0.230$ (*SE = 0.032, P < .001*), which translates to a cost of $4,300 per job-quarter in 1942 dollars. This translates

[21]Note that because a state's pre-war share of agricultural employment is time-invariant, F_i is collinear with state fixed effects. Thus it would be redundant to include F_i separately in the estimating equation.

[22]The cost per job-quarter is $\$1,000/\beta$ in 1942 dollars, which translates into a cost per job-year of $14.54 * 4 * \$1,000/\beta$ in 2015 dollars.

[23]The point estimates on β are significantly smaller when weighting is not used, and income results lose statistical significance. For the baseline regression on employment, the unweighted regression produces $\beta = 0.151$ (*SE = 0.056, P < .010*), or a cost per job-year of $386,000 in 2015 dollars.

Table 1.2: Employment Effects of War Production Spending by Sector

Total	Total Private	Manufacturing	Private Non-Manufacturing	Government
(SIC 1–7)	(SIC 1–6)	(SIC 3)	(SIC 1–2, 4–6)	(SIC 7)
0.230***	0.310***	0.274***	-0.00738	-0.0416
(0.0318)	(0.0300)	(0.0296)	(0.0210)	(0.0364)

All effects estimated using two-year differences in both spending and employment. Standard errors clustered by state, and regressions weighted by states' 1939 populations. Time fixed effects and interaction between time fixed effects and pre-war agricultural employment shares estimated but not shown. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

to an estimated cost of about $253,000 per job-year in 2015 dollars.[24] This figure is high, but not terribly far outside the range of the literature. For comparison, it is instructive to consider recent estimates of job creation from the 2009 American Recovery and Reinvestment Act (ARRA). At the high end, Feyrer & Sacerdote (2011) find an overall cost of $170,000 per job-year and Wilson (2012) finds an initial cost of $125,000 per job-year, both using fairly general measures of ARRA spending distributed through states. At the low end, Chodorow-Reich et al. (2012) find a $26,000 cost per job-year for the Medicaid expansion in the ARRA, or 3.8 job-years per $100,000 of spending, and Shoag (2013) finds a cost of $22,000 per job-year using shocks to state pension fund returns during the Great Recession.

However, these vast differences in top-line figures mask underlying similarities in the results. For instance, Chodorow-Reich et al. find that 3.2 out of those 3.8 job-years per $100,000 of fiscal stimulus appear outside the health, education, and government sectors (the sectors in which one would expect to see direct effects from the spending increase they study), implying a "direct" cost per-job year of $167,000. In contrast, the effects of war production spending are largest for manufacturing employment (see Table 1.2 below), with a "direct" cost of $212,000 per manufacturing job-year.[25] Thus the vast difference in the estimated cost per job-year is principally driven by differences in the *indirect* effects of fiscal stimulus: in the more recent case the indirect effects appear to be large and positive, while in World War II they appear to be negative. The forces interfering with the "usual" multiplier process during World War II are discussed at length in sections 1.7 and 1.8.

[24] An even simpler approach is to ignore timing variation and simply look at the cross-sectional effect of spending on employment by state. Because war production peaks in 1944, the most sensible way to do this is to regress state-level changes in employment (scaled by population) between 1940 and 1944 on total per capita state-level war spending over the same period. This approach shows that each additional $1,000 per capita in war spending (about $14,500 in 2015 dollars) was associated with a 1.6% increase in a state's employment/population ratio over the same period, with a t-statistic of 3.34 and P-value of .002.

[25] This figure is almost certainly an underestimate of the "true" direct cost, since much of the crowd-out during World War II took place *within* the manufacturing sector.

Income Responses

For income, my baseline specification is:

$$\frac{Y_{it} - Y_{i,t-k}}{Y_{i,t-k}} = \beta \frac{G_{it} - G_{i,t-k}}{Y_{i,t-k}} + \alpha_i + \gamma_t + F_i * \gamma_t + \epsilon_{it} \tag{1.2}$$

where Y is personal income per capita (in thousands of 1942 dollars) and G and F are war production spending and the pre-war share of agricultural employment, defined as above. Because state-level personal income data is annual rather than quarterly, income responses are estimated using annual data. With $k = 2$ years, this specification is quite similar to that used by Nakamura & Steinsson, and β can be interpreted as the open-economy relative multiplier for the fiscal shocks from war production spending.[26]

Using the baseline specification with $k = 2$ years, standard errors clustered by state, and weighting by states' 1939 populations, I find an open-economy relative multiplier of $\beta = 0.248$ (*SE = 0.053, P < .001*). When separate time fixed effects for farm states are not included, $\beta = 0.041$ (*SE = 0.059, P = .483*). These estimates are summarized in Table 1.1.

What does my finding of an open economy relative multiplier of 0.25 imply about the aggregate multiplier for World War II? As discussed in Section 1.4, the aggregate multiplier should be approximately 20% larger than the open economy relative multiplier under economic conditions similar to those prevailing during World War II. This suggests an aggregate multiplier around 0.3 for World War II, only slightly below Hall's estimate of 0.36 for the same period.

The time series literature on fiscal multipliers has estimated lower aggregate multipliers for periods including World War II than for periods excluding it, though in many cases estimates excluding World War II lack statistical significance. Barro & Redlick (2011) find an aggregate multiplier of 0.69 when beginning their sample in 1950, but 0.44–0.66 for samples including World War II (depending on how one interprets their results)Barro and Redlick (2011). Hall (2009) finds an aggregate multiplier of just 0.36 for 1939–44, but 0.47 for 1948–2008. Ramey (2011) finds the opposite, a *smaller* multiplier when excluding World War II,[27] but Ramey & Zubairy's (2016) findings are in keeping with the rest of the literature: they find evidence that the multiplier is larger when the economy is weak only when *excluding* World War II from their sample.

When interpreting these estimates of the aggregate multiplier, it is essential to understand just how strongly they are influenced by World War II. The Second World War effectively dominates even time series spanning many decades because the scale of military spending shocks was so large compared to any other time in U.S. history. This means that certain wartime observations are always highly leveraged in any standard estimation procedure—a point made with great clarity in Figure 1 of Hall (2009). Thus understanding the effects of

[26] The advantages of this particular approach—using percentage change instead of log differences and dividing the change in G by the initial level of Y rather than the initial level of G, effectively scaling the change in G by G/Y—have been argued persuasively by both Bob Hall and Valerie Ramey.

[27] See Appendix A for a discussion of what factors influence this estimate and others in the literature.

fiscal stimulus in World War II is a key context for interpreting the time series literature on the stimulative effects of military spending in the U.S.

Impulse Responses

While the baseline approach has the advantages of clarity and simplicity, economic stimulus is a dynamic process, and the process itself is also of interest. I construct impulse responses for employment and personal income using an estimation procedure broadly similar to Ramey & Zubairy's (2016) modified Jordà projection. This approach traces out *cumulative* impulse responses. That is, at each time horizon the total effect and not the incremental effect is estimated. Importantly, the time spans used for the variables of interest on both the right and left hand sides are always the same, so that the computed impulse in the outcome variable will appropriately account for sequential spending shocks on the right hand side.

For employment I estimate:

$$\Delta^h E_{i,t+h} = \beta \Delta^h G_{i,t+h} + \eta_1 \Delta^k E_{it} + \eta_2 \Delta^k G_{it} + \alpha_i + \gamma_t + F_i * \gamma_t + \epsilon_{it} \tag{1.3}$$

The impulse is computed by running a separate regression for every time horizon $h = 1, ..., 12$ quarters. Long difference length k is once again set to 8 quarters, and applied to the lagged controls. β can once again be used to directly compute cost per job-quarter.

For personal income I estimate:

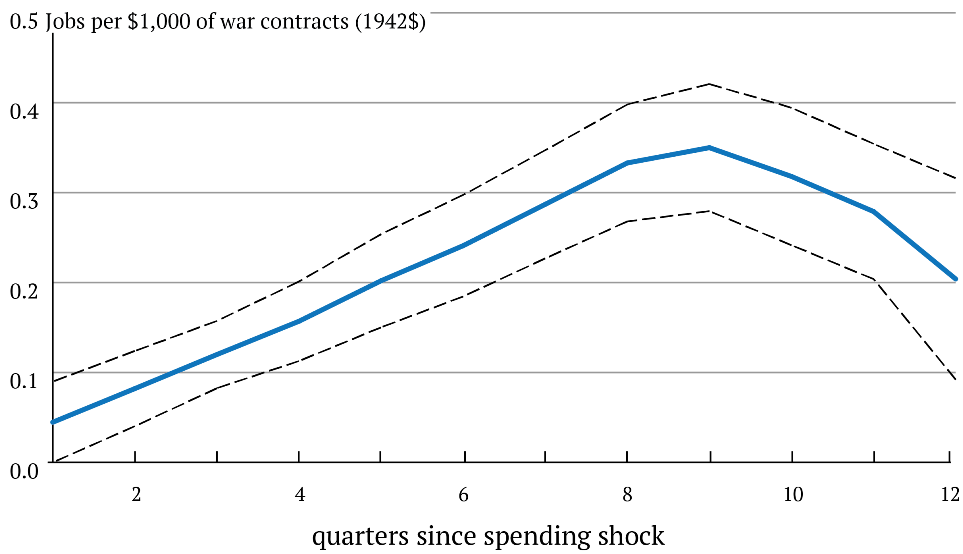
$$\frac{Y_{i,t+h} - Y_{it}}{Y_{it}} = \beta \frac{G_{i,t+h} - G_{it}}{Y_{it}} + \eta_1 \frac{Y_{it} - Y_{i,t-k}}{Y_{i,t-k}} + \eta_2 \frac{G_{it} - G_{i,t-k}}{Y_{i,t-k}} + \alpha_i + \gamma_t + F_i * \gamma_t + \epsilon_{it} \tag{1.4}$$

Again, the impulse is computed separately for each time horizon, this time with $h = 1, 2, 3$ years. Long difference length k is once again set to 8 quarters, and applied to the lagged controls, and β can be interpreted as the fiscal multiplier (on personal income).

As explained by Ramey & Zubairy, an advantage of constructing the impulse by expanding the time horizon on both sides is its simplicity: as a one-step estimation procedure, the standard error is computed directly for each time horizon. This avoids the complexity and loss of precision endemic to less direct procedures. However, a crucial difference between my procedure and Ramey & Zubairy's is that instead of summing quarterly changes in both variables over each relevant time horizon, I simply compute the net change from time t to time $t+h$.[28] When using differences to compute employment responses (again because they allow a direct calculation of cost per job) this distinction is irrelevant. When using percentage change, as I do for income, my approach is appropriate because the scale of spending shocks in World War II is so large that updating the denominator introduces significant endogeneity into the calculation—and with it, bias. In most other cases the scale of spending shocks is

[28] Ramey & Zubairy scale by potential output rather than actual output in their baseline approach. Between the Great Depression, World War II, and data limitations, there is no reasonable way to estimate state-level potential output for the 1940s.

Figure 1.7: Impulse Response of Total Non-Farm Employment to War Production Spending



Source: Author's calculations based on data from Civilian Production Administration, BLS, and Census. Dotted lines show 95% confidence interval.

small relative to other factors altering output, making it appropriate to update output in the denominator, but that is simply not the case for World War II production.

Controlling for lagged values of both variables is a standard part of the Jordà projection approach, and aims to separate the "true" innovation in each variable from its prior trend. However, instead of using the more common one-period difference in each lagged control, I follow the spirit of my baseline approach and use the change in each variable over the preceding two years. Again, the longer differences are less influenced by noise and therefore contain more information. In practice they have stronger predictive power than more standard one-period differences.

The resulting impulses for various kinds of employment are shown in Figure 1.7 and Table 1.3. The impulse response for personal income is given in Table 1.4.

The effects of war production spending on total (non-farm), private, and manufacturing employment are positive and significant at every time horizon out to 12 quarters (a limit chosen given the short length of the panel). Effects peak 8–9 quarters after the initial shock. The effects on manufacturing employment are consistently larger than on total non-farm employment, implying that crowd-out is active at every time horizon. At peak impact, these estimates imply a cost per job-year of $147,000 in manufacturing and $166,000 overall, both in 2015 dollars.

The effects on private non-farm employment outside of manufacturing and on government are generally negative and but rarely significant. This is consistent with the baseline results

Table 1.3: Cumulative Impulse Responses of Employment by Sector to War Production Spending

horizon (quarters)	Total (SIC 1–7)	Total Private (SIC 1–6)	Manufacturing (SIC 3)	Private Non-Manufacturing (SIC 1–2, 4–6)	Government (SIC 7)
1	0.0447*	0.0815***	0.0744***	-0.0107	-0.0148**
	(0.0226)	(0.0278)	(0.0238)	(0.0123)	(0.00692)
2	0.0821***	0.127***	0.135***	-0.0229*	-0.0193*
	(0.0209)	(0.0263)	(0.0250)	(0.0122)	(0.00972)
3	0.120***	0.161***	0.169***	-0.0290**	-0.0236*
	(0.0187)	(0.0258)	(0.0283)	(0.0134)	(0.0131)
4	0.157***	0.203***	0.208***	-0.0264	-0.0265
	(0.0221)	(0.0303)	(0.0297)	(0.0165)	(0.0173)
5	0.202***	0.253***	0.259***	-0.0227	-0.0314
	(0.0259)	(0.0341)	(0.0305)	(0.0188)	(0.0238)
6	0.241***	0.298***	0.305***	-0.0164	-0.0365
	(0.0282)	(0.0331)	(0.0283)	(0.0180)	(0.0303)
7	0.287***	0.344***	0.346***	-0.0100	-0.0352
	(0.0301)	(0.0302)	(0.0291)	(0.0169)	(0.0313)
8	0.333***	0.380***	0.379***	-0.00140	-0.0317
	(0.0325)	(0.0342)	(0.0348)	(0.0182)	(0.0288)
9	0.350***	0.395***	0.374***	0.00876	-0.0270
	(0.0353)	(0.0366)	(0.0370)	(0.0193)	(0.0230)
10	0.318***	0.359***	0.330***	0.00225	-0.0233
	(0.0382)	(0.0396)	(0.0390)	(0.0181)	(0.0157)
11	0.279***	0.310***	0.287***	-0.0142	-0.0214**
	(0.0376)	(0.0408)	(0.0256)	(0.0187)	(0.0105)
12	0.204***	0.264***	0.249***	-0.0145	-0.0258**
	(0.0560)	(0.0528)	(0.0273)	(0.0201)	(0.0125)

Standard errors clustered by state, and regressions weighted by states' 1939 populations. Time fixed effects and interaction between time fixed effects and pre-war agricultural employment shares estimated but not shown. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

given in Table 1.2.

With only annual data, the impulse response for personal income is less detailed, but broadly consistent with the results for employment. The effects are largest at a horizon of 2 years, and roughly equal at horizons of 1 and 3 years. The open economy relative multiplier is estimated at 0.31 over two years: somewhat larger than the baseline estimate of 0.25, but with broadly similar economic implications. This approach implies an aggregate multiplier of 0.38 for World War II.

Table 1.4: Cumulative Impulse Response of Personal Income To War Production Spending

horizon (years)	1	2	3
Spending (impulse)	**0.179*** **	**0.309*** **	**0.409*** **
	(0.0405)	(0.0555)	(0.105)
Lagged spending change	0.0817**	0.170*	0.355*
	(0.0401)	(0.0845)	(0.182)
Lagged personal income	-0.158***	-0.642***	-1.002***
	(0.0363)	(0.0747)	(0.120)
Observations	294	245	196
Within R-Squared	0.744	0.845	0.894

Standard errors clustered by state, and regressions weighted by states' 1939 populations. Time fixed effects and interaction between time fixed effects and pre-war agricultural employment shares estimated but not shown.
*** p<0.01, ** p<0.05, * p<0.1

Robustness Checks

Government Employment as a Placebo Test

As discussed in Section 1.4, one possible identification concern is that the distribution of government spending often results from political power, and that it may be correlated with other forms of political patronage as a result. A placebo case suggests that this was not the case for World War II. Table 1.2 shows the effects of war spending on various categories of employment (using the baseline specification). As one would expect given that the war production contracts are all with private firms and the vast majority of contracts were in manufacturing,[29] the employment effects are stronger for private-sector (non-farm) employment, and are strongest for manufacturing.

Notably, however, war production spending is not positively related to growth in government employment. If anything, war production spending had a negative effect on government employment, though those effects are not significant at most time horizons. Government employment did grow significantly during the war, and roughly half of that increase was in federal employment, so if war spending was systematically biased by political patronage, one might expect a positive correlation between war production spending and government employment. However, the estimated effect of war spending is small (*-.032*), and not statistically significant.

It seems likely that the sheer scale of war spending reduced competition for contracts: in cases where two or more firms were capable of producing a specific good at scale, war contracts generally went to both firms. In fact, the government's main concern was with

[29] A small proportion of contracts in the data are for services, generally education and training or storage services. Because the data also include purchases of raw materials, there are also some contracts that would be classified as mining employment.

filling *enough* contracts. While rationing and crowd-out of civilian production are discussed at length in Sections 1.7 and 1.8, it is important to remember that firms receiving military contracts often had to curtail (or even suspend, in some industries) non-war production. Additionally, war contract placements were generally channeled through the War Production Board and the supply arms of the armed services, seemingly reducing the power of members serving on key congressional committees. Thus the usual dynamics of pork barrel spending do not seem particularly relevant to World War II, and the geographic placement of war contracts appears less political than it would have been under ordinary circumstances.

Wartime Population Flows

A potential concern with the econometric approach is that wartime population movements systematically went towards states with higher levels of war spending. Because all variables of interest are scaled by state population, this could potentially bias the results. I test for the effects of wartime population movements in two ways.

First, I run the baseline specification (analogous to that for employment) with the level change in population as the outcome variable.

$$\Delta^k Population_{it} = \beta \Delta^k G_{it} + \alpha_i + \gamma_t + F_i * \gamma_t + \epsilon_{it} \tag{1.5}$$

The estimated coefficient on war production spending in this regression is not statistically distinguishable from zero, and is in fact slightly negative: $\beta = -0.024$ (*SE = 0.023, P = .283*). When all fixed effects (state and and time) are excluded, the coefficient becomes slightly positive (0.17) but remains statistically insignificant.

Second, I rerun the baseline specifications for employment and income with all population denominators fixed to states' 1939 populations. If the results were significantly different, the difference could in some sense be interpreted as the influence of migration. However, as can be seen in Table 1.5, the results are extremely similar no matter whether the population measure used to scale variables is allowed to evolve over time (as in the baseline specification) or is held fixed at 1939 levels. In both cases, the difference in the point estimate is smaller than the standard error on the estimates.

Together, these tests suggest that wartime population movements are not an important driver of my results, and unlikely to be a source of bias.

Controlling for Tax Changes

Another possible concern is the influence of tax policy. Federal income taxes were increased substantially during World War II, both to finance the war and as a tool to control inflation (since interest rates were kept low to finance the war).

Statutory increases in federal taxes were of course the same in all states, so a large portion of the effects of wartime tax changes should be accounted for by time fixed effects. This implies that the influence of taxes on the open economy relative multiplier should be relatively small, though still positive since taxes increase with income.

Table 1.5: Effects of War Production Spending on Total Non-Farm Employment and Personal Income with Population Adjusting vs. Fixed to 1939 Levels

	Employment		Personal Income	
β	0.230***	0.221***	0.248***	0.261***
	(0.0318)	(0.0641)	(0.0526)	(0.0726)
Population Measure	baseline	1939	baseline	1939
Observations	980	980	294	294
Within R-squared	0.877	0.816	0.791	0.755

Standard errors clustered by state. Regressions weighted by states' 1939 populations. Time fixed effects and interaction between time fixed effects and pre-war agricultural employment shares estimated but not shown.
*** p<0.01, ** p<0.05, * p<0.1

However, federal income taxes had a highly progressive rate structure throughout the period. Additionally, war production spending tended to be concentrated in industrial states where income levels were higher to begin with. Both of these factors should lead to relatively larger effective tax rate increases in states that received more war spending.

For these reasons, not accounting for differential changes in federal taxation could lead to a downward bias in estimates of the open economy relative multiplier during wartime. I gathered state-level data on federal income tax filings from the historical *Statistics of Income* and used it to build a state-level panel of total federal income taxes by the year in which taxes were collected.[30]

Contemporaneous tax increases are highly endogenous to war production spending, so I use lagged tax changes instead. For employment I estimate:

$$\Delta^k E_{it} = \beta \Delta^k G_{it} + \tau \Delta^k T_{i,t-k} + \alpha_i + \gamma_t + F_i * \gamma_t + \epsilon_{it} \tag{1.6}$$

where T is per capita taxes filed. Note that because income tax data is annual, I use annual data for employment in this robustness check. With annual data, β corresponds to cost per job-year rather than cost per job-quarter, and will therefore be mechanically smaller (by a factor of 4) than estimates produced with annual data. For personal income I estimate:

$$\frac{Y_{it} - Y_{i,t-k}}{Y_{i,t-k}} = \beta \frac{G_{it} - G_{i,t-k}}{Y_{i,t-k}} + \tau \frac{T_{i,t-k} - T_{i,t-2k}}{Y_{i,t-2k}} + \alpha_i + \gamma_t + F_i * \gamma_t + \epsilon_{it} \tag{1.7}$$

[30]The relationship between the year in which income was earned and the tax filing year changed in the middle of World War II. Tax data from 1942 is omitted because filings for 1942 bore very little relationship to taxes actually paid on income from that year due to the mechanics of the transition. (Actual 1942 tax payments depended on income over multiple years.)

Table 1.6: Effects of War Production Spending on Employment and Income with Controls for Tax Changes

	Employment (total)		Personal Income	
β	0.0583***	0.0832***	0.253***	0.256***
	(0.00795)	(0.0122)	(0.0521)	(0.0608)
τ		-0.140		-1.277**
		(0.0852)		(0.552)
Taxes included?	No	Yes	No	Yes
Observations	245	98	343	98
Within R-squared	0.881	0.888	0.847	0.883

Standard errors clustered by state. Regressions weighted by states' 1939 populations. Time fixed effects and interaction between time fixed effects and pre-war agricultural employment shares estimated but not shown. *** p<0.01, ** p<0.05, * p<0.1

Estimates of β and τ from these specifications are reported in Table 1.6. Controlling for income taxes noticeably increases the employment effect of war production spending, but has no noticeable effect on income. The estimated cost per job-year is \$249,000 (in 2015 \$) under the baseline specification (very close to the \$253,000 baseline estimate using quarterly data), but only \$175,000 (in 2015 \$) when controls for federal income tax increases are included. However, the estimated open economy relative multiplier (on personal income) increases by only 0.003 when controls for income taxes are added, even though the estimated income response to tax changes is substantial.

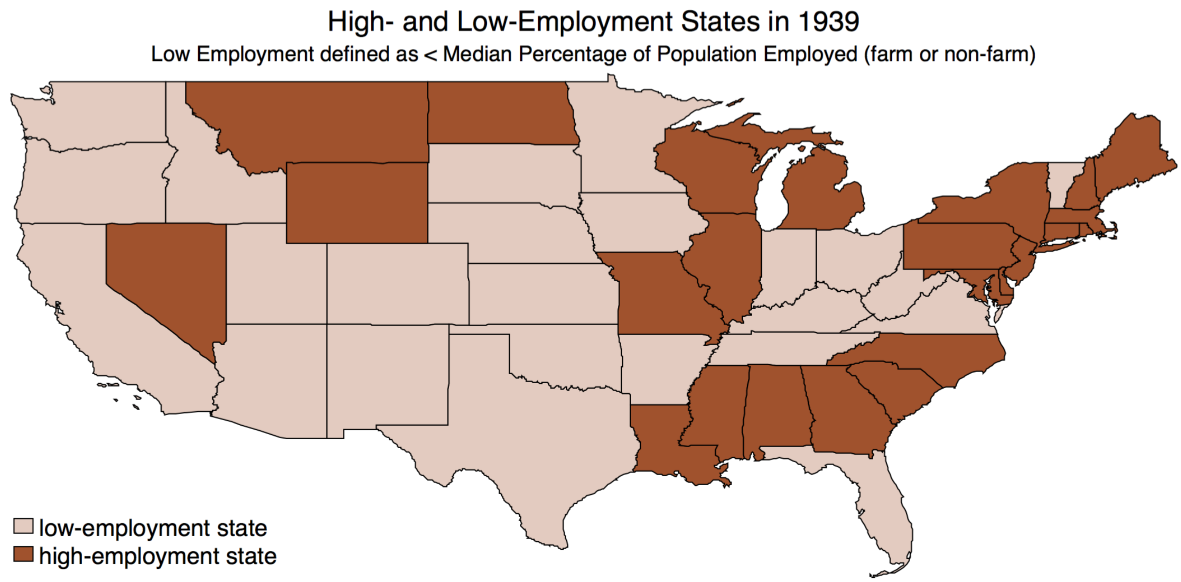
1.6 Evidence for State Dependence

Recently the literature has focused on variation in the size of the fiscal multiplier over the business cycle. When the economy is weak, there is higher "slack" in the economy: more potentially productive workers lack jobs, and output is below potential. Intuitively, higher slack should make it easier for government spending to stimulate the economy.

Using a postwar sample of aggregate data for the U.S., Auerbach & Gorodnichenko (2012) find strong evidence for state dependence in the multiplier. They find particularly large differences for defense spending: a fiscal multiplier of 0.8 for expansions versus 3.6 for recessions. But when Ramey & Zubairy (2016) use time series data on U.S. defense spending that begins much earlier, they find no evidence for state dependence in the multiplier *except* when World War II is omitted from their sample.

Because the U.S. was still recovering from the Great Depression when it entered the war (initiating the largest positive spending shocks), World War II contributes high slack (high unemployment) observations. These shocks will be highly leveraged even in a large sample because the scale of spending shocks in World War II was so large relative to all others (even

Figure 1.8: Pre-War Variation in State Employment Rates



World War I). Thus they strongly influence the estimation of the high-slack multiplier.

Since World War II spending shocks are so large and also strongly correlated across time, using a measure of slack that responded to wartime changes would invite substantial endogeneity problems. Instead, I construct an indicator for slack based on pre-war employment levels. This indicator is 1 for all states whose total employment rates were at or above the median in 1939, and 0 otherwise.

Crucially, the measure of employment used to construct the indicator for slack includes agricultural employment. Slack should be measured by total economic activity, and not influenced by differences in sectoral composition between states—as would implicitly be the case for any measure based solely on non-farm employment. I use farm employment statistics for 1939 from the 1940 Census of Agriculture, adjusted to exclude agricultural workers who also worked off-farm, and combine these figures with BLS estimates of non-farm employment.

The resulting measure of slack, shown in Figure 1.8, has a correlation of -.08 with pre-war shares of agricultural employment, suggesting that it captures pre-war variation between states unrelated to agricultural intensity.

To estimate the effects of state dependence, I interact the spending shock with the indicator for slackness. For employment, the baseline specification is augmented as follows:

$$\Delta^k E_{it} = \beta_1 \Delta^k G_{it} + \beta_2 \Delta^k G_{it} * \mathbb{1}(\textit{high employment 1939}) + \alpha_i + \gamma_t + F_i * \gamma_t + \epsilon_{it} \qquad (1.8)$$

The baseline specification for personal income is extended analogously. Note that $\mathbb{1}(\textit{high employment 1939})$ does not enter into the specification independently because it is time-invariant and therefore collinear with the state fixed effects.

Thus the effect of war production on spending is estimated as β_1 for states with low initial employment rates (high slack), and $\beta_1 + \beta_2$ for states with high initial employment

Table 1.7: Effect of War Production Spending on Income and Employment, Controlling for Initial Employment Rates

	Employment (total)		Personal Income	
β_1	0.230***	0.313***	0.253***	0.361***
	(0.0318)	(0.0332)	(0.0521)	(0.0821)
β_2		-0.124***		-0.191***
		(0.0294)		(0.0680)
Interaction?	No	Yes	No	Yes
Observations	980	980	294	294
Within R-squared	0.877	0.886	0.847	0.860

Standard errors clustered by state. Regressions weighted by states' 1939 populations. Time fixed effects and interaction between time fixed effects and pre-war agricultural employment shares estimated but not shown. *** p<0.01, ** p<0.05, * p<0.1

rates (low slack). If there is state dependence in the multiplier β_2 should be negative and significant—and indeed it is.

I find significantly larger stimulative effects from war production spending in states with below-median employment levels in 1939. The results are summarized in Table 1.7. For employment, I estimate $\beta_1 = .313$ and $\beta_2 = -.124$, which translates to a cost per job-year of $186,000 for states with low initial employment and $308,000 for states with high initial employment (both in 2015 $). For personal income, the implied open economy relative multipliers are .376 and .17, respectively.

An alternative approach would be to use a continuous measure of pre-war employment (combining both agricultural and non-agricultural employment) rather than constructing an indicator. Because this alternative approach involves the interaction of two continuous variables, interpreting the estimates from this approach is less straightforward. Moreover, interactions between continuous variables can exaggerate the influence of outliers, potentially biasing results. However, this approach produces qualitatively similar estimates.[31]

These results are far from trivial. The estimated spending per net job is more than 50% higher in states that started out with higher employment rates. The multiplier is more than twice as large in states with low initial employment than in states with high initial employment—smaller than the effects of state dependence estimated by Auerbach & Gorodnichenko for aggregate time series data, but roughly comparable.

Small multipliers for World War II should not be interpreted as evidence against state dependence in the multiplier. I use cross-sectional variation in pre-war employment levels to show that state dependence does in fact hold as theorized *within* World War II. The

[31] For employment, interacting government spending with the share of the state's population employed (in farm or non-farm jobs) in 1939 produces $\beta_1 = 1.152$ (*SE = 0.375, P = .003*) and $\beta_2 = -2.347$ (*SE = 0.995, P = .022*). For personal income, this approach produces $\beta_1 = 2.001$ (*SE = 0.553, P = .001*) and $\beta_2 = -4.554$ (*SE = 1.477, P = .003*).

aggregate effects of state dependence are simply dwarfed by the influence of other features of the wartime economy.

1.7 Conversion of Manufacturing Capacity and Output

It is essential to remember that war production was not conceived as an economic stimulus—or to fulfill any economic objective at all. Or in the words of Donald Nelson, Chairman of the War Production Board, "the war production program was not a form of the [Works Progress Administration], and could not be used as one" (Nelson, 176). The goal of war production was to achieve total victory over the Axis powers as quickly as possible. Planners did their best to minimize economic harms where possible, but they could not always be avoided.

Before U.S. entry into the war, the American defense industry was tiny. Many defense production facilities from WWI had been abandoned, or in some cases intentionally demolished (to avoid paying capital taxes) during the interwar years.[32] While many new production facilities were built (some $17 billion was spent on wartime facilities, including new factories) and the existing defense industry expanded its production capabilities, these additions were far from adequate for supplying military needs in the necessary timeframe.

Thus *conversion* of existing manufacturing facilities from civilian to military production was necessary. In this context, conversion is defined as the retooling of facilities previously in use for civilian production to produce goods for the war effort. Conversion came about in two ways.

First, production of certain civilian goods was expressly prohibited through the duration of the war. The most famous example of this is the automotive industry, which was required to stop producing vehicles for civilian use in early February 1942 and did not resume production until well into 1945. The entire automotive industry converted to war production, and by the end of the war was estimated to have contributed roughly a fifth of all war production.[33]

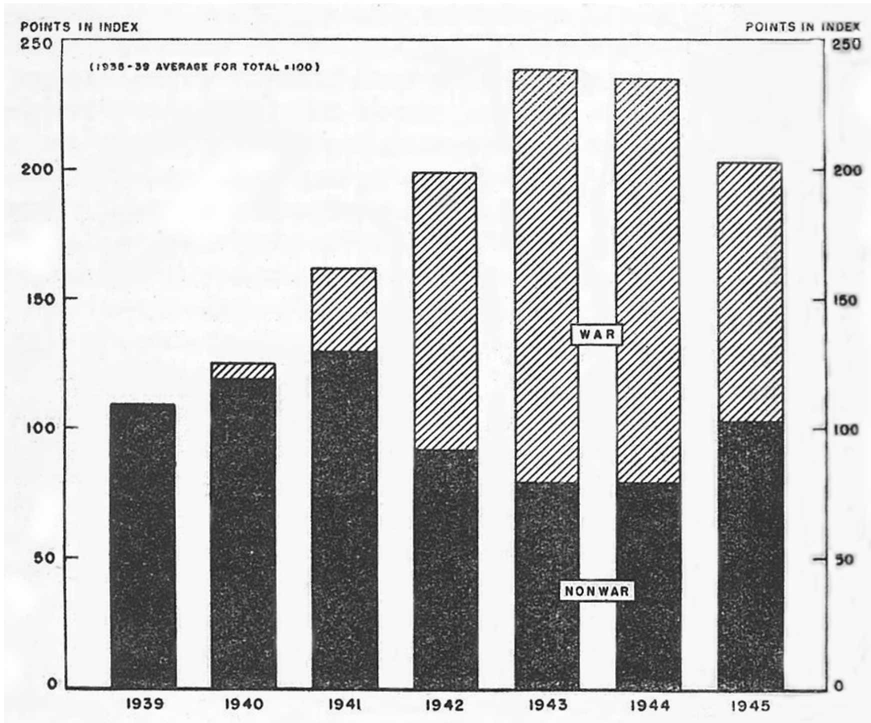
But outside of a few large industries such as automobile and airplane manufacturing, conversion was incentivized through the allocation system for strategic materials, which made it increasingly difficult for firms manufacturing civilian goods to obtain necessary inputs. Shortages of raw materials—especially steel, aluminum, rubber, copper, and iron—were already growing acute in 1941, and became even more critical as the war progressed. The allocation system for strategic materials was set up so that defense production and production deemed essential for the smooth functioning of the civilian economy were prioritized.

In practice, it was nearly impossible for firms to obtain raw materials with strategic value for any other purpose. Yet rationed materials were necessary for producing many civilian

[32] See Nelson, p. 32.

[33] See Nelson, p. 217. Computations using the figures presented on p. 70 of *Automobile Facts and Figures (1947)* suggest that the correct figure was $29 billion, or about 16%.

Figure 1.9: Industrial Production 1939–1945 (United States)



Source: Bureau of the Budget, based on data from the Federal Reserve Board of Governors

goods, and in many cases (especially metals) there were also shortages of substitutes. Many firms then faced a choice between sitting idle for lack of inputs or retooling to take on defense production.

While the distinction between production via conversion and production via new (or previously unused) industrial capacity was meaningless in terms of the production of war matériel, it is an important distinction in terms of the macroeconomic effects of war production.

Conversion largely represents a change in the use of economic resources—a movement along the production possibilities frontier towards guns and away from butter. One should expect dollars spent on war production via conversion to have a much smaller (if any) stimulative effect compared to dollars spent on production at new facilities, expansions, or previously unused industrial capacity.

A Case Study of Conversion: Airplane Manufacturing

The aircraft industry provides a clear illustrative example of how conversion affected economic activity. Prior to World War II, airplane production was a very small industry with little standardization: a craftsman's art rather than a streamlined production process. The demand for airplanes spurred by the war forced the industry to adopt mass production techniques and drove massive expansion. The number of airplanes produced in the U.S. grew by a factor of 44 between 1939 and 1944, a figure which understates growth in the industry given the increases in the size and complexity of planes produced.

But the manner in which production capacity grew varied systematically by region. World War II airplane production was concentrated in three geographic regions: on the west coast (from Seattle to San Diego), around the Great Lakes, and in upstate New York and Connecticut. The contrast between the first two regions is particularly striking.

On the west coast, which had relatively low levels of manufacturing pre-war, nascent production facilities expanded hugely. This is the story of Douglas, Lockheed, and Boeing. Over the course of the war, Boeing produced nearly \$1.5 billion in airplanes and airplane parts (\$19.7 billion in 2015 dollars) in Washington state. Douglas and Lockheed's production in the greater Los Angeles area totaled \$2.4 billion and \$2.5 billion, respectively (\$35.0 and \$35.6 billion in 2015 dollars). Most of this production came from greatly expanded facilities. The 1947 Census of Manufactures shows 47,754 production workers (average over the year) manufacturing "aircraft and parts" (SIC 372) in just the Los Angeles metropolitan area—almost equalling the total for the entire United States in 1939.[34] Aircraft and parts are omitted from the data on Washington state for 1947 in order to avoid disclosure of figures for individual companies, but Boeing's annual reports reveal that the company had a total of 6,791 employees in 1939, of which 5,971 were employed in Seattle. At the end 1947, Boeing employed about 18,700 employees between its Seattle and Wichita plants. While the company's annual report for that year does not separate employment totals for the two locations, reported data on employment in transportation manufacturing and subsectors in the Seattle metropolitan area suggests that around 15,000 of those employees likely worked in Washington state.

But around the Great Lakes, automotive companies played a significant role in aircraft production. Ford produced \$1.9 billion in aircraft and parts in Dearborn, MI, and another \$1.3 billion in Ypsilanti, for a total of \$47.1 billion in 2015 dollars. General Motors' production of airplanes and parts was largest in Indianapolis, where production totaled \$1.4 billion (\$20.3 billion in 2015 dollars), with another \$39 million (\$560 million today) scattered across Michigan, and other large production centers spread from Illinois to New York. Nor was aircraft production at automotive companies limited to the largest firms. Packard produced \$1.4 billion in airplanes and parts in Detroit (\$20.5 billion in 2015 dollars), while Studebaker's production in South Bend, IN totaled \$650 million (\$9.4 billion in 2015 dollars).

This is not to say that there were no aircraft manufacturers located in the midwest, with trajectories similar to those of west coast firms. Curtiss-Wright produced \$550 million of airplanes and parts in Columbus and another \$500 million in Indianapolis (\$7.9 billion and \$7.2 billion in 2015 dollars, respectively). Its subsidiary, Wright Aeronautical, produced \$1.8

[34] I have not been able to find metropolitan area employment statistics by sub-industry for 1939. In 1947, the average number of production workers across the U.S. had jumped to 162,596. While available statistics are far less complete than one might wish, it seems likely that the increases were concentrated in California and Washington.

billion in airplanes and parts ($25.9 billion in 2015 dollars) in Lockland, OH. But more than half of the value of aircraft and parts produced for the war in Illinois, Indiana, Michigan, and Ohio came from firms in the automotive industry. This is reflected in the post-war industry composition of those states. In Michigan, the number of production workers (average over the year) manufacturing "aircraft and parts" (SIC 372) grew from 961 to 1,407 between 1939 and 1947—a much more modest increase than those seen in areas where production increases were not driven by conversion.[35]

In California and Washington, the rapid expansion of facilities for airplane production led to the creation of many new jobs and relatively large growth in the local economy. But in Michigan and Indiana, a large share of airplane production took place in factories that had previously been used to produce cars, at facilities that had already employed large numbers of workers before the war. So when production occurred through conversion as it did in airplane manufacturing in Michigan, the net effects on the local economy were noticeably smaller.

Measuring the Economic Influence of Conversion

There isn't data on which contracts were produced in new or expanded facilities, much less an ideal measure of conversion, which would distinguish which contracts superseded the production of civilian goods. However, pre-war manufacturing shares can be used as a rough measure of where conversion was more likely. A dollar of war production spending would be more likely to go to conversion in a state that had more manufacturing (relative to population) to begin with than in a state that entered the war with less manufacturing.

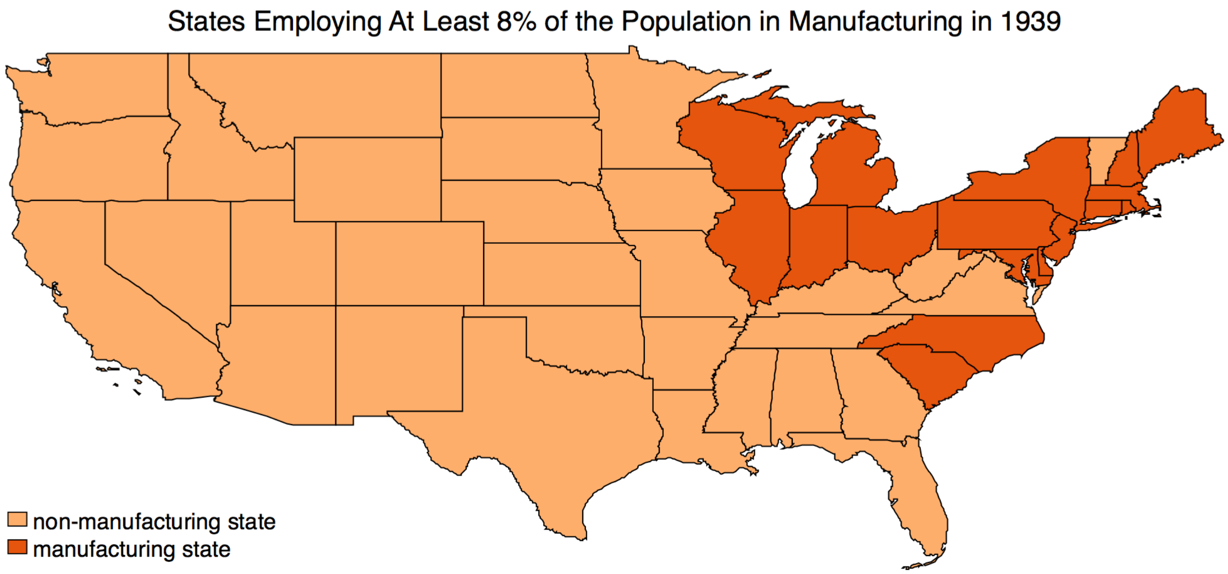
As a proxy for conversion, I construct an indicator for pre-war manufacturing intensity, shown in Figure 1.10. The measure is 1 if at least 8% of a state's population was employed in manufacturing in 1939 and 0 otherwise.[36] Once again, I use an indicator variable both because the interpretation of the coefficients is more straightforward and to avoid leveraging outliers in the interaction of two continuous variables However, the results are qualitatively similar when substituting the fraction of the state's population employed in manufacturing in 1939 for the indicator.[37]

[35]Employment in the manufacture of aircraft and parts was not reported for either Illinois or Indiana in the 1947 Census of Manufactures, and in both cases the residual number of production workers in transportation manufacturing after subtracting totals for reported categories were quite small. In Ohio, home of Curtiss-Wright, the number of production workers manufacturing airplanes and parts jumped from 820 in 1939 to 10,991 in 1947.

[36]I choose 8% as a threshold because it captures a group of states that were manufacturing intensive, but not entirely concentrated in the northeast. The distribution of manufacturing was so skewed during this time period that splitting the sample at the median would not be appropriate.

[37]For employment, interacting government spending with the share of the state's population employed (in farm or non-farm jobs) in 1939 produces $\beta_1 = 0.274$ (*SE = .111, P = .017*) and $\beta_2 = -0.129$ (*SE = 0.277, P = .644*). For personal income, this approach produces $\beta_1 = 0.481$ (*SE = 0.119, P < .001*) and $\beta_2 = -.659$ (*SE = 0.311, P = .039*).

Figure 1.10: Pre-War Concentration of Manufacturing Employment



To estimate the effects of conversion, I interact the spending shock with the indicator for pre-war manufacturing intensity. For employment, the baseline specification is augmented as follows:

$$\Delta^k E_{it} = \beta_1 \Delta^k G_{it} + \beta_2 \Delta^k G_{it} * \mathbb{1}(\textit{manufacturing 1939}) + \alpha_i + \gamma_t + F_i * \gamma_t + \epsilon_{it} \qquad (1.9)$$

The baseline specification for personal income is extended analogously.

Thus the effect of war production on spending is estimated as β_1 for states with low initial manufacturing intensity (where conversion was less common), and $\beta_1 + \beta_2$ for states with high initial manufacturing intensity (where conversion was more common). The results are summarized in Table 1.8.

Table 1.8: Effects of Government Spending on Employment and Income Controlling for Initial Manufacturing Industry Share

	Employment (total)		Personal Income	
β_1	0.230***	0.324***	0.253***	0.496***
	(0.0318)	(0.0452)	(0.0521)	(0.0711)
β_2		-0.117***		-0.293***
		(0.0389)		(0.0575)
Interaction?	No	Yes	No	Yes
Observations	980	980	294	294
Within R-squared	0.877	0.883	0.847	0.867

Standard errors clustered by state. Regressions weighted by states' 1939 populations. Time fixed effects and interaction between time fixed effects and pre-war agricultural employment shares estimated but not shown. *** p<0.01, ** p<0.05, * p<0.1

The effects are particularly striking for personal income. The open economy relative multiplier is estimated at .5 for states that had low manufacturing employment rates pre-war, but only .2 for states that had larger shares of manufacturing employment pre-war, suggesting that war spending had significantly smaller economic effects in areas where conversion was common.

An Alternate Approach to Measuring Conversion

Another way to think about conversion is that it is more necessary as the scale of spending increases: with a spending shock equivalent to 1-2% of output, it is much easier for those contracts to be filled without diverting workers and industrial capacity from other economic activities than with a shock equivalent to 10 or 20% of output. To estimate the effects of these nonlinearities, I estimate the effects of war production spending including the square of the spending shock variable. The results are shown in Table 1.9.

Table 1.9: Allowing for Nonlinear Effects of Spending

	Employment (total)		Personal Income	
Spending	0.268***	0.308***	0.238***	0.280***
	(0.0310)	(0.0358)	(0.0661)	(0.0770)
Spending Squared		-0.809***		-0.0818
		(0.181)		(0.0823)

Standard errors clustered by state. Time fixed effects (and separate time fixed effects for farm states in personal income results) estimated but not shown.
*** p<0.01, ** p<0.05, * p<0.1

For both employment and personal income, the coefficients on the square of spending are strongly negative (though only significant for employment) while the coefficients on spending are strongly positive. Other factors may also contribute to nonlinearities in the effects of spending. In particular, the stronger results for employment than income may indicate that smaller spending shocks have larger extensive margin effects, but larger spending shocks generate proportionally larger responses along the intensive margin (dampening the effects on personal income). But there are limits to how much labor can increase along both the intensive and extensive margins, and these results are broadly consistent with direct crowd out of non-defense production through conversion.

Interpreting the Economic Effects of Conversion

The textbook Keynesian fiscal multiplier is expressed as $1/(1 - mpc)$. Because conversion reduces the initial stimulative effect of government spending, it should be understood as a reduction in the numerator. Somewhat more formally, let c be the fraction of war production spending that directly displaces civilian manufacturing via conversion. The wartime fiscal multiplier is then $(1-c)/(1-mpc)$, which is strictly smaller than the standard multiplier for any positive value of c. If $c = 1/3$ in the aggregate, as suggested by Figure 1.9, conversion should reduce the aggregate multiplier by one-third. The results for personal income given in Table 1.8 suggest that conversion has a similar influence on the open economy relative multiplier.

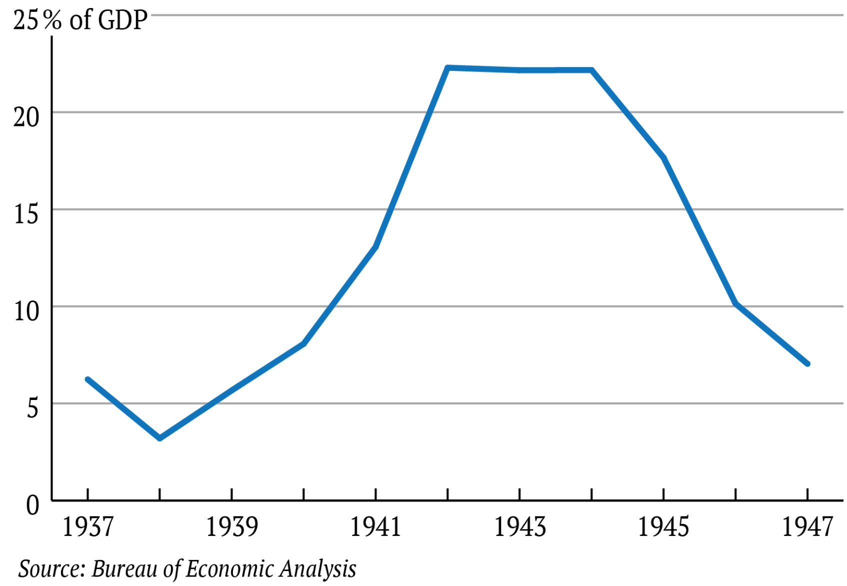
1.8 Savings and Consumption During Wartime

The textbook Keynesian multiplier is larger than 1 because household consumption responses magnify the initial effects of government spending. Even if conversion reduced the initial effects of spending because a significant fraction of spending went into replacing civilian production rather than expanding overall production, one would expect consumer responses to magnify the macroeconomic effects of spending, potentially bringing the total multiplier effect back above 1. But this Keynesian channel relies on households increasing their consumption in response to rising incomes. I argue below that this is not a realistic assumption for the U.S. economy during World War II.

Wartime Savings Behavior

Savings rates increased dramatically during World War II. In 1939, the net private saving rate was 6% of GDP. Net private saving jumped to more than 20% of GDP in 1941, 1942, and 1943, only to fall back to 10% in 1946 and 7% in 1947—back in line with pre-war savings behavior (see Figure 1.11). Since GDP was influenced by the massive growth of government spending, it is also useful to consider the behavior of net private saving as a share of the non-government portion of GDP. In these terms, net private saving grew from 5% in 1939 to

Figure 1.11: Net Private Saving (United States)



more than 30% in 1942, 1943, and 1944, peaking at 36.8% of non-government GDP in 1943. This measure of saving also fell back to 5% of GDP in 1947.[38]

In his classic *A Study of Saving in the United States*, Goldsmith shows a similar pattern for personal savings behavior during World War II. For most of the half-century preceding the war (excepting the depths of the Great Depression), the personal saving rate hovered around one-eighth of after-tax income. But for 1942–1945, personal savings jumped to around one-quarter of after-tax income, roughly doubling. Immediately after the war, the personal savings rate returned to pre-war levels.

Why did savings behavior alter so dramatically during the war? There are two main explanations: patriotism and desire to win the war, and wartime restrictions on purchases of durable goods through rationing.

Everyone from soldiers to school children was encouraged to save during World War II. Saving was presented as a patriotic duty. Purchases of government bonds—a major form of saving—helped finance the war effort, and large advertising campaigns encouraged everyone to participate. Pop stars wrote songs encouraging the purchase of war bonds, and posters advertised them to all segments of society. Children collected booklets of war savings stamps that, when filled, could be traded for war bonds, and many schools organized bond drives for their students. Workers purchased war bonds through payroll deductions, and soldiers purchased them while stationed overseas. In total, 85 million Americans purchased $185 billion in government securities between 1941 and 1946, of which $54 billion were war bonds, and $34 billion in series E bonds.[39]

[38] Calculations based on statistics from BEA.

[39] See *A History of the United States Savings Bonds Program* (1991) for details.

War bond purchases were an important way in which all civilians could aid the war effort, and a tangible way for people on the home front to support family and friends fighting overseas. Perhaps equally important, advertising campaigns tied war bonds to victory, suggesting that buying war bonds would help hasten the end of the war and return to ordinary peacetime life. These motives provided a strong incentive for saving during World War II.

Another explanation for wartime saving was that many consumer durables were largely unavailable. Conversion did not only mean reallocation of labor and capital. It also meant that many civilian goods were not produced and consequently could not be bought. As discussed in the previous section, production of a few broad categories of civilian goods was explicitly prohibited, but many other goods, while not explicitly prohibited, could not be produced due to the rationing of inputs necessary for their production.

Whether their production was explicitly or implicitly restricted, the goods whose manufacture required the same strategic materials needed for war production were overwhelmingly durable goods. The historical narrative is remarkably specific on this point. As early as July 1941, civilian planners for war production were aware that "manufacturers of consumers' durable goods requiring raw materials, principally metals, were facing increasing shortages and must be prepared to curtail their output of non-defense items."[40]

Rationing of automobiles was particularly acute. When all production of civilian automobiles was halted in February 1942 under Rationing Order 2, there was a stock of 521,000 new passenger cars nationwide, 28,000 of which were sales in process as of January 1 (which were allowed to go forward).[41] Between February 1942 and June 1943, a total of 328,000 passenger cars were released for civilian purchase nationwide. For comparison, there were 29.5 million passenger cars registered in the U.S. in 1941, and nearly 3.8 million new passenger car purchases in 1941 alone.[42] The effects of this rationing on spending are clearly shown in Figure 1.12.

But shortages of consumer durables for purchase did not end with automobiles. The materials—particularly metals—which were needed for war production were also needed to produce washing machines, refrigerators, and other household appliances. Figure 1.13 shows spending on household appliances. The pattern of wartime spending is strikingly similar to that for automobiles.

Because the manufacture of durable goods was so restricted due to materials shortages during the war, households could not respond to rising incomes by increasing their purchases of durables. Under more normal circumstances, purchases of durable goods are an important component of consumer responses to fiscal stimulus (see Parker et al., 2013; Hausman, 2016).

Consumption of nondurable goods was less affected by rationing, though few sectors were completely untouched by the war effort. But nondurables are not close substitutes for durables. It appears that many consumers saved instead—and likely used those savings to purchase durable goods after the war.

[40] Nelson, p. 145.

[41] *Automobile Facts and Figures 1943*, p. 56.

[42] *Automotive News (1952 Almanac Issue)*, p. 24. Total registrations also listed in *Automobile Facts and Figures 1943*, p. 48.

Figure 1.12: New Automobile Purchases (United States)

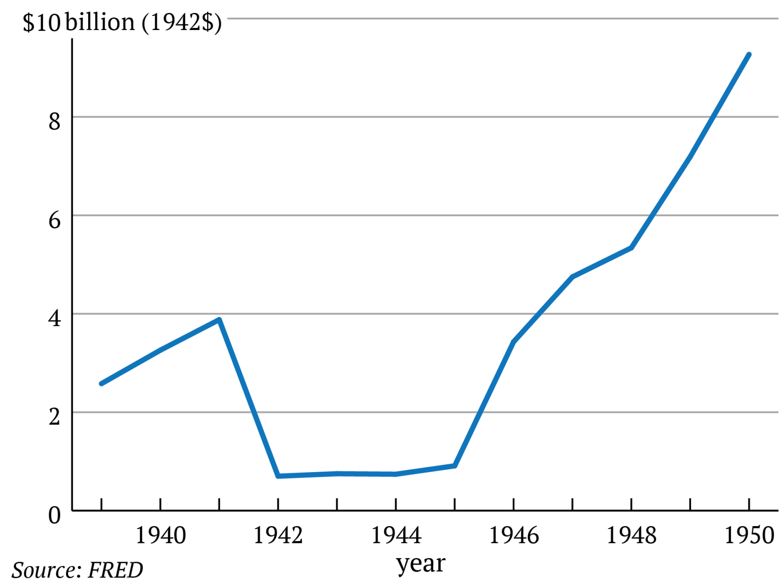
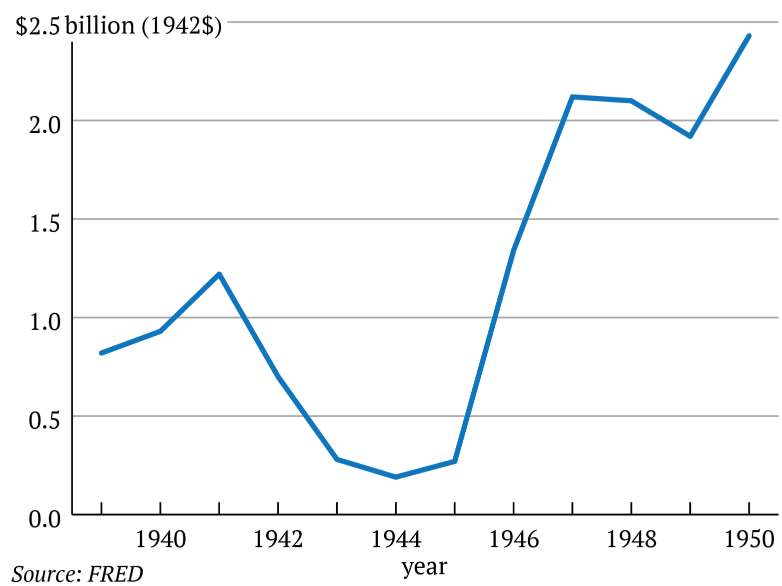


Figure 1.13: Personal Consumption Expenditures: Household Appliances (United States)



Effects of Wartime Savings & Consumption on the Multiplier

While the scale of the aggregate shift in consumption and savings behavior certainly suggests that consumer behavior may be an important part of the story and may influence the *aggregate* multiplier, aggregate shifts in consumption and savings cannot explain the size of the open economy relative multiplier discussed in sections 1.5 and 1.6 on their own.

If consumption is equally constrained everywhere, the cross-sectional effects should be small. This is especially true in the case of durable goods, because their manufacture is concentrated in specific geographic regions. When consumers respond to income shocks by buying durable goods, the add-on effects from the Keynesian consumption channel should be concentrated in the places where durables are manufactured rather than showing up in the cross section.

On the savings side, changes to savings behavior will only affect the relative multiplier to the extent that savings responses are larger in states receiving larger fiscal shocks. Yet this is exactly what happened.

To find the differential effect of war spending shocks on savings behavior, I use my baseline specification for personal income but replace the personal income variable on the left-hand side with a measure of saving. The specification becomes:

$$\frac{S_{it} - S_{i,t-k}}{Y_{i,t-k}} = \beta \frac{G_{it} - G_{i,t-k}}{Y_{i,t-k}} + \alpha_i + \gamma_t + F_i * \gamma_t + \epsilon_{it} \tag{1.10}$$

where S is a measure of saving (per capita, in thousands of 1942 dollars). Specifically, S is either combined time and demand deposits of individuals, partnerships, and corporations,[43] *or* series E war bond purchases.[44]

The effects of war production spending on saving are summarized in Table 1.10. (The baseline result for personal income is also included for comparison.) These results imply that if war production spending increased by 10% of income in a given state (relative to other states), deposit holdings in that state grew by 1.1% of income relative to trend, and purchases of series E war bonds grew by another 0.3%.

These responses may seem small, but so was the total effect of war spending on income. When war production spending increased by 10% of income in a given state, that state's income grew by only 2.5% relative to trend. Together, deposits and E-bond purchases account for 1.4 percentage points of that income growth, or just under 60% of total income growth attributable to war spending.[45] Moreover, while bank deposits and series E war bonds were major savings vehicles, they were far from the only savings vehicles used during the war. These estimates do not cover postal savings or purchases of either other war

[43] This measure of saving is constructed from combined balance sheets of all active banks in the United States, reported for December 31 each year in the annual *Reports of the Comptroller Comptroller of the Currency*. I digitized portions of these publications from 1939–1948. This measure excludes all inter-bank deposit holdings and government deposit holdings (US, state, and local). Ideally, this measure of saving would include postal savings, but postal savings are excluded because they are not reported separately from US government deposit holdings in some years.

[44] E-bond purchases are reported by state in the *Monthly Treasury Bulletins* during 1942–1947. Series E, F, and G bonds were all war bonds. E-bonds could only be purchased by individuals, while F and G bonds could be purchased by institutional investors.

[45] Because E-bonds could only be purchased by individuals, these responses can be added without fear of double counting.

Table 1.10: Response of Saving and Tax Payments to War Spending Compared to Personal Income

	Personal Income	Savings Deposits	E-Bonds	Income Taxes
β	0.248***	0.115***	0.0277***	0.0429***
	(0.0526)	(0.0362)	(0.00810)	(0.0112)
Observations	294	245	294	196
Within R-squared	0.791	0.481	0.952	0.971

Standard errors clustered by state. Regressions weighted by states' 1939 populations. Time fixed effects and interaction between time fixed effects and pre-war agricultural employment shares estimated but not shown.
*** p<0.01, ** p<0.05, * p<0.1

bonds or other government securities[46] nor do they account for the wartime increase in homeownership (documented by Fetter, 2016). The total cross-sectional response of savings to war production is certainly large in relation to the income response.

One possible explanation for why saving may have increased by more in places that saw more war spending is the salience of wartime saving programs. As discussed above, war bonds were a major form of household saving during the war. The government instituted a payroll deduction program for war bond purchases—which is particularly remarkable given that income tax withholding in payrolls was also introduced during World War II. But it was up to individual firms to participate in these programs. Anecdotally, the participants appear to have been larger firms—which were also particularly likely to have defense contracts. Wartime annual reports from Boeing, Chrysler, Douglas, and General Motors all discuss employee participation in payroll deduction programs for war bond purchases. At General Motors, 95% of employees were enrolled in the payroll deduction program in both 1943 and 1944. The public finance literature on salience and retirement savings choices[47] suggests that the payroll deduction program could have increased the salience of saving, potentially with spillovers to other savings vehicles.

Outside of savings responses, increases in federal income tax payments account for a further share of the income response to war spending. Personal income is a pre-tax measure of income, so when thinking about the allocation of personal income increases one should consider taxes as well as consumption and saving. Federal income taxes are progressive, so taxes tend to increase more in places where income growth is faster. Congress enacted large federal income tax increases during World War II to help finance the war, reinforcing

[46]Bond series E, F, and G were all designated as war bonds. Series F and G bonds were available for purchase to institutional investors as well as individuals, while only individuals could purchase E-bonds. While E-bond sales accounted for the majority of war bonds, they totaled less than one-fifth of all purchases of government securities over 1941–1945.

[47]For example, see Duflo et al. (2006).

the effects of progressive taxes. In a relative sense these effects were modest: when war production spending increased by 10% of income in a given state, the relative increase in federal income taxes was 0.4% of income.[48] While small in an absolute sense, this differential increase in taxation accounts for a further 17% of the income response to war spending.

1.9 The Big Picture: An Accounting Exercise

In previous sections I have discussed several key factors affecting the size of the multiplier during World War II. Since all of these factors affected the multiplier at the same time, it is helpful to understand how they fit together.

The literature suggests that the open-economy relative multiplier with neutral monetary policy (the most relevant comparison for World War II, as explained in Section 1.4) is approximately 1.6 during normal times.[49] For purposes of comparison, this estimate is shown in the first column of Figure 1.14.

In even the most basic models of the multiplier, the size of the multiplier depends on the marginal propensity to consume. I can bound the relative mpc using a basic identity: together, savings, consumption, and taxes must account for the total income increase due to war production spending. The findings presented in Section 1.8 suggest that increased savings and taxes account for about 75% of the relative increase in income caused by war production spending, and potentially more, depending on relative responses to war spending through other savings vehicles and forms of taxation. The add-on effect of the Keynesian multiplier applies *only* to the portion of income increases that are consumed. While state-level data on wartime saving is too incomplete to draw precise conclusions, the relative mpc appears to be between 0 and 0.25, suggesting that the consumption add-on effect during World War II was significantly smaller than has been found for other periods.

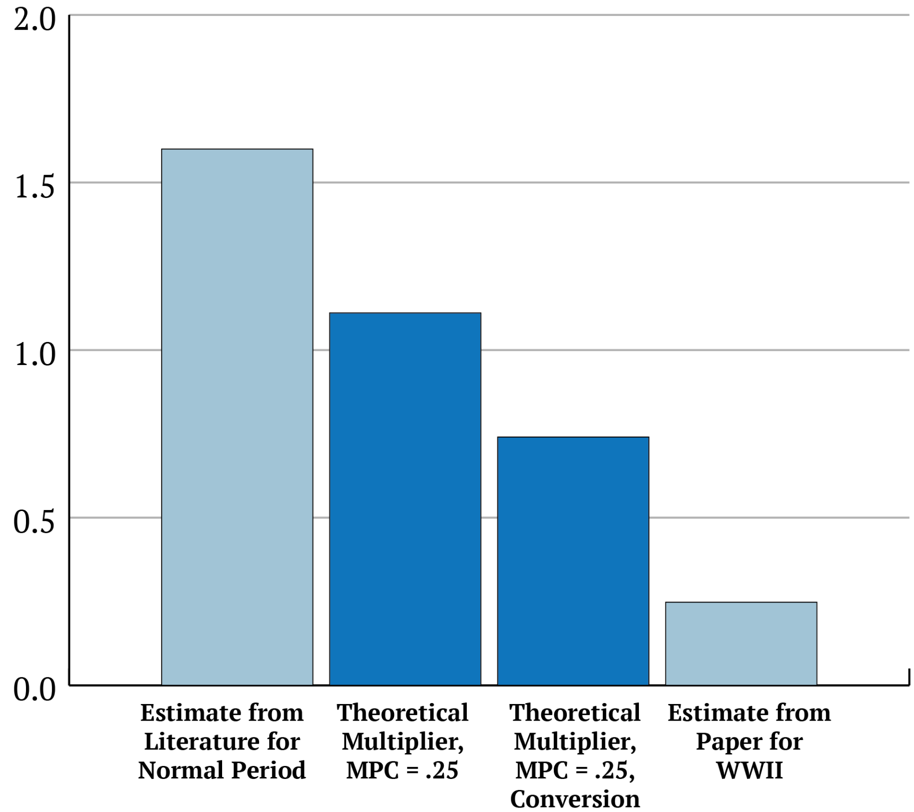
The second column of Figure 1.14 shows how the theoretical relative multiplier is affected by an unusually small mpc. Specifically, I compute a simple aggregate multiplier with an mpc of 0.25 ($1/(1 - mpc) = 1/.75 = 4/3$). As discussed in Section 1.4, the best evidence is that the aggregate multiplier should be about 20% larger than the open economy relative multiplier given the economic conditions prevailing during World War II. Accordingly, I scale the multiplier estimate downwards (dividing by 1.2) to account for the relative multiplier generally being smaller than the aggregate multiplier due to consumption spillovers and relative price effects. This produces a theoretical relative multiplier of 1.11, already considerably smaller than the "normal" relative multiplier estimate of 1.6.

In Section 1.7, I argued that conversion of manufacturing capacity reduced the initial effects of war production spending on the economy. Both the aggregate statistics and the

[48] I construct data on total federal income taxes per capita by state using the historical *Statistics of Income* for 1939–1946 reported by the IRS. Figures for 1942 are excluded because income tax filings (what the SOI data actually records) were dramatically different from taxes paid in that year due to changes in the tax code and a transition in which year of income was taxed.

[49] This estimate comes from Nakamura & Steinsson.

Figure 1.14: Comparing Theoretical and Empirical Estimates of the Relative Multiplier



relative estimates shown in Section 1.7 suggest that conversion reduced the initial stimulative impulse of spending by about 1/3. That is, when the government spent \$1 on war production, about 33¢ of that dollar directly displaced preexisting civilian economic activity, so the initial impulse to total economic activity was \$0.67 rather than \$1. Thus conversion reduces the size of the initial shock, depressing both the aggregate and relative fiscal multipliers proportionally. Mathematically, conversion can be interpreted as reducing the numerator of the textbook Keynesian multiplier from 1 to $1 - c$, where $c \approx 1/3$ in both the aggregate and relative cases.

The third column of Figure 1.14 shows the low-mpc relative multiplier from the second column, now scaled downward by a factor of 1/3 to account for the effects of conversion. Together, the small mpc and conversion imply a theoretical relative multiplier of 0.74. This is less than half the empirical estimate of the relative fiscal multiplier for "normal" times, though still considerably larger than this paper's estimate of 0.25 for World War II. Together, the small mpc and conversion explain about 60¢ of the difference between Nakamura & Steinsson's estimate of the relative multiplier for the post-war period and my estimate of the relative multiplier for World War II.

There are several possible factors that could account for the remaining difference between the observed multiplier for World War II and the combined theoretical effects of a high marginal savings rate and the costs of conversion. None of these factors are possible to measure directly (at least with the data used in this paper).

First, as discussed in Section 1.3, if there was systematic subcontracting from states receiving large amounts of war contracts to states receiving fewer war production contracts, the multiplier estimate (constructed from contract data) would be biased downwards. Second, if there were large spillovers between states, those spillovers would show up in the time fixed effects, also causing a downward bias in the estimated multiplier. This interpretation is certainly possible, as the estimated effects of war production spending are much larger when time fixed effects (and their interactions) are omitted,[50] but without time fixed effects identification is at best highly questionable, so these results imply only that this possibility should not be ruled out. Third, inefficiencies and misallocation in the wartime economy may have been systematically larger in places that received larger proportions of war production spending, which would counterbalance the stimulative effect of spending shocks in those areas.

1.10 Conclusions

What have we learned about the U.S. economy during World War II?

First, while the popularly accepted narrative that World War II helped the U.S. economy rebound from the Great Depression may be true in the broadest sense, it is true because the spending increase was so massive, not because the fiscal multiplier on wartime spending was large. War production itself appears to have been fairly inefficient as economic stimulus due to the constraints on civilian economy resulting from the war effort. The reality is that the real goal of war production—winning the war—necessitated economic policies that constrained the civilian economy, even as government spending increased dramatically. The net result was economic growth, but the immediate effects were much more modest than those claimed by popular imagination.

Second, key dynamics of the wartime economy deviate from standard macroeconomic models in important ways. Conversion replaced non-war production, especially in sectors previously producing durable goods. Entire industries converted to war production for the duration, a reality at odds with models in which both consumers and government purchase overlapping continuums of goods. Moreover, wartime savings behavior—likely motivated in part by constraints on consumption—had significant macroeconomic consequences. The presence of these constraints causes significant deviations from the predictions of standard models. World War II should be understood as a special case in many respects.

What have we learned about fiscal stimulus more generally?

[50] Without time fixed effects, the baseline estimate of the multiplier on personal income is $\beta = .528$ (*SE = 0.062, P < .001*) for the weighted regression and $\beta = .636$ (*SE = 0.074, P < .001*) for the unweighted regression.

Empirical estimates of the fiscal multiplier derived from samples including World War II will be biased downwards due to the unusual economic conditions that prevailed during wartime. Because the scale of war spending was so large, observations from World War II will be highly leveraged in any standard estimation procedure. But conversion and the associated rationing of strategic materials do not apply to other periods in U.S. history, except the final years of World War I, and even then the constraints on the civilian economy were much less severe. Since these constraints significantly altered the economy's response to military spending shocks during World War II, these observations are not directly comparable to those from other periods.

Perhaps more importantly, World War II provides evidence that savings behavior—possibly driven by durables consumption—plays an important role in the economy's response to economic stimulus. In macro models of fiscal stimulus, consumers are usually assumed to have stable consumption preferences conditional on current income, expected future income, and the relevant interest and discount rates. But in practice, large macroeconomic shocks can be accompanied by shocks to savings and consumption even conditional on those variables. World War II provides an example in which a shock to savings behavior works against the usual response of consumption to an increase in income: either due to patriotism or because goods are not available, consumers reduce their consumption as a share of income even as their incomes increase, dampening the effectiveness of fiscal stimulus.

While World War II demonstrates that a shift in savings and consumption behavior can significantly alter the effects of fiscal stimulus, war is far from the only circumstance in which such a shock to savings and consumption may accompany a fiscal shock. Financial crises often cause large increases in consumer debt burdens, which may cause a negative shock to consumption preferences, reducing the short-run effectiveness of fiscal stimulus in the wake of a financial crisis. Savings and consumption preferences seem particularly likely to shift in the wake of large macroeconomic events such as war or crisis. Attempts to gauge the effectiveness of fiscal stimulus are likely to become both more meaningful and more directly comparable when they take possible shifts in saving and consumption preferences into account.

Chapter 2

When Does Government Spending Matter? Evidence from a New Measure of Spending

2.1 Introduction

A large literature in macroeconomics studies the effect of government spending on short-run economic activity. However, economists have focused less on the question of how government spending should be measured in this context.

We tend to think of government spending as occurring at a single point in time, as recorded in the National Income Product Accounts (NIPA). This is surely good enough for many purposes (not least national income accounting), but in reality government spending is a long, complex process. For settings in which precise timing is particularly important, such as measuring the fiscal multiplier, the measurement of government spending deserves further attention.

Recent literature has suggested that the timing of government spending is very important to the measurement of the fiscal multiplier. However, rather than considering the measurement of government spending, the literature has focused on the idea that shocks to government spending may be anticipated—that is, that responses to fiscal policy are driven by shocks to expectations about future spending.

This paper takes an alternative approach to the timing of government spending shocks, focusing on the question of how—that is, when in the process—to measure government spending. Specifically, I introduce a new measure of government spending, called *budget authority*, which measures the government's commitment to spend. Budget authority measures spending when it is authorized, before funds are dispersed from the Treasury.

Like many papers in the fiscal multiplier literature, I focus on military spending because it is more plausibly exogenous to short-run economic fluctuations. Budget authority by category of spending is available from 1976 onward from the Office of Management and

Budget, including budget authority for spending on national defense. I use historical budget documents to construct defense budget authority going back to 1938.[1]

I show that budget authority produces much more precise estimates of the fiscal multiplier than the traditional NIPA measure, especially over short time horizons. Moreover, while the point estimates implied by NIPA and budget authority are fairly similar, the shapes of the impulse responses they produce are not.

The concept of spending shock reflected by budget authority is very different from the concept reflected in the measure of anticipated defense spending most commonly used in the literature, Ramey's narrative measure. Ramey's measure aims to capture changes in expected spending very broadly, and in doing so implies a model in which households and firms ignore uncertainty when optimizing. In contrast, budget authority captures changes in expectations only in a narrow sense, and implies a model in which households and firms either follow adaptive expectations about future spending or make decisions only when uncertainty is resolved.

These conceptual differences between budget authority and Ramey's measure lead to large differences in estimates of the fiscal multiplier. Using data annualized to federal fiscal years, I estimate an aggregate fiscal multiplier of 0.8 for the post-World War II period when shocks to defense spending are measured with budget authority, but a multiplier of 0.08 for the same period when using Ramey's narrative measure.

Finally, I examine the effects of shocks to each spending measure on each of the major components of output. When controls for tax changes are included, the NIPA measure of defense spending never produces statistically significant effects on either consumption or investment. In contrast, I find that shocks to budget authority produce positive, significant responses in consumption and, over a one-year time horizon, investment. Ramey's narrative measure does not produce significant consumption responses on an annual basis, but produces significant investment responses over all time horizons.

While total government spending responds strongly to all three measures, it appears that both budget authority and Ramey's measure understate direct increases in government spending because of the differences in timing between these measures and NIPA. To address this issue, I introduce a fourth measure of defense spending: budget outlays. Budget outlays are conceptually identical to budget authority, except that their timing reflects the disbursement of funds from the Treasury rather than authorizations. The timing of budget outlays is similar to the timing of NIPA, so budget outlays can be used to estimate a lower bound on the attenuation bias affecting estimates made with budget authority. When using budget outlays to correct for attenuation bias in the response of total government spending to changes in defense budget authority, budget authority implies a lower bound on the aggregate fiscal multiplier of 1.3 to 1.4.

This paper is structured as follows. Section 2.2 provides an overview of the relevant

[1] While it would not be possible to construct historical budget authority for all types of federal spending due to changes in accounting practices, these changes have minimal relevance to defense spending. See Appendix A for further detail.

literature and the motivation for this paper. Section 2.3 presents budget authority, a new approach to measuring anticipated spending, and explains how it differs from other spending measures. Section 2.4 provides baseline specifications and results. Section 2.5 examines the underlying empirical differences between the three spending measures by separately examining their effects on each major component of output. Section 2.6 concludes.

2.2 Literature and Motivation

Because government spending is often intentionally countercyclical, naïve estimates of the multiplier will be biased toward zero. A large empirical literature on fiscal stimulus uses military spending to measure the fiscal multiplier because it is less likely to be driven by short-run fluctuations in the domestic economy.[2] These papers include Hall (2009), Ramey & Shapiro (1998), Ramey (2011), Barro & Redlick (2011), and Nakamura & Steinsson (2014), among many others.

More recently, the fiscal stimulus literature has turned towards using panel data to construct cross-sectional estimates of the relative or local multiplier.[3] The advantage of this panel data approach is that it allows for far clearer identification and much more precise estimates than the traditional time series approach. However, this precision comes at a significant cost. The relationship between the aggregate multiplier and the relative or local multiplier varies systematically based on monetary policy responses, as argued persuasively by Nakamura & Steinsson (2014), making it difficult to recover precise estimates of the aggregate multiplier from even the most precise estimates of the relative multiplier.

In many circumstances, the ultimate object of interest is the aggregate fiscal multiplier, since it measures the government's ability to stimulate the economy as a whole. While estimating relative and local multipliers is surely important to our understanding of the multiplier process, improving estimates of the aggregate multiplier is still a valuable exercise.

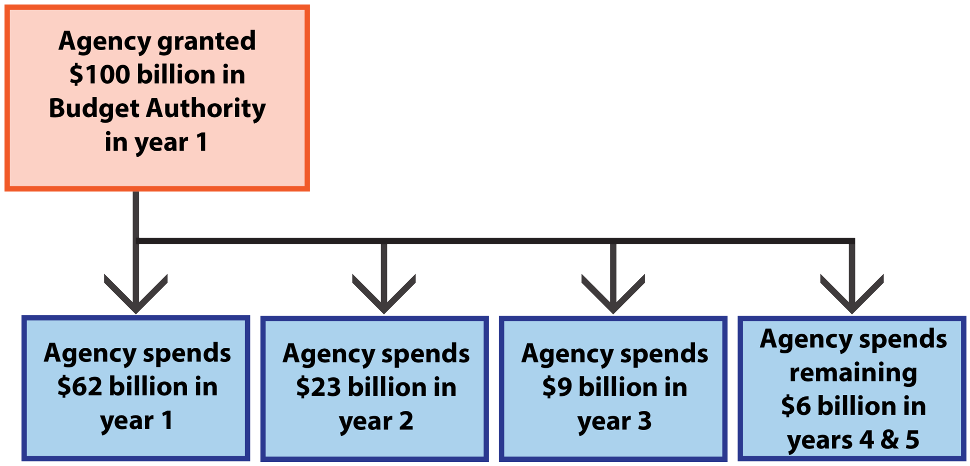
In recent years, a number of papers in the fiscal multiplier literature have focused on anticipation effects, or the idea that responses to fiscal shocks may come before the shocks themselves are observed in the data. The best known papers in this literature are Ramey & Shapiro (1998), Ramey (2011), and Mertens & Ravn (2012). Implicitly or explicitly, all of these papers assume that anticipation effects work through an information channel, with firms and households receiving news and acting upon it before government spending occurs.

This paper introduces another explanation for anticipation effects in the fiscal stimulus literature: significant time lags between the point in time at which government spending affects economic activity and the point in time at which government spending is measured.

[2]For a detailed argument on this point, see Hall (2009).

[3]The distinction between local and relative multipliers depends on the unit of observation. Research using state-level panel data tends to use the term relative multiplier, while papers using county-level data tend to discuss local multipliers.

Figure 2.1: How Does Budget Authority Translate into Spending?



Illustrative example uses aggregate spend out rates for defense implicit in CBO's May 2013 budget forecast. Spend out rates for major shocks likely to be even slower.

Economists often approach government spending as if it occurred at a single moment in time.[4] 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 dollar spent must appear precisely once as government spending in the NIPA. In reality, 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 may be 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.

2.3 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.[5] Budget authority directly measures those authorizations.

[4] A notable exception 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.

[5] I will sometimes say "obligations" to refer to [budget] authorizations, and especially may use the verb "obligate" instead of "authorize." For my purposes these terms are essentially synonymous.

Table 2.1: Budget Authority Predicts NIPA Defense Spending

	Nominal	Real	Share of GDP
budget authority (contemporaneous)	0.583***	0.360***	0.390***
	(0.124)	(0.0883)	(0.0403)
budget authority (lagged 1 year)	0.312*	0.442***	0.372***
	(0.182)	(0.124)	(0.0512)
budget authority (lagged 2 years)	0.0752	0.107	0.106
	(0.184)	(0.115)	(0.0838)
budget authority (lagged 3 years)	0.228*	-0.0540	-0.00560
	(0.119)	(0.0801)	(0.0868)
Observations	75	75	75
R-squared	0.992	0.841	0.929

Outcome is always defense spending as measured by NIPA.
*** p<0.01, ** p<0.05, * p<0.1

Budget authority tracks very closely to actual spending over time—i.e. the commitment to spend leads and predicts the eventual transfer of funds. Budget authority tends to lead outlays (the dispersement of funds from the Treasury) by about a year, and shocks to budget authority are generally sharper than shocks to outlays. Over time, however, $1 of budget authority seems to translate fairly directly into $1 of actual spending. (See Table 2.1 and Figures 2.1 & 2.2.)

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[6] At the federal level, budgeting has always been done on an annual basis. Budget authority is granted for the federal fiscal year; it does not need to be spent by the end of the year, but it must be obligated by the end of the year or it is returned to the Treasury's general fund. It is similar to but broader than appropriations.[7]

Annual budget authority totals are available from OMB for 1976 onward, disaggregated by category of spending. I constructed a historical series of budget authority for 1938-1975 using the Budgets of the United States, which have been issued annually since 1923.[8] For a detailed description of the data construction, see Appendix A.

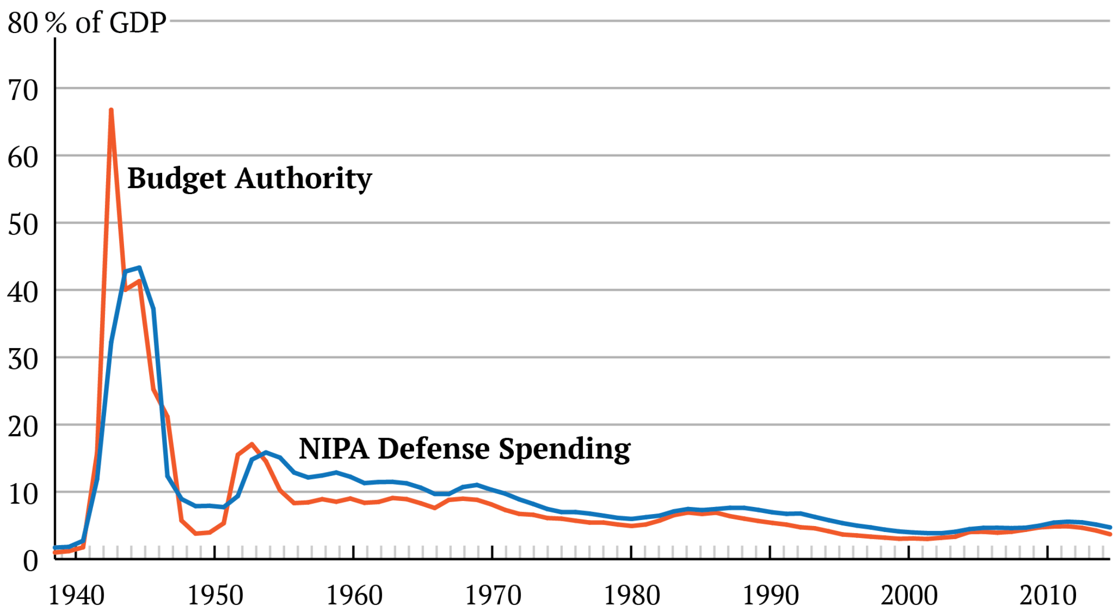
To place budget authority in context, it is important to understand how it differs conceptually from other common measures of defense spending.

[6]See the Appropriations Committee history published in 2010.

[7]Unlike appropriations, budget authority extends to the entirety of the federal budget, including mandatory spending, which is not controlled by the annual appropriations process. For the appropriated portions of the budget, budget authority generally matches appropriations, except that budget authority captures rescissions (i.e. when previously appropriated funds that remain unspent are rescinded). Thus budget authority is slightly less likely to overstate eventual spending than appropriations.

[8]I begin my series in 1938 because prior to the New Deal budget totals were regarded as upper bounds rather than as targets, so budgetary information cannot be understood in the same way.

Figure 2.2: Defense Spending Measured with Budget Authority & NIPA as Shares of GDP



Source: author's calculations using data from BEA, OMB Historical Tables, Budgets of the U.S. Government

How Does Budget Authority Differ from Defense Spending as Measured in the NIPA?

The most common measure of government spending used by economists is from the National Income and Product Accounts (NIPA). NIPA estimates of defense spending are constructed using one of three methods, depending on the type of spending.[9]

Spending on many types of military goods, including purchases of aircraft, missiles, vehicles, and petroleum products, is recorded by the direct pricing method whenever possible. Under the direct pricing method, the timing of spending reflects the delivery of goods to the military. As explained by the Bureau of Economic Analysis:

> An example of the direct-pricing method is the derivation of the F-15 aircraft estimates. The F-15 aircraft is usually purchased by DOD [the Department of Defense] as component parts. The components—such as engines, airframes, guns, and various electronic subsystems—are supplied as government-furnished equipment to the airframe contractor who performs the final integration and assembly. Quarterly estimates of the value of component parts are obtained

[9] For a complete discussion of these methods and their applications, see BEA 2005, pp. II: 33-45.

by multiplying the quantity delivered to DOD by the price paid by DOD. The estimates for each component, including integration and assembly, are summed.[10]

The other two methods for measuring military spending, the ratio method and direct estimation, both rely on the disbursement of funds from the Treasury—that is, when contractors are paid.

Since budget authority captures authorizations for spending, it necessarily leads disbursements, since contracts cannot be made without prior authorization and payments cannot start until contracts are granted. Deliveries often come later, sometimes much later (though in many cases final disbursements are made after delivery).

If government spending were a quicker process, these distinctions would not matter empirically. However, government spending is often quite slow in practice, especially when it comes to complex, capital-intensive goods like aircraft carriers, which take many years to build.

Figure 2.2 shows defense spending over time measured with both budget authority and NIPA. Two things should stand out in this figure. First, there is an extraordinarily close relationship between the two measures of spending over time. Second, shocks generally appear in budget authority before they can be seen in NIPA, and are often more sharply defined in budget authority. This is most evident in World War II and the Korean War, when the shocks were particularly large, but the same pattern can be seen in the Vietnam buildup in the mid-to-late 1960s and again with the Reagan buildup in the early 1980s. Given the spending process shown in Figure 2.1, it is reasonable to conceptualize NIPA as a noisy moving average of past budget authority. So by construction, shocks to NIPA will almost always lag shocks to budget authority.

Outside of timing differences, however, the two measures are extremely close in what they measure. First, the government cannot spend without authorization, and budget authority accounts for all types of authorizations (including for so-called mandatory spending programs, such as Social Security, though defense spending is funded through the discretionary appropriations process). Second, budget authority differs from appropriations (the authorizations for spending) in that when appropriations are not used—i.e. when spending is authorized but then does not materialize—and the authorization for spending expires, budget authority is adjusted downwards to include only the authorizations for spending that were actually used.

There are small differences between how spending is recorded in the budget and in the NIPA, but most of the differences are specific to the treatment of items like intergovernmental transfers, pension contributions, revolving loan funds, and loan guarantees.[11] The only areas in which the measures diverge which significantly affect defense spending are the treatment of military pensions and the treatment of income from the sale of surplus equipment. However,

[10]BEA (2005), II: 33-34.

[11]CBO (2016) lays out the conceptual differences between NIPA and budget measures, and the corresponding adjustments accounting for the differences between them. NIPA Table 3.18B contains the same adjustments to historical data, as computed by the BEA.

these areas are fairly small relative to total defense spending; differences in timing account for a large majority of the differences between the two measures.

How Does Budget Authority Differ from Ramey's Narrative Measure?

While both budget authority and Ramey's narrative measure capture shocks to anticipated spending, they correspond to fundamentally different concepts of spending shocks. In addition to more superficial differences, the two measures differ in two deep (and related) ways. Specifically, they differ in how broadly they attempt to capture changes in expectations and in their implicit treatment of uncertainty.

The goal of Ramey's narrative measure is to capture broad shifts in expectations about future spending. Her measure includes shocks to expectations about spending several years in the future (discounted using present values), as well as multi-year shocks (e.g. if spending is now expected to increase by \$20 billion next year and then stay \$20 billion higher for the next three years, Ramey's measured shock will be the PDV of that \$20 billion increase in each relevant year, summed across years). In that sense, her narrative measure is fundamentally forward-looking, as any broad-based measure of shocks to expectations must be.

In contrast, to the extent that budget authority captures changes in expectations, it does so very narrowly. Because budget authority is based on authorizations, it will not capture changes in expectations about spending several years in the future unless those expectations are reflected in current authorizations. If one thinks of budget authority as correcting for a measurement error problem in the NIPA series, budget authority is a direct measure of spending shocks rather than a measure of expectations. Alternatively, budget authority can be understood as implicitly assuming adaptive expectations for future spending: that is, the best (or at least most certain) information about spending after the coming year is information about spending in the coming year. (This is consistent with assumptions about future spending levels in the federal budget process.)

As a consequence of the differences in how Ramey's narrative measure and budget authority treat changes in expectations, the two measures also make very different implicit assumptions about uncertainty. Expectations about future discretionary spending levels rarely contain a high degree of certainty, due to the annual legislative process which determines spending. Moreover, changes to expectations about future spending often involve either additional uncertainty or the resolution of uncertainty. Thus any broad measure of changes in expectations must reflect some view of uncertainty, whether implicitly or explicitly.

Ramey's narrative measure directly measures changes in *Business Week*'s best guesses about the path of future spending. Thus she implicitly assumes that agents ignore policy uncertainty when making decisions based on their expectations about changes to future government spending. However, some of the shocks in her series reflect much higher degrees of uncertainty than others. For instance, some shocks reflect changes that have already been incorporated into appropriations (and therefore also show up in budget authority). Since the government has already committed to those changes in spending, those near-term shocks are very certain. However, her measure also incorporates changes to expectations about spending several years in the future, such as beliefs about when Nazi Germany will fall (a key factor in several of Ramey's World War II shocks) and the length of American ground presence in Vietnam. In Ramey's narrative measure, the only difference between these two types of shocks is discounting to reflect present values (generally small adjustments). The second type of shock involves much higher levels of uncertainty than the first, but they are essentially treated the same. Implicitly, this reflects a view that households and firms are forward-looking and optimize based on the best information they have at any time, ignoring uncertainty.

Budget authority implicitly handles uncertainty in a very different way. Because it measures actual authorizations, budget authority measures changes to spending more or less at the moment at which uncertainty is resolved. Rather than treating all shocks to expected future spending identically regardless of associated uncertainty, budget authority essentially ignores non-adaptive shocks to expectations unless and until uncertainty has been resolved (but incorporates all realized shocks into adaptive expectations of future spending). In that sense, budget authority is far more conservative about lending credence to shocks to expectations.

Which measure is preferable depends on beliefs about how households and firms respond to uncertainty. Ramey's measure assumes that households and firms are forward-looking and make decisions based on the best information they have about the future, ignoring uncertainty. In contrast, budget authority reflects an assumption that household and firms care about certainty and put off decisions until uncertainty is resolved. Notably, the timing of shocks associated with major wars is quite similar under both measures, though the correlation between the two series is only 0.55.

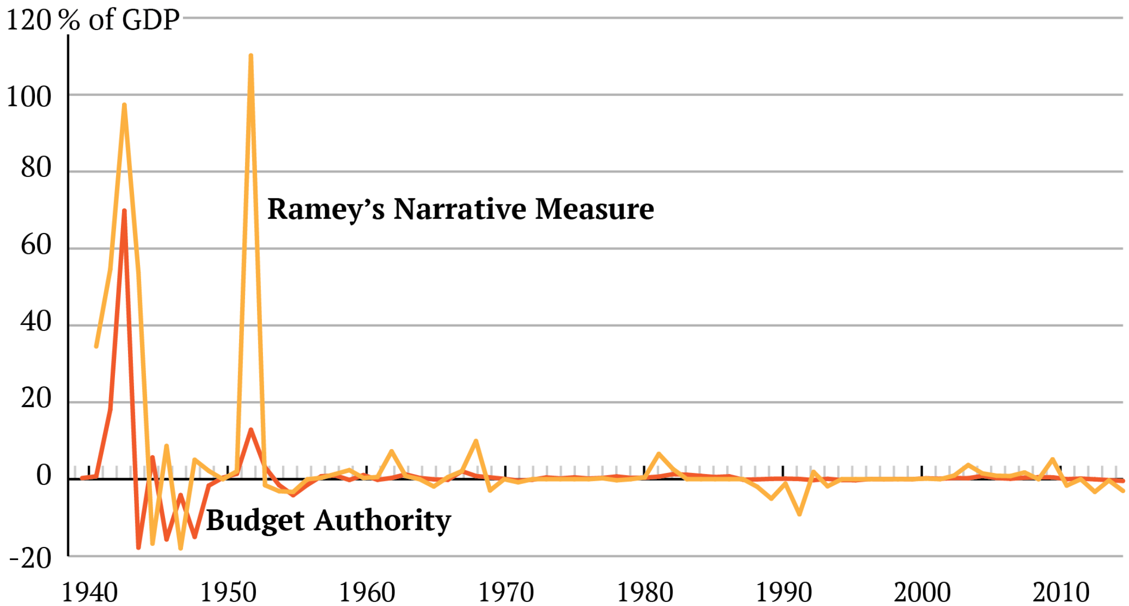
Figure 2.3 shows shocks to military spending as measured by both budget authority and Ramey's narrative measure. Three things stand out in this figure.

First, outside of the drawdown from World War II, it is striking how closely the two measures match on timing. Both the U.S. entrance into World War II and the start of the Korean War show up at the same time. This is notable because these shocks appear a year or two later in NIPA (as shown in Figure 2.2). Despite the conceptual differences between them, the timing of major shocks is very similar in Ramey's narrative measure and budget authority.

Second, there are more significant shocks in Ramey's series than in budget authority. There are several reasons for this. One is that Ramey includes shocks to spending on NASA and non-military foreign aid, both of which are excluded from defense budget authority because they are not classified as defense spending in the federal budget. Thus Ramey's measure includes shocks relating to the space race while budget authority does not, and so forth. In addition, because Ramey's measure captures broader changes to expectations, some shocks were reversed before they were fully realized, and so never appear as large shocks in budget authority.

Finally, shocks in Ramey's series are usually (though not always) larger than shocks to

Figure 2.3: Shocks to Budget Authority and Ramey's Narrative Measure, as Shares of GDP



Source: author's calculations using data from BEA, OMB Historical Tables, Budgets of the U.S. Government, Ramey's published data appendix

budget authority, sometimes (as with the start of the Korean War) by a large margin. Where the two measures capture the same underlying shocks to spending, this is usually due to the way in which Ramey accounts for the expected persistence of the shock. Conversely, the magnitude of shocks matches more closely where Ramey does not assume persistence in her measure.

2.4 Baseline Results

Having clarified the conceptual differences between budget authority and other standard measures of defense spending, I now examine how the choice of spending measure influences empirical estimates of the fiscal multiplier.

My baseline specification follows Ramey & Zubairy (forthcoming). Because budget authority is an inherently annual measure, all data (including other measures of government spending, for comparison, and GDP) is aggregated to fiscal years.[12] The benefit of this

[12]Before 1947, quarterly data on GDP and government spending is not available, so calendar year estimates are used. From 1947 through 1976, the fiscal year began with the third quarter (in July) and lasted through the end of June (the end of the second quarter). Later the fiscal year was changed to start with the fourth quarter on October 1 and last through the end of the third quarter. See the Appendix for details.

Table 2.2: Baseline Fiscal Multiplier Estimates Using Budget Authority

	one-year horizon	two-year horizon	three-year horizon	# Obs
Without Controls for Tax Changes				
Post-WWII Sample (1948 - 2013)	0.747***	0.609***	0.596***	65
	(0.108)	(0.0940)	(0.162)	
Post-Korea Sample (1955 - 2013)	2.036***	1.734***	1.371	58
	(0.376)	(0.547)	(0.917)	
Full Sample (1939 - 2013)	0.173**	0.426***	0.535***	74
	(0.0748)	(0.0418)	(0.0294)	
With Controls for Tax Changes				
Full Sample (1946 - 2007)	0.634***	0.805***	0.851***	60-62
	(0.0800)	(0.124)	(0.204)	
Post-Korea Sample (1955 - 2007)	1.923***	1.446**	0.848	50-52
	(0.423)	(0.633)	(0.972)	

All estimates use Newey-West standard errors with 4 lags. AIC and BIC suggest that anywhere from 2 to 4 lags is optimal, depending on specification, so 4 lags are allowed in all estimates for purposes of consistency. Tax change variable is taken from Romer & Romer (2010) and is available for 1946 - 2007. Tax variable includes shocks classified as exogenous by Romer & Romer, plus shocks classified as driven by changes to defense spending. *** p<0.01, ** p<0.05, * p<0.1

specification is that the fiscal multiplier (and hence its standard error) is estimated directly. I estimate:

$$\sum_{i=0}^{h} \tilde{y}_{t+i} = \alpha + \beta_h \sum_{i=0}^{h} \tilde{g}_{t+i} + \gamma_1 \tilde{y}_{t-1} + \gamma_2 \tilde{g}_{t-1} + \epsilon_t \tag{2.1}$$

where

$$\tilde{y}_t = \frac{y_t - y_{t-1}}{y_{t-1}} \text{ and } \tilde{g}_t = \frac{g_t - g_{t-1}}{y_{t-1}}. \tag{2.2}$$

By summing shocks to both output and government spending on both sides, it is possible to estimate the multiplier directly over longer time horizons while still updating the denominator to allow for short-run changes to output driven by factors other than changes to defense spending.

In this approach, β_h estimates the aggregate fiscal multiplier at a time horizon of $h+1$ years. I estimate β_h for $h \in \{0, 1, 2\}$, or time horizons of 1, 2, and 3 years. I also include controls for lagged changes to both output and military spending.

While it is not possible to control for an extensive list of possible omitted variables in a time series analysis, one of the most important omitted variables is changes to taxes, as tax changes are often legislated to offset the effects of shocks to spending. I control for tax changes using the measure provided by Romer & Romer (2010). Specifically, I include those tax shocks they categorize as exogenous and tax shocks which they find to have been driven by changes in defense spending. When tax changes are included, the specification becomes:

$$\sum_{i=0}^{h} \tilde{y}_{t+i} = \alpha + \beta_h \sum_{i=0}^{h} \tilde{g}_{t+i} + \eta_h \sum_{i=0}^{h} \tilde{\tau}_{t+i} + \gamma_1 \tilde{y}_{t-1} + \gamma_2 \tilde{g}_{t-1} + \epsilon_t \tag{2.3}$$

where

$$\tilde{\tau}_t = \frac{\textit{Romer tax shock}}{y_{t-1}} \tag{2.4}$$

Table 2.2 shows estimates of the aggregate fiscal multiplier computed with defense budget authority over several time periods, with and without controls for tax changes. Estimates of the fiscal multiplier are positive and statistically significant at (at the 95% level or higher) for all sub-samples and time horizons, with the exception of the three-year time horizon in the post-Korean War period.

Additionally, when defense spending is measured using budget authority, the fiscal multiplier appears to be smaller when World War II is included in the sample. This is consistent with the findings presented in chapter one, which argues that features of the wartime economy reduced the fiscal multiplier during World War II.

While these estimates appear reasonable, it is instructive to consider how fiscal multipliers estimated using budget authority compare to estimates using other measures of defense spending.

Comparative Results: Budget Authority and NIPA

Table 2.3 shows comparable estimates of the aggregate fiscal multiplier on defense spending generated using both NIPA and budget authority, again over various time horizons, samples, and with and without controls for tax changes. The specifications are identical except for the measure of defense spending. Several patterns emerge in these estimates.

First, with NIPA there does not appear to be a strong immediate response of output to spending in the first year. That is, when NIPA is used the fiscal multiplier is much more statistically significant over longer (two- and three-year) time horizons than over the first year. The point estimates of the multiplier are much smaller over one year than over two years in most specifications, and standard errors tend to be quite large relative to point the estimates for year one. In contrast, budget authority generally shows much larger and more precise point estimates for the fiscal multiplier over a one-year time horizon. Interestingly, this relative difference in the point estimates does not persist at longer time horizons.

Second, budget authority consistently produces more precise multiplier estimates than NIPA. This can be seen clearly in Figure 2.4, which shows the estimated impulse response of output to a shock to defense spending over 1946–2007, when controls for tax changes are included in the estimation. Remarkably, the 95% confidence interval for the impulse of the fiscal multiplier measured with budget authority lies entirely within the 95% confidence interval for the impulse of the fiscal multiplier measured by NIPA, though the shape of the

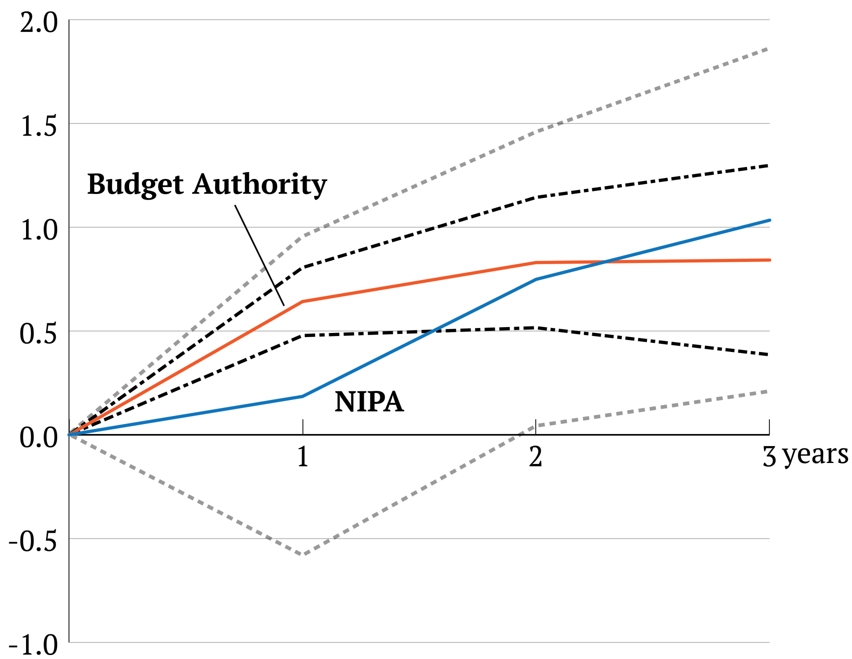
Table 2.3: Fiscal Multiplier Estimates with Budget Authority & NIPA

	Without Tax Controls		**With Tax Controls**		
	NIPA	Budget Authority	NIPA	Budget Authority	# Obs
Post-WWII Sample (1946 - 2007)					
one-year horizon	0.399	0.624***	0.185	0.642***	61
	(0.269)	(0.0517)	(0.384)	(0.0819)	
two-year horizon	0.776***	0.660***	0.749**	0.830***	60-61
	(0.188)	(0.0817)	(0.354)	(0.157)	
three-year horizon	0.902***	0.645***	1.034**	0.842***	59-61
	(0.191)	(0.106)	(0.413)	(0.228)	
Post-Korea Sample (1955 - 2007)					
one-year horizon	0.982*	2.045***	0.761	1.923***	52
	(0.546)	(0.414)	(0.541)	(0.423)	
two-year horizon	0.985*	1.653**	0.732	1.446**	51-52
	(0.556)	(0.631)	(0.592)	(0.633)	
three-year horizon	1.146	1.191	0.851	0.848	50-52
	(0.798)	(1.085)	(0.707)	(0.972)	
Full Sample (1939 - 2013)					
one-year horizon	0.331**	0.173**	–	–	74
	(0.149)	(0.0748)			
two-year horizon	0.577***	0.426***	–	–	74
	(0.120)	(0.0418)			
three-year horizon	0.698***	0.535***	–	–	74
	(0.0955)	(0.0294)			

All estimates use Newey-West standard errors with 4 lags. AIC and BIC suggest that anywhere from 2 to 4 lags is optimal, depending on specification, so 4 lags are allowed in all estimates for purposes of consistency. Tax change variable is taken from Romer & Romer (2010) and is available for 1946 - 2007. Tax variable includes shocks classified as exogenous, plus shocks classified as driven by changes to defense spending.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 2.4: Impulse Responses with NIPA and Budget Authority



Impulse responses reflect estimates for post-WWII sample with controls for tax changes (1946 - 2007). 95% confidence intervals shown.

impulse is significantly different. In light of this greater precision, it is perhaps not surprising that multiplier estimates from budget authority retain more statistical significance in the post-Korean War sample, whether or not controls for tax changes are included.

Some readers may prefer a more direct comparison between the two spending measures. Table 2.4 presents the results of horserace regressions with both NIPA and budget authority. Results are presented both with and without controls for tax shocks, and always over baseline post-World War II sample period (1946 - 2007).

At the one- and two-year time horizons, the regression clearly prefers budget authority to NIPA as a measure of defense spending. At the three-year time horizon, NIPA wins out, though neither measure looks particularly strong.

Comparative Results: Budget Authority and Ramey's Measure

Ramey's narrative measure is a series of shocks to expected spending, while budget authority and NIPA both measure spending levels and must be differenced in order to measure shocks. In order to make the spending measures comparable, $\tilde{g}_t$ must be constructed in a slightly different way for Ramey's measure. When Ramey's narrative measure is used for government

Table 2.4: Horserace Regressions with Budget Authority & NIPA

	Without Tax Controls			**With Tax Controls**		
One-year time horizon						
NIPA	0.399		-0.237	0.185		-0.236
	(0.269)		(0.248)	(0.384)		(0.259)
Budget Authority		0.624***	0.657***		0.642***	0.722***
		(0.0517)	(0.106)		(0.0819)	(0.207)
Two-year time horizon						
NIPA	0.776***		0.170	0.749**		0.198
	(0.188)		(0.207)	(0.354)		(0.219)
Budget Authority		0.660***	0.503***		0.830***	0.682***
		(0.0817)	(0.0996)		(0.157)	(0.247)
Three-year time horizon						
NIPA	0.902***		0.581*	1.034**		0.609*
	(0.191)		(0.324)	(0.413)		(0.311)
Budget Authority		0.645***	0.269		0.842***	0.458
		(0.106)	(0.222)		(0.228)	(0.368)

All estimates use Newey-West standard errors with 4 lags. AIC and BIC suggest that anywhere from 2 to 4 lags is optimal, depending on specification, so 4 lags are allowed in all estimates for purposes of consistency. Sample is always the full period for which the tax change variable is available, 1946 - 2007. Tax change variable is taken from Romer & Romer (2010). Tax variable includes shocks classified as exogenous, plus shocks classified as driven by changes to defense spending.
*** p<0.01, ** p<0.05, * p<0.1

spending, I define:

$$\tilde{g}_t = \frac{Ramey\ shock}{y_{t-1}}. \tag{2.5}$$

This follows Ramey's own specifications.

In all other respects the specifications used for Ramey's narrative measure are the same as those used for NIPA and budget authority. Table 2.5 shows comparable estimates of the aggregate fiscal multiplier on defense spending generated using both Ramey's measure and budget authority, once again over various time horizons, samples, and with and without controls for tax changes.

At all time horizons, the multiplier estimated using Ramey's narrative measure are always statistically significant at the 95% level or higher when the Korean War is included in the sample. In this sense it is very similar to budget authority. However, two major differences arise. First, Ramey's measure does not have predictive power for the post-Korean War sample. Second, the fiscal multiplier implied by Ramey's narrative measure is consistently much smaller than the fiscal multipliers implied by budget authority, often by an order of magnitude.

Table 2.5: Fiscal Multiplier Estimates with Budget Authority & Ramey's Narrative Measure

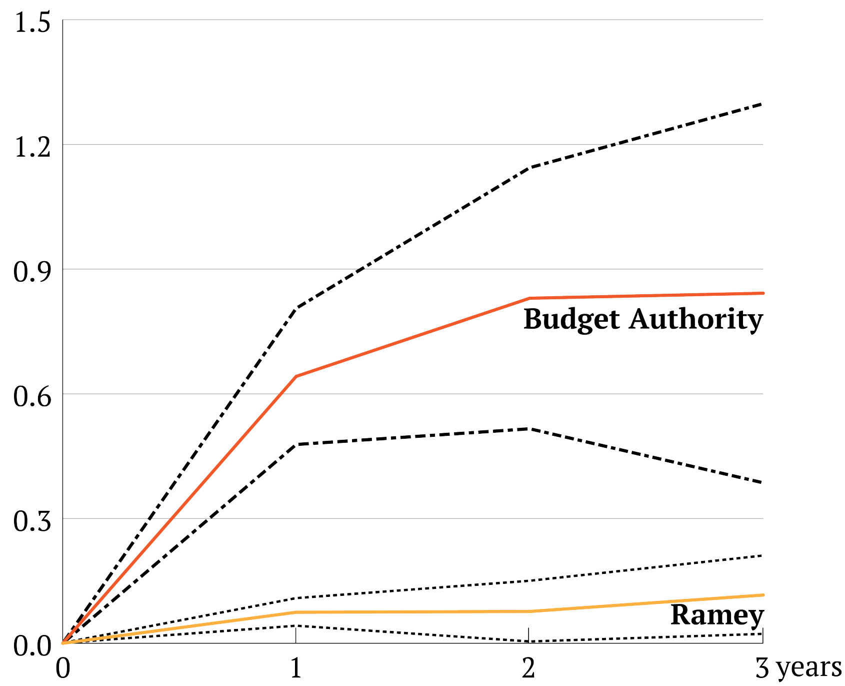
	Without Tax Controls		**With Tax Controls**		
	Ramey	Budget Authority	Ramey	Budget Authority	# Obs
Post-WWII Sample (1946 - 2007)					
one-year horizon	0.0643***	0.624***	0.0746***	0.642***	61
	(0.0117)	(0.0517)	(0.0166)	(0.0819)	
two-year horizon	0.0736***	0.660***	0.0765**	0.830***	60-61
	(0.0116)	(0.0817)	(0.0365)	(0.157)	
three-year horizon	0.0890***	0.645***	0.116**	0.842***	59-61
	(0.0145)	(0.106)	(0.0471)	(0.228)	
Post-Korea Sample (1955 - 2007)					
one-year horizon	-0.211	2.045***	-0.208	1.923***	52
	(0.154)	(0.414)	(0.146)	(0.423)	
two-year horizon	-0.127	1.653**	-0.122	1.446**	51-52
	(0.212)	(0.631)	(0.197)	(0.633)	
three-year horizon	-0.0308	1.191	-0.0486	0.848	50-52
	(0.217)	(1.085)	(0.205)	(0.972)	
Full Sample (1939 - 2013)					
one-year horizon	0.0908***	0.173**	–	–	73-74
	(0.0211)	(0.0748)			
two-year horizon	0.138***	0.426***	–	–	72-74
	(0.0325)	(0.0418)			
three-year horizon	0.167***	0.535***	–	–	71-74
	(0.0354)	(0.0294)			

All estimates use Newey-West standard errors with 4 lags. AIC and BIC suggest that anywhere from 2 to 4 lags is optimal, depending on specification, so 4 lags are allowed in all estimates for purposes of consistency. Tax change variable is taken from Romer & Romer (2010) and is available for 1946 - 2007. Tax variable includes shocks classified as exogenous, plus shocks classified as driven by changes to defense spending. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

This second point can be seen clearly in Figure 2.5, which shows shows the estimated impulse response of output to a shock to defense spending over 1946–2007, when controls for tax changes are included in the estimation. Remarkably, the impulse responses for the two measures are completely separate, implying that the two measures suggest significantly different fiscal multipliers at every time horizon.

As discussed in Secion 2.3, there are significant theoretical differences between budget authority and Ramey's narrative measure. Specifically, the two measures capture changes to expectations in very different ways, and fit very different models of how households and firms respond to uncertainty. Table 2.5 shows that these theoretical differences have significant implications for the size of the multiplier.

Figure 2.5: Impulse Responses with Ramey's Narrative Measure and Budget Authority



Impulse responses reflect estimates for post-WWII sample with controls for tax changes (1946 - 2007). 95% confidence intervals shown.

Table 2.6 presents the results of horserace regressions with both Ramey's narrative measure and budget authority. Results are presented both with and without controls for tax shocks, and always over baseline post-World War II sample period (1946 - 2007). The regression clearly prefers budget authority at all time horizons, and in several specifications with shorter time horizons the inclusion of Ramey's measure has almost no effect on the point estimate of the multiplier on budget authority.

2.5 Mechanism: Effects on Components of Output

In the previous section I showed that budget authority produces much more precise estimates of the fiscal multiplier than NIPA, with much larger responses over a one-year horizon. I also showed that budget authority produces much larger estimates of the fiscal multiplier than Ramey's narrative measure. While Section 2.3 discusses the conceptual differences between budget authority and the other two measures, so far I have not shown evidence for why or how the various measures produce such different estimates.

I now turn to the components of GDP, particularly consumption and investment, and examine how each of the spending measures affects the components of GDP. Specifically, I

Table 2.6: Horserace Regressions with Budget Authority & Ramey's Measure

	Without Tax Controls			**With Tax Controls**		
One-year time horizon						
Ramey's Measure	0.0643***		-0.00796	0.0746***		0.000990
	(0.0117)		(0.0146)	(0.0166)		(0.0233)
Budget Authority		0.624***	0.655***		0.642***	0.647***
		(0.0517)	(0.0885)		(0.0819)	(0.0898)
Two-year time horizon						
Ramey's Measure	0.0736***		-0.0302	0.0765**		-0.00601
	(0.0116)		(0.0205)	(0.0365)		(0.0267)
Budget Authority		0.660***	0.771***		0.830***	0.845***
		(0.0817)	(0.112)		(0.157)	(0.138)
Three-year time horizon						
Ramey's Measure	0.0890***		-0.00852	0.116**		0.0254
	(0.0145)		(0.0372)	(0.0471)		(0.0464)
Budget Authority		0.645***	0.681***		0.842***	0.771***
		(0.106)	(0.235)		(0.228)	(0.242)

All estimates use Newey-West standard errors with 4 lags. AIC and BIC suggest that anywhere from 2 to 4 lags is optimal, depending on specification, so 4 lags are allowed in all estimates for purposes of consistency. Sample is always the full period for which the tax change variable is available, 1946 - 2007. Tax change variable is taken from Romer & Romer (2010). Tax variable includes shocks classified as exogenous, plus shocks classified as driven by changes to defense spending.
*** p<0.01, ** p<0.05, * p<0.1

will return to my baseline specification, but replace the change in output with the change in consumption, investment, total government spending, or net exports as a fraction of output in the starting period. This scaling of components by total output simplifies interpretation of coefficients and allows for direct comparisons between these estimates and estimates of the fiscal multiplier. The baseline specification becomes:

$$\sum_{i=0}^{h} \tilde{x}_{t+i} = \alpha + \beta_h \sum_{i=0}^{h} \tilde{g}_{t+i} + \gamma_1 \tilde{x}_{t-1} + \gamma_2 \tilde{g}_{t-1} + \epsilon_t \tag{2.6}$$

where

$$\tilde{x}_t = \frac{x_t - x_{t-1}}{y_{t-1}} \tag{2.7}$$

and x is defined as either consumption, investment, total government spending, or net exports.

Table 2.7 shows the effects of shocks to all three spending measures on consumption, investment, total government spending, and net exports. Once again, the sample period is

Table 2.7: Effects on Components of GDP, by Spending Measure

	Without Tax Controls			**With Tax Controls**		
	NIPA	Budget Authority	Ramey	NIPA	Budget Authority	Ramey
Consumption						
One-year horizon	-0.320***	0.303***	0.00833	-0.345	0.378***	0.00591
	(0.110)	(0.108)	(0.00651)	(0.207)	(0.0591)	(0.0131)
Two-year horizon	-0.0676	0.207*	0.00422	0.0253	0.415***	0.00406
	(0.0878)	(0.114)	(0.00817)	(0.182)	(0.109)	(0.0253)
Three-year horizon	0.0606	0.141	0.00161	0.225	0.345***	0.0132
	(0.101)	(0.0889)	(0.00704)	(0.186)	(0.127)	(0.0295)
Investment						
One-year horizon	-0.0647	0.322***	0.0484***	-0.262	0.290***	0.0535***
	(0.265)	(0.0867)	(0.00573)	(0.209)	(0.0911)	(0.0124)
Two-year horizon	0.0730	0.146**	0.0294***	-0.0451	0.145	0.0370*
	(0.113)	(0.0613)	(0.00622)	(0.185)	(0.122)	(0.0190)
Three-year horizon	0.0649	0.0827	0.0246***	0.0403	0.120	0.0415*
	(0.0998)	(0.0567)	(0.00746)	(0.204)	(0.123)	(0.0222)
Total Government Spending (NIPA "G")						
One-year horizon	0.973***	0.112***	0.0133***	1.049***	0.140**	0.0108
	(0.0501)	(0.0422)	(0.00102)	(0.0440)	(0.0622)	(0.00716)
Two-year horizon	0.933***	0.291***	0.0443***	1.023***	0.305***	0.0475***
	(0.0395)	(0.0461)	(0.00896)	(0.0509)	(0.0825)	(0.0130)
Three-year horizon	0.936***	0.401***	0.0765***	1.010***	0.409***	0.0776***
	(0.0367)	(0.0533)	(0.00945)	(0.0587)	(0.0966)	(0.0140)
Net Exports						
One-year horizon	0.0641	-0.0520***	-0.00577***	-0.00503	-0.0620**	-0.0149***
	(0.0671)	(0.0161)	(0.00117)	(0.103)	(0.0237)	(0.00501)
Two-year horizon	-0.0438	-0.0201	-0.00623	-0.129*	-0.0538	-0.0198***
	(0.0507)	(0.0170)	(0.00382)	(0.0698)	(0.0343)	(0.00737)
Three-year horizon	-0.108**	-0.0566**	-0.0169***	-0.171**	-0.0879*	-0.0322***
	(0.0458)	(0.0223)	(0.00527)	(0.0790)	(0.0474)	(0.00897)

All estimates use Newey-West standard errors with 4 lags. AIC and BIC suggest that anywhere from 2 to 4 lags is optimal, depending on specification, so 4 lags are allowed in all estimates for purposes of consistency. Sample is always the full period for which the tax change variable is available, 1946 - 2007. Tax change variable is taken from Romer & Romer (2010). Tax variable includes shocks classified as exogenous, plus shocks classified as driven by changes to defense spending. *** p<0.01, ** p<0.05, * p<0.1

always 1946 - 2007, and results are always shown both with and without controls for tax changes.

Effects on Consumption

The effects of a defense spending shock on consumption look very different depending on how shocks to defense spending are measured. When defense spending is measured using NIPA, the effects of a spending shock are strongly negative in the first year—a drop of slightly over 3% in consumption as a share of GDP if defense spending is increased by 10% of GDP. Without controls for tax shocks, this effect is highly significant. The statistical significance vanishes when controls for tax shocks are included, but the point estimate remains essentially unchanged. Over longer time horizons, the effects of NIPA defense spending on consumption become positive, but never statistically significant.

In contrast, when budget authority is used to measure defense spending, a shock to defense spending has a strongly positive effect on consumption, both over the first year and over longer time horizons. When controls for tax changes are included, the estimated effect of a defense spending shock on consumption hovers around 0.4 across all estimated time horizons.

Ramey's narrative measure does not appear to have much power for capturing consumption responses. At all time horizons the coefficients are positive but small, and they are never statistically significant.

Effects on Investment

When considering investment, it is NIPA which appears to be lacking in statistical power. Over one year, NIPA appears to have a negative but not statistically significant effect on investment. Over time, the effect of NIPA on investment appears to be closer to zero, but the coefficients never have statistical significance.

In contrast, both budget authority and Ramey's narrative measure do much better at capturing effects of defense spending shocks on investment. Budget authority produces strong, positive effects on investment over a one-year time horizon, but the magnitude of those effects shrinks by half and they become statistically insignificant at longer time horizons. Ramey's narrative measure, on the other hand, shows significant effects on investment over a one-year time horizon which persist over longer time horizons.

If one considers the different ways in which budget authority and Ramey's narrative measure treat both expectations and uncertainty, these differences in how the two measures pick up shocks to investment are fairly intuitive. Budget authority captures immediate, certain changes: the shocks to which firms are likely to respond immediately. Ramey's measure, on the other hand, captures shocks to expectations well into the future, and shocks associated with higher levels of uncertainty. One might expect her measure to capture longer-run shifts in investment behavior for these reasons, which indeed it appears to do.

Effects on Net Exports

All measures of spending show relatively small, negative effects on net exports at all time horizons (at least when controls for tax changes are included). With NIPA these effects grow stronger and more statistically significant over longer time horizons, while with budget authority they appear strongest over a one-year horizon (but also marginally significant over three years). Ramey's measure shows small but significant negative effects on net exports at all time horizons.

Effects on Total Government Spending

It is well established that short-term fluctuations in government spending are largely driven by fluctuations in defense spending, as other types of government spending tend to grow fairly smoothly over time, especially in aggregate. Thus it is not surprising that shocks to defense spending as measured by NIPA account for shocks to total government spending on a nearly one-for-one basis.

It is much more surprising that shocks to budget authority and Ramey's measure produce such small coefficients when used to predict total government spending. For instance, over a three-year time horizon, if defense budget authority increases by 10% of GDP, total government spending would show an increase of only 4%. This relationship is even more skewed with Ramey's measure: shocks amounting to 10% of output in Ramey's measure imply an increase in government spending of less than 1% of output.

To the extent that the differences between NIPA's measure of defense spending and other measures of defense spending (budget authority and Ramey's narrative measure) reflect differences in timing and not differences in the definition of government spending, the smaller coefficients on those other measures reflect attenuation bias in the measurement of the response of total government spending. Both the BEA and the Congressional Budget Office (CBO) publish tables detailing the adjustments needed to move between the budget and NIPA estimates of total federal government expenditures, but these tables do not distinguish which adjustments affect estimates of defense spending. Appendix B explains the differences between budget and NIPA definitions of spending in detail, and outlines the specific areas in which the two measures treat defense spending differently.

The important empirical question is the extent to which the smaller coefficients on budget authority (with total government spending as the left-hand side variable) are attributable to timing differences—in which case the smaller coefficients reflect attenuation bias—versus the extent to which the smaller coefficients on budget authority reflect conceptual differences between the different measures of spending. To answer this question, I introduce a fourth measure of defense spending: budget *outlays*. Like budget authority, budget outlays are derived from the federal budget process. However, while budget authority is timed to reflect obligations, budget outlays are timed to reflect the disbursement of funds from the Treasury, so the timing of outlays is much closer to the timing of NIPA (though because they derive from different sources and methodologies—and because NIPA does not always follow the

Table 2.8: Effects on Total Government Spending: NIPA, Budget Authority, and Budget Outlays as Measures of Defense Spending

	Without Tax Controls			**With Tax Controls**		
	NIPA	Budget Authority	Outlays	NIPA	Budget Authority	Outlays
Total Government Spending (NIPA "G")						
One-year horizon	0.973***	0.112***	0.712***	1.049***	0.140**	0.740***
	(0.0501)	(0.0422)	(0.0776)	(0.0440)	(0.0622)	(0.104)
Two-year horizon	0.933***	0.291***	0.804***	1.023***	0.305***	0.883***
	(0.0395)	(0.0461)	(0.0489)	(0.0509)	(0.0825)	(0.105)
Three-year horizon	0.936***	0.401***	0.825***	1.010***	0.409***	0.910***
	(0.0367)	(0.0533)	(0.0452)	(0.0587)	(0.0966)	(0.114)

All estimates use Newey-West standard errors with 4 lags. AIC and BIC suggest that anywhere from 2 to 4 lags is optimal, depending on specification, so 4 lags are allowed in all estimates for purposes of consistency. Sample is always the full period for which the tax change variable is available, 1946 - 2007. Tax change variable is taken from Romer & Romer (2010). Tax variable includes shocks classified as exogenous, plus shocks classified as driven by changes to defense spending. *** p<0.01, ** p<0.05, * p<0.1

disbursement of funds, as explained in Section 2.3—even NIPA and budget outlays may not match precisely in their timing).

Because differences between budget authority and outlays are completely attributable to the differences in timing between them, and because outlays are much closer in timing to NIPA than to budget authority, outlays can be used to establish a lower bound for the role of timing differences in driving the small coefficients on budget authority. Table 2.8 presents coefficients on defense outlays when that measure is used to predict the response of total government spending.[13] The corresponding coefficients estimated using NIPA and budget authority are reproduced from Table 2.7 for purposes of comparison.

At all time horizons, the coefficient on budget outlays is always larger than the coefficient on budget authority by at least 0.5. Since the difference between budget authority and outlays is one of timing, this implies that budget authority significantly understates the effects of government spending due to attenuation bias. Note that while the timing of outlays is significantly closer to the timing of NIPA than budget authority, the timing of NIPA and budget outlays may not match exactly, so the coefficients on outlays should be understood as lower bounds for the coefficients one would estimate using budget authority if it were possible to correct completely for the timing differences between NIPA and budget authority.

It is more difficult to put a meaningful bound on the role of attenuation bias in producing small coefficients when estimating the effects of Ramey's narrative measure on total govern-

[13]Complete estimates using budget outlays are not shown, but are generally quite similar—though not identical—to the results using NIPA.

ment spending, not least because Ramey's measure is constructed as a series of shocks and so not directly comparable to the NIPA. However, given that the timing of major shocks in Ramey's measure is closer to budget authority than to NIPA, it seems reasonable to assume that attenuation bias is non-trivial in this case as well.

Given the scaling of components by total output (see equation 2.6), one can construct an estimate of the fiscal multiplier by summing the coefficients on each major component of output. Indeed, doing so produces estimates of the fiscal multiplier very close to those presented earlier in the paper: using budget authority, this approach produces fiscal multipliers of 0.75 to 0.8 over the post-World War II period when tax controls are included, versus comparable estimates of 0.65 to 0.85 reported in Tables 2.3 and 2.5.

One can also use this approach to recalculate the fiscal multiplier with corrections for attenuation bias in the response of total government spending to changes in defense spending. Substituting the coefficients on outlays for the coefficients on budget authority from Table 2.8—that is, correcting for the timing differences between budget authority and the dispersement of funds when measuring the response of total government spending to defense shocks—implies a fiscal multiplier of 1.3 to 1.4 (with the responses to all other components of spending measured with budget authority).

Even this estimate can be understood as a lower bound, given that the timing of budget outlays do not match NIPA perfectly. Outside of timing differences, there is no clear reason to assume that the conceptual differences between NIPA and budget concepts of spending should cause total government spending to increase by either more or less than \$1 in response to a \$1 shock in defense budget authority, especially if the identifying variation in defense spending is not driven by either military pensions or sales of surplus equipment (the areas in which accounting differs between NIPA and budget measures). If one believes that the "true" response to a \$1 shock in defense budget authority is \$1 of total government spending, budget authority implies an aggregate fiscal multiplier of 1.4 to 1.6, depending on the time horizon.

2.6 Conclusion

A longstanding puzzle in the empirical literature on the aggregate fiscal multiplier has been the question of why empirical estimates of the multiplier almost always fall below one, even when many models suggest that the multiplier should be (perhaps modestly) above unity. The above analysis suggests that correctly estimating the fiscal multiplier is, indeed, "all in the timing."

Moreover, Section 2.5 suggests that the appropriate measure—and implicitly timing—of spending may differ systematically for each set of agents. Budget authority was the only measure to produce positively statistically significant responses in consumption, suggesting that the timing of authorizations and concept of spending shocks reflected in budget authority are a good match for consumer behavior. In contrast, budget authority produces strong investment responses over a short (one-year) time horizon—perhaps reflecting intertemporal shifting in firms' choices—while Ramey's narrative measure, with its broader estimates of shocks to expectations, seems to do a better job of capturing firms' investment decisions over longer time horizons. Net exports respond in similar ways under all three measures of spending, suggesting that precise timing is less important for that channel. Finally, government spending itself responds most strongly to shocks as measured by NIPA.

Because the timing of major shocks as measured by NIPA is so different from the timing of budget authority and Ramey's narrative measure, no one measure of government spending can produce an estimate of the fiscal multiplier that does not suffer from attenuation bias due to incorrect timing. In some sense every measure faces a Catch-22: if the timing of a measure correctly captures consumption and investment responses, the measure will systematically understate the response of total government spending. Conversely, if the timing of a measure correctly captures the response of total government spending, it will systematically miss the responses of consumption and investment.

One way around this problem is to use a measure such as budget authority to correctly measure the responses of other components of output, and then to correct for bias in the measurement of the response of government spending to the shock. Budget authority in particular lends itself to this approach because budget outlays can be used to set a lower bound on the bias due to timing in the response of government spending. Ultimately, this approach suggests that in the post-World War II United States, the "true" aggregate fiscal multiplier on defense spending is at least 1.3 to 1.4, and possibly 1.4 to 1.6.

Chapter 3

After the War: Effects of Wartime Spending on Household Consumption and Saving, 1946–1949

3.1 Introduction

World War II was the largest fiscal event of the twentieth century in the United States. It was also an extremely unusual event in an economic sense: the war effort required immense resources—labor, materials, and industrial capacity—with relatively little advance notice. It was impossible to increase productive capacity fast enough to meet the demands of the war effort without constraining the civilian economy. Instead, the government rationed consumer goods, some of which were not produced at all for the duration of the war, and strictly controlled the allocation of strategic materials.

After the war, the American economy boomed. The housing and automobile markets—two sectors which had been particularly affected by wartime constraints—were especially strong. Post-war American prosperity has long been attributed to the war, and particularly to the war effort on the home front for its role in stimulating the economy. But little direct evidence has linked the post-World War II boom to wartime spending. This paper provides that evidence.

Previous work on the U.S. macroeconomy during the war has shown that the stimulative effect of World War II spending was unusually small and argued that rationing played an important role in diminishing the size of the fiscal multiplier. This paper explores the macroeconomic effects of World War II spending on households *after* the war ended—and rationing with it. If war spending had positive effects on post-war consumption, that would imply that rationing *delayed* the consumption response to war spending, and in turn that the total stimulative effect of war spending was larger than previously believed.

While many people believe that World War II spending contributed to the post-war boom, it has been hard to find causal evidence, largely because timing cannot be used for identification. As is so often the case, geographic variation in the relevant historical data requires some effort to find—but it exists.

Total World War II spending is recorded by county in the *1947 County Data Book*. I link this information to households in the Surveys of Consumer Finances (SCFs, 1947–1950) according to the county in which each household lived at the time it was surveyed for the SCF.

I find that in locations that received more war spending, households surveyed over 1947 to 1950 were significantly more likely to have bought a car or a house in the previous year. Higher war spending was also predictive of larger holdings of liquid assets and higher levels of total household debt. With the notable exception of household debt, all of these effects were stronger for households headed by a person age 45–64. This is the age demographic most likely to have been employed in war production. In addition, households in this age group were also significantly more likely to purchase household durables when they lived in locations that had received more war spending.

The remainder of the paper is organized as follows. Section 3.2 discusses the motivation and relevant literature in greater detail. The empirical approach and identification are discussed in Section 3.3. Section 3.4 discusses the effects of war spending on consumption of cars and other durable goods. In Section 3.6 I turn to the effects of war spending on liquid saving and debt, and in Section 3.5 I examine the relationship between war spending and both home ownership and home purchases. Finally, Section 3.7 concludes.

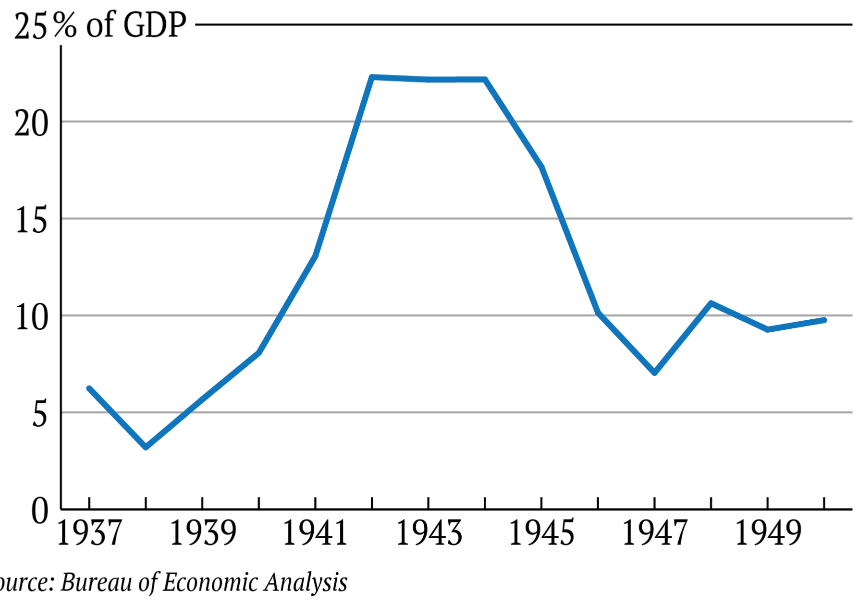
3.2 Motivation and Literature

The short-run stimulative effect of World War II spending appears to have been quite small. One reason for the small wartime multiplier was that the private savings rate was extremely high during the war (see Figure 3.1). Wartime rationing limited opportunities for wartime consumption. In this context, the wartime savings boom is not surprising. A natural question then is whether wartime saving produced delayed consumption responses after the war ended—and rationing with it.

Not only did aggregate saving increase dramatically during World War II, but previous work has shown that wartime saving increases were larger in locations that received more war spending. Fishback & Cullen (2013) show a county-level relationship between war spending and Series E war bond purchases, while chapter one shows relationships between war spending and both war bond purchases and bank deposit flows at the state level.

This suggests that households in areas receiving more war spending accumulated larger stocks of assets during World War II than households in areas that were less involved in war production. It is, of course, possible that after the war households in these areas simply continued to hold their wartime accumulation of assets as savings—given the accumulation of public debt during this period, continued saving would be consistent with Ricardian behavior. However, if rationing placed binding constraints on wartime consumption—especially consumption of durables—one would expect to see higher levels of consumption during the

Figure 3.1: Net Private Saving as a Share of GDP (United States)



immediate post-war years in areas that received higher levels of war spending: a delayed consumption response to wartime stimulus. This would suggest that while the immediate stimulative effect of war spending was quite small, the total stimulative effect of war spending was somewhat larger—and imply some degree of causation between government spending on the war and the post-war durables boom.

To establish the empirical relationship between war spending and post-war household behavior one needs geographically disaggregated information on both war spending and post-war household consumption and savings. Because it is so detailed and contains information not disaggregated elsewhere, household-level survey data can be particularly useful for building a complete picture of consumption and savings.

An extensive literature uses data from household-level surveys to measure consumption and savings responses to fiscal shocks. For example, Hsieh (2003) uses the Consumer Expenditure Survey to study the effects of changes to anticipated payments from the Alaska Permanent Fund on household consumption behavior and finds evidence for consumption smoothing when income shocks are large and transparent. Parker et al. (2013) also uses the Consumer Expenditure Survey to measure consumption responses to the 2008 economic stimulus payments, while Johnson et al. (2006) does the same for the 2001 federal income tax rebates. Shapiro & Slemrod (2009, 2003) explore the same questions (respectively) using data from the Michigan Survey of Consumers.

More recent papers in this literature sometimes use geographic variation in household survey data in their identification strategies. One example Hausman (2016), which uses the 1935–1936 Study of Consumer Purchases to measure consumption responses to the World

War I Veterans' Bonus of 1936.

3.3 Empirical Approach and Identification

Data

In this paper I use household-level survey data from the early Surveys of Consumer Finances (SCFs) to study the effects of World War II production spending on household consumption and saving in the immediate post-war years. While geographic information from the SCF is not commonly used, the early SCFs included codes for the location of each household in the sample, generally the county (or in a few instances, a small set of adjacent counties, the largest being New York City). While difficult to work with because the geographic coding was not consistent from year to year, each year's survey contains households located in 70–80 different counties (or county groups), chosen to create a nationally representative sample.

Most counties in the sample are repeated from year to year, increasing the average number of observations per included county. Over 1947 to 1950 the SCF included households from 92 geographic units covering 97 counties. While these counties represent only a small fraction of all U.S. counties, they include the most important U.S. metropolitan areas of the era and a broad mix of smaller cities and rural areas. It is worth recalling that the sample was designed to be nationally representative (at least for the purposes of the SCF).

The SCF is a relatively small survey, containing roughly 3,000 households per survey year. Because of the small sample size, I pool observations across years, for a total of 12,726 observations. Table 3.1 compares summary statistics for the pooled sample to summary statistics for households in the 1950 Census. Compared to the population of households in the Census, SCF households are much more educated and less likely to be headed by an elderly person—unsurprising for a survey of that era designed to measure consumption—but in other respects fairly similar to the U.S. population as a whole.

To measure the effects of war spending on household consumption and saving, I link county-level data—including total war spending— from the *1947 County Data Book* to SCF household observations based on the county in which the household is observed in the SCF. Ideally I would link households to their locations during the war, but that information is not available. That said, household locations during World War II should be highly correlated to household locations in the late 1940s. However, the identification problem posed by migration is one of several strong reasons for cutting off the sample after the 1950 SCF: the problems introduced by migration are only likely to grow more significant as the time horizon widens.

War spending varied dramatically among the counties included in the SCF.[1] Table 3.2 shows per capita war spending in the highest, median, and lowest-spending groups of five

[1]For geographic areas composed of multiple counties, I sum war spending across counties and then divide by the counties' combined population to generate average war spending for the geographic area defined in the SCF.

Table 3.1: SCF 1947-1950 Summary Statistics, Compared with 1950 Census

	(1) SCF 1947-1950, pooled	(2) 1950 Census (1% sample)
Black	0.06	0.08
WWII veteran	0.24	0.24
Age 18-64	0.89	0.85
Age 65 plus	0.11	0.15
Less than high school	0.42	0.65
High school	0.38	0.27
College	0.20	0.08
Observations	12,726	111,241

counties (calculated using county populations from the 1940 Census).

Baseline Specification

I estimate the following specification:

$$y_{ic} = \alpha + \beta w_c + \gamma' X_{ic} + \epsilon_{ic} \tag{3.1}$$

where y_{ic} is outcome y for household i in county c, w_c is total war production spending in county c, divided by the county's 1940 population, and X_{ic} is a vector of controls. X includes both household demographics—including the age, race, and education level of the head of household and total wage income earned by the household[2]—and county-level controls, such as the county's 1940 manufacturing employment rate. The baseline specification also includes controls for the size of city in which the household lives—in theory a household-level control but in practice often a county-level control for particularly urban or particularly rural counties. I also include dummies for the survey year.

It is impossible to include location fixed effects in this specification because they would be collinear with war spending. This makes the inclusion of controls such as city size all the more important.

Because many variables of interest measured in dollar values include observed zeros (especially for certain types of savings outcomes), I use the inverse hyperbolic sine transformation

[2] I use wage income as a control because total household income from all sources was not recorded in dollar amounts in the 1950 SCF. It is important to control for income both because of its direct relationship to household saving and consumption and because war production could in theory have had persistent effects on equilibrium wages in local labor markets. Since I am specifically trying to measure the effects of asset accumulation due to saving *during* the war, it is particularly important to control for post-war wages. I also include an indicator for whether the household has zero wage income, as households with no wage income are particularly likely to have income from other sources, such as self-employment or pensions.

Table 3.2: County per capita WWII spending, 1940-1945: Highest, Median, and Lowest Spending Counties in SCF Sample

County	Per Capita WWII Spending (nominal $)
Oakland, Michigan	10,004.78
Genesee, Michigan	8,096.83
Erie, New York	5,936.38
Wayne, Michigan	5,831.31
Portage, Ohio	5,377.93
Seneca, Ohio	892.12
New York City, New York	859.98
Lancaster, South Carolina	607.27
Wood, West Virginia	564.30
Gray's Harbor, Washington	511.00
Beadle, South Dakota	0
Geneva, Alabama	0
Haskell, Texas	0
Telfair, Georgia	0
Posey, Indiana	0

for all continuous variables. In this transofrmation, y_i is replaced with $log(y_i + (y_1^2 + 1)^{1/2})$. This transformation avoids the problems for values near zero introduced by more basic logarithmic transformations, but for larger values is approximately equal to $log(2y_i) = log(2) + log(y_i)$, so can be interpreted as if it were a standard logarithmic variable. Standard errors are clustered at the county level.

An alternate approach would be to take averages of all variables at the county level, and then run analogous regressions using county-level data. With this approach it is necessary to weight either by the number of observations per county or by the county's 1940 population, since some counties in the sample (especially in less populated areas) have particularly small samples, and within-county variation is significant enough to make county-level averages very noisy when the number of households is too small. Weighting either by the number of observations or by county populations (in either 1940 or 1950) produces results similar to those presented below.

Identification

Because there is substantial variation between households in income and wealth within any given area, this identification strategy relies on having enough observations *per location.*

While the sample size is small, the important thing is that it includes enough locations—92 of them—and enough household observations per location (an average of 138 per location).

Given this approach, the key identifying assumption is that the geographic distribution of war spending is not driven by unobserved location characteristics that also influence household consumption and saving behavior.

Crucially for the validity of this identifying assumption, all specifications include controls for the location's pre-war manufacturing employment (i.e. manufacturing employment as reported in the 1940 Census divided by 1940 Census population), the fraction of the population employed in agriculture in 1940, and the type of area (i.e. major metropolitan area, large city, small city, town, or rural area) in which the household resides. Together, these factors capture the key economic factors governing contract placement.

War contracts were placed by the military and the military was neither interested in nor concerned about the civilian economy. As a result, differences in local labor market conditions or differential recovery from the Great Depression were generally ignored when the military placed contracts, as discussed at length in chapter one. The military's main consideration was reliability: both the quality and timeliness of production were key considerations when the military placed contracts. This led to a higher dollar volume of contract placements in places with more initial manufacturing capacity—which might have received differential post-war shocks. The one function of the civilian economy which the military (and other war planners) intentionally protected from war production was food production, which was of course concentrated in rural areas. Controlling for both the share of the population employed in agriculture employed in 1940 and city size (which in some cases contains extra information about a household's location, since some counties include households in several city size designations) accounts for this concern in contract placement. Since these appear to have been the key economic factors influencing contract placement, including controls for these variables is essential for identification.

Another important consideration is that I want to measure the effects of military spending *during* World War II, not the effects of later military spending. This is another key reason for ending the sample with the 1950 SCF. As shown by Nakamura & Steinsson (2014), the geographic distribution of military spending is remarkably stable over time.

Military spending (and particularly the production of military goods) fell precipitously after World War II, but grew substantially (and unexpectedly) after the start of the Korean War in the summer of 1950. The 1950 SCF was collected from January through March 1950 and primarily questioned respondents about consumption and saving in 1949, so it is safe to include. However, including later SCFs in the sample would raise concerns that results stemmed from the immediate effects of Korean War spending rather than the longer-run effects of spending from World War II. Limiting the sample to the SCFs from 1947–1950 (covering 1946–1949) avoids the interpretation problem posed by the Korean War.

In the next sections I examine the effects of war spending on three sets of household-level outcomes: purchases of durable goods, the accumulation of liquid assets, and home ownership.

3.4 Results: Consumption of Durable Goods

Many of the materials used to produce durable goods were also critically important to producing war goods. This was particularly true of metals, including copper, steel, iron, and aluminum. These materials were themselves strictly rationed during the war, and in most cases only available to manufacturing firms for war production. Thus for much of World War II, consumers could not buy many types of durable goods because they were not produced.

The most extreme case of wartime rationing of consumer goods was in the automobile industry. Automobile production for civilian purchase was curtailed entirely in February 1942 and did not resume until the summer of 1945. Inventories of passenger cars totaled around 521,000 nationwide when the rationing order went into effect. These cars were slowly released for sale over the course of the war, but no new automobiles were produced for civilian purchase for over three years.

It is well known that purchases of cars and other consumer durables increased dramatically after World War II ended and rationing was relaxed. The important question for this paper is whether consumer purchases of cars and other durable goods were significantly higher in locations that received higher levels of war spending. Table 3.3 shows the estimated effects of World War II spending on household purchases of cars and other consumer durables.

The outcome in the first column is an indictor for whether a household bought a car in the past year. Thanks to the post-war automotive boom, nearly 21 percent of households surveyed in any given year had bought a car *in the previous year*. The coefficients shown in Table 3.3 are estimated using OLS. The linear probability model implies that if the per capita war spending received by the location where the household lived at the time of sampling were doubled, the probability that a household bought a car in the year prior to being surveyed would increase by 3.2 percentage points. For context, the average cost of a new car was around $1,500 in the immediate postwar years, while the median county in the sample received just over $600 per capita in war spending. An alternative approach to measuring the effect of war spending on the probability of buying a car would be to use a probit model. Probit estimation produces a positive and significant coefficient, and the marginal effect of war spending is very similar to the effect estimated by OLS—3.1 percentage points versus 3.2 percentage points.

The second column of Table 3.3 shows the effect of war spending on the cost of the automobile(s) (if any) purchased by the household in the previous year. The estimated effect is large and modestly significant, though it is driven by car purchases (the variable in the first column). The estimated effect is still positive when the sample is restricted only to those who purchased a car, but much smaller (a point estimate of 0.0478 and a p-value of 0.115).

The outcome in column three is an indicator for whether the household purchased one or more durable goods *other* than a car in the previous year. Specifically, this outcome measures whether the household bought a refrigerator, washing machine, stove, other electrical appliance (such as a vacuum cleaner or toaster), furniture, radio, or television. 36 percent

Table 3.3: Effects of County-Level War Spending on Post-War Household Durables Purchases

	Bought Car Last Year	Spending on Cars	Purchased Non-Car Durables Last Year	Borrowing for Durables Purchases
War Spending (per capita)	0.0322**	0.168*	0.00555	0.0507
	(0.0128)	(0.0945)	(0.0143)	(0.0855)
Black	-0.108***	-0.577***	0.0193	0.547***
	(0.0125)	(0.0828)	(0.0190)	(0.126)
Wage Income	0.0679***	0.497***	0.0734***	0.278***
	(0.00446)	(0.0373)	(0.00504)	(0.0337)
Zero Wage (Indicator)	0.612***	4.412***	0.570***	1.938***
	(0.0359)	(0.312)	(0.0462)	(0.273)
Pre-war Manufacturing Emp	-0.0812	-1.163	-0.351	-2.659*
	(0.182)	(1.513)	(0.222)	(1.511)
WWII Veteran in Household	0.0473***	0.354***	0.0534***	0.160***
	(0.0102)	(0.0662)	(0.0101)	(0.0605)
HH Age 25–34	0.0382**	0.0886	0.124***	0.277**
	(0.0158)	(0.112)	(0.0195)	(0.128)
HH Age 35–44	0.0180	-0.254**	0.0666***	-0.167
	(0.0164)	(0.114)	(0.0179)	(0.114)
HH Age 45–64	-0.00333	-0.537***	-0.0378**	-0.718***
	(0.0157)	(0.103)	(0.0184)	(0.118)
HH Age 65+	-0.111***	-1.300***	-0.151***	-1.018***
	(0.0199)	(0.124)	(0.0238)	(0.133)
Observations	11,444	12,587	12,594	9,202
R-squared	0.101	0.079	0.069	0.084

All variables in dollar amounts (war spending, income, etc.) use inverse hyperbolic sine transformation. In addition to controls shown, regression also includes education bins, bins for city size, and fixed effects for survey years. Standard errors clustered by county.
*** p<0.01, ** p<0.05, * p<0.1

of households in the sample bought one or more of these items in the year prior to being sampled. Once again, Table 3.3 estimates a linear probability model. The estimated effect of war spending on the probability of purchasing household durables is near zero and not statistically significant. Estimating a probit model instead of using OLS produces similar results.

Finally, the fourth column of Table 3.3 shows the effects of war spending on borrowing for durables purchases. There are significantly fewer observations for this outcome because this variable was not recorded in the 1950 SCF.[3] War spending does not appear to have had

[3]The 1950 SCF recorded installment debt for 1949 durables purchases, but other types of debt incurred for that purpose were reported jointly with debts incurred for other purposes.

Table 3.4: Effects of County-Level War Spending on Post-War Household Durables Purchases, Head of Household Aged 45–64

	Bought Car Last Year	Spending on Cars	Purchased Non-Car Durables Last Year	Borrowing for Durables Purchases
War Spending (per capita)	0.0418***	0.175	0.0286**	0.0136
	(0.0152)	(0.113)	(0.0144)	(0.0885)
Black	-0.0963***	-0.392***	0.0448	0.775***
	(0.0211)	(0.120)	(0.0335)	(0.186)
Wage Income	0.0674***	0.472***	0.0579***	0.0938*
	(0.00700)	(0.0535)	(0.00967)	(0.0530)
Zero Wage (Indicator)	0.628***	4.414***	0.439***	0.401
	(0.0592)	(0.452)	(0.0797)	(0.437)
Pre-war Manufacturing Emp	-0.391*	-2.978*	-0.386*	-2.595*
	(0.228)	(1.695)	(0.209)	(1.443)
WWII Veteran in Household	-0.0180	0.163	-0.0205	-0.0530
	(0.0252)	(0.158)	(0.0214)	(0.0907)
Observations	3,964	4,332	4,334	3,204
R-squared	0.159	0.068	0.025	0.038

All variables in dollar amounts (war spending, income, etc.) use inverse hyperbolic sine transformation. In addition to controls shown, regression also includes education bins, bins for city size, and fixed effects for survey years. Standard errors clustered by county.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

a significant effect on consumer borrowing for the purpose of buying durable goods.

Table 3.4 shows the same outcomes, but this time with the sample restricted to households headed by a person of age 45–64. This age group is the cohort most likely to have accumulated assets through war production: young enough to be working, but old enough that the men were much more likely to have worked in war production than to have served overseas in the military. Compared to households headed by members of younger age cohorts, these households were less likely to purchase a car or household durables (while households headed by a person aged 25–34 were most likely to buy both).

Interestingly, war spending has a stronger effect on purchases of both cars and household durables for this subsample of households which were more likely to have receive income from war production. As estimated with a linear probability model, doubling war spending in a location made households in this age bin 4.2 percentage points more likely to buy a car and 2.9 percentage points more likely to buy household durables. Probit estimation produces positive coefficients with nearly identical marginal effects.

Meanwhile, the estimated effect of war spending on the cost of the car(s) (if any) purchased by the household in the previous year is essentially unchanged, though no longer statistically significant. Similarly, war spending still does not appear to have a significant effect on household borrowing for the purpose of buying durable goods.

3.5 Home Ownership

Along with car purchases, there was a particularly large increase in home ownership directly after the war. Nationally, home construction boomed after World War II (after being strictly limited during the war years). Construction on slightly more than 1 million new housing units began in 1946, compared to just 325,000 housing starts in 1945 and a pre-war high of 620,000 in 1942. New housing starts rose in every year from 1946 through 1950, hitting 1,900,000 new housing units begun in 1950. In this context, it is natural to consider the relationship between war spending and the post-war housing boom.

Home ownership is an unusual good in that it spans both consumption and investment. But a house is considered a form of investment, and indeed often proves a valuable asset, but the household also consumes the (imputed) rental value of the house if it resides in it. So when a household buys a house that is of higher quality than its previous dwelling—larger, in a better neighborhood, etc.—it is increasing its consumption while also making an investment.

Table 3.5 shows the effect of war spending in a location on the probability that a household in that location owns its home, and also on the probability that a household bought a house in the past year. Columns 1 and 3 use OLS while columns 2 and 4 use probit estimation.

War spending appears to have had economically large effects on home ownership. The linear probability model suggests that doubling war spending in a location was associated with a 5.6 percentage point increase in the location's homeownership rate and a 2 percentage point increase in the fraction of households who bought a house in the previous year. The marginal effects implied by the probit approach are nearly identical: a 5.6 percentage point increase in the homeownership rate and a 1.9 percentage point increase in home purchases. (Note that some households who purchased homes were already home owners, so one would expect the increase in the homeownership rate to be smaller than the cumulative mass of households who bought over 1946–1949.)

Table 3.6 shows the same outcomes, but restricted to households headed by an individual age 45–64. For all housing-related outcomes, the results are stronger for this subset of households. Note that while these households were as a group less likely to buy cars and household durable goods than other age groups, they were more likely to own houses—and yet the estimated effects of war spending on car purchases, household durables purchases, and homeownership are larger for these households than for households in other age categories.

Tables 3.5 and 3.6 suggest that the post-war housing boom was likely fueled at least in part by war spending.

Table 3.5: Effects of County-Level War Spending on Post-War Homeownership

	Household Owns Home		Household Purchased Home Last Year	
	OLS	Probit	OLS	Probit
War Spending (per capita)	0.0561**	0.163**	0.0198***	0.182***
	(0.0255)	(0.0741)	(0.00520)	(0.0550)
Black	-0.117***	-0.355***	-0.0204***	-0.218**
	(0.0253)	(0.0769)	(0.00724)	(0.0933)
Wage Income	0.0652***	0.200***	0.0142***	0.158***
	(0.00818)	(0.0233)	(0.00370)	(0.0420)
Zero Wage (Indicator)	0.604***	1.837***	0.137***	1.516***
	(0.0723)	(0.205)	(0.0296)	(0.349)
Pre-war Manufacturing Emp	-0.175	-0.530	-0.152*	-1.491*
	(0.304)	(0.870)	(0.0811)	(0.769)
WWII Veteran in Household	-0.00663	-0.00461	0.0255***	0.221***
	(0.0116)	(0.0339)	(0.00636)	(0.0440)
HH Age 25–34	0.181***	0.769***	0.0224***	0.198**
	(0.0196)	(0.0685)	(0.00822)	(0.0831)
HH Age 35–44	0.352***	1.235***	0.0207**	0.189**
	(0.0233)	(0.0745)	(0.00865)	(0.0878)
HH Age 45–64	0.452***	1.506***	-0.00418	-0.0414
	(0.0267)	(0.0778)	(0.00854)	(0.0936)
HH Age 65+	0.461***	1.528***	-0.0353***	-0.519***
	(0.0279)	(0.0872)	(0.00982)	(0.138)
Observations	11,668	11,668	11,479	11,479
R-squared	0.154		0.020	

In addition to controls shown, regression also includes education bins, bins for city size, and fixed effects for survey years. Variable for whether household purchased a home in the previous year taken from Fetter data set (published by the *American Economic Journal: Economic Policy*). Standard errors clustered by county. *** p<0.01, ** p<0.05, * p<0.1

3.6 Results: Liquid Savings

The preceeding sections show that households in locations which received more war spending were more likely to buy cars and houses than households in places which received less war spending.

Previous studies of the effects of World War II military spending show a positive relationship between war spending and liquid forms of wartime saving, in particular war bond purchases and bank deposit flows. Buying a house usually requires substantial liquid assets (for the down payment), so would be easier for households with more liquid assets to attain—explaining how war spending may have driven increases in homeownership.

Table 3.6: Effects of County-Level War Spending on Post-War Homeownership, Head of Household Aged 45–64

	Household Owns Home		Household Purchased Home Last Year	
	OLS	Probit	OLS	Probit
War Spending (per capita)	0.0881***	0.247***	0.0245***	0.259***
	(0.0322)	(0.0879)	(0.00588)	(0.0695)
Black	-0.145***	-0.384***	0.0168	0.169
	(0.0342)	(0.0935)	(0.0151)	(0.146)
Wage Income	0.0638***	0.176***	0.00715	0.0748
	(0.0115)	(0.0324)	(0.00538)	(0.0635)
Zero Wage (Indicator)	0.589***	1.634***	0.0718	0.758
	(0.103)	(0.290)	(0.0459)	(0.548)
Pre-war Manufacturing Emp	-0.573	-1.664	-0.146	-1.596
	(0.375)	(1.020)	(0.102)	(1.162)
WWII Veteran in Household	0.0507**	0.144**	0.0139	0.130
	(0.0240)	(0.0683)	(0.0138)	(0.126)
Observations	4,092	4,092	3,885	3,885
R-squared	0.081		0.012	

In addition to controls shown, regression also includes education bins, bins for city size, and fixed effects for survey years. Variable for whether household purchased a home in the previous year taken from Fetter data set (published by the *American Economic Journal: Economic Policy*). Standard errors clustered by county. *** p<0.01, ** p<0.05, * p<0.1

At the same time, households usually take on significant mortgage debt when they buy a house, increasing their borrowing. This was already the case in the 1940s. Given household purchases of houses and other assets, it is useful to understand the relationship between wartime saving and postwar accumulations of assets and debt. Table 3.7 shows the estimated effects of World War II spending on household holdings of a range of liquid assets.

The outcome in the first column is series A-F bond holdings, which includes most war bonds held by individuals (series E, F, and G bonds were originally referred to as war bonds and then as victory bonds, but series G was generally purchased by institutional investors). The point estimate for the coefficient on war spending is small and positive, implying that post-war A-F bond holdings would increase by 5 percentage points if war spending in a location were doubled, but it is not statistically significant. This is consistent with Fishback & Cullen, who find a positive and significant effect of war spending on growth of series E war bond holdings (the type of war bond which could *only* be purchased by individuals) over 1940–1944 but essentially no effect on E-bond holdings over 1940–1949. E-bonds were not introduced until 1941, so all holdings were zero in 1940. Thus the change in E-bond holdings over 1940–1949 is equivalent to E-bond holdings in 1949, and so directly comparable to the

Table 3.7: Effects of County-Level War Spending on Post-War Household Savings Outcomes

	Series A-F Bond Holdings	Savings Account Balance	Total Liquid Saving	Total Debt
War Spending (per capita)	0.0493	0.175	0.169*	0.501***
	(0.120)	(0.176)	(0.0967)	(0.146)
Black	-1.799***	-1.179***	-2.265***	0.272
	(0.138)	(0.122)	(0.148)	(0.188)
Wage Income	0.635***	0.469***	0.641***	0.436***
	(0.0605)	(0.0479)	(0.0565)	(0.0746)
Zero Wage (Indicator)	5.246***	3.855***	5.834***	3.363***
	(0.534)	(0.425)	(0.474)	(0.599)
Pre-war Manufacturing Emp	1.936	5.797**	1.194	-6.315**
	(1.793)	(2.271)	(1.473)	(2.673)
WWII Veteran in Household	-0.00916	0.211***	0.105	0.406***
	(0.0686)	(0.0679)	(0.0645)	(0.138)
HH Age 25–34	-0.330*	-0.625**	-0.484***	-0.0751
	(0.188)	(0.251)	(0.145)	(0.257)
HH Age 35–44	-0.458**	-0.767**	-0.290	0.580**
	(0.225)	(0.360)	(0.189)	(0.264)
HH Age 45–64	2.053***	0.992***	1.979***	0.434*
	(0.102)	(0.128)	(0.113)	(0.228)
HH Age 65+	1.622***	0.836***	1.443***	-1.130***
	(0.152)	(0.153)	(0.165)	(0.214)
Observations	12,593	12,590	12,582	5,826
R-squared	0.178	0.244	0.223	0.100

All variables in dollar amounts (war spending, income, etc.) use inverse hyperbolic sine transformation. In addition to controls shown, regression also includes education bins, bins for city size, and fixed effects for survey years. Standard errors clustered by county.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

results of the SCF shown here.

The second column shows the effects on savings account balances in the late 1940s. Once again the coefficient on war spending is positive and statistically insignificant. It suggests that doubling war spending in a location is associated with a 17.5 percentage point increase in post-war saving account balances, but the standard error is very large. Interestingly, the coefficient on pre-war manufacturing employment is both large and statistically significant, implying that households in locations with proportionally more manufacturing employment before the war held more wealth in their bank accounts during the immediate post-war years. (Households in these areas likely held more bank savings to begin with; bank deposit holdings and manufacturing employment rates in 1939 were highly correlated at the state level.)

The outcome in the third column is total liquid savings.[4] This variable includes both Series A-F bond holdings and assets in saving accounts, but also accounts for several other types of liquid savings commonly held by households in the 1940s: other U.S. government bonds, postal savings, deposits held in building and loan associations, and deposits in checking accounts. The coefficient on war spending implies that if war spending in a county had been doubled, households in that county would have held approximately 17 percent more total liquid savings in the late 1940s. This coefficient is significant at the 90 percent confidence level.

It is not surprising to find a (moderately) significant effect of war spending on total liquid asset holdings without finding significant effects on discrete types of liquid assets: there is no particular reason to think that war spending in an area should affect post-war asset composition, and there is significant variation in asset composition between households. Therefore total liquid assets are likely to be substantially less noisy than their individual components, as suggested by the standard errors.

Taken with the more precise effects of war spending on contemporary liquid savings found in other papers, these findings suggest that at least some households which benefited from war spending spent down a significant portion of their wartime accumulation of liquid assets immediately after the war, likely purchasing houses and cars.

The final column of Table 3.7 shows the relationship between war spending and total household debt. There are significantly fewer observations for this variable because it was not recorded in either the 1947 or 1948 SCFs. Total household debt was reported directly in the 1949 SCF, and the 1950 SCF included all of the components used to calculate total debt in the 1949 SCF (following the worksheet in the survey questionnaire), making it possible to compute total debt indirectly for households in the 1950 SCF.

Interestingly, the coefficient on war spending is large, significant, and positive. It implies that a 10 percent increase in war spending in a location is associated with a 5 percent increase in debt for households in that location. The correlation between total household debt and total liquid asset holdings is -0.01, or virtually zero. One possible explanation for these findings would be that the households which benefitted directly from war spending both held more assets and increased their consumption—or, very probably, bought houses—while the households which did not directly benefit from war spending may have borrowed in order to buy houses and cars like their friends and neighbors whose incomes had benefitted from war production.

Table 3.8 shows the same outcomes, but this time with the sample restricted to households headed by a person of age 45–64. As noted previously, this age group is the cohort most likely to have accumulated assets through war production: young enough to be working, but old enough that the men were much more likely to have worked in war production than to have served overseas in the military. The coefficient on war spending for total liquid saving

[4]In the 1947–1949 SCFs, this variable was called *total liquid asset holdings*. In the 1950 SCR it was called *total savings*. Survey worksheets included in the codebooks make it clear that the variables are equivalent.

Table 3.8: Effects of County-Level War Spending on Post-War Household Savings Outcomes, Head of Household Aged 45–64

	Series A-F Bond Holdings	Savings Account Balance	Total Liquid Saving	Total Debt
War Spending (per capita)	0.255	0.354	0.364***	0.396**
	(0.157)	(0.213)	(0.129)	(0.174)
Black	-2.481***	-1.356***	-2.738***	0.977***
	(0.192)	(0.190)	(0.259)	(0.345)
Wage Income	0.831***	0.433***	0.692***	0.436***
	(0.0805)	(0.0648)	(0.0648)	(0.147)
Zero Wage (Indicator)	7.044***	3.740***	6.393***	3.296***
	(0.714)	(0.587)	(0.561)	(1.155)
Pre-war Manufacturing Emp	-0.227	2.463	-0.926	-6.349*
	(2.119)	(3.048)	(2.019)	(3.618)
WWII Veteran in Household	-0.00973	0.0303	-0.151	-0.136
	(0.185)	(0.204)	(0.176)	(0.314)
Observations	4,334	4,333	4,328	2,098
R-squared	0.169	0.221	0.220	0.031

All variables in dollar amounts (war spending, income, etc.) use inverse hyperbolic sine transformation. In addition to controls shown, regression also includes education bins, bins for city size, and fixed effects for survey years. Standard errors clustered by county.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

is roughly double the coefficient for the full sample, while the standard error is only very slightly larger. This suggests that the effect of war spending on liquid saving is driven by households headed by a person age 45–64.

In contrast, the coefficient on total debt is 0.396 for the subsample, somewhat smaller than the coefficient for the full sample (0.501). This suggests that the effect of war spending on household debt is *not* driven by households in the age group most likely to have been employed in war production (though the effect is still large and significant for this age cohort). More generally, the effects of war spending on household debt appear to be larger for younger age cohorts and smaller for older age cohorts.

Taken together, these results suggest that war spending had a positive effect on households' liquid asset holdings in the late 1940s (1946–1949), but was also associated with higher levels of household debt.

One possible explanation is that because older households had more total assets, they were able to buy houses while taking on lower levels of debt. Meanwhile, war spending enabled some younger households to buy houses, but because those households had fewer total assets, war spending prompted them to take on more debt in order to buy those houses.

3.7 Conclusion

The data strongly suggest that households in locations which received more war spending bought cars and houses at higher rates than similar households in areas which received less World War II spending. Households in these locations also held moderately more liquid assets and, conversely, significantly higher levels of debt.

This evidence supports the popular narrative that World War II spending contributed to post-World War II economic prosperity. Moreover, it suggests that the total stimulative effect of World War II spending was larger than estimates of the short-run fiscal multiplier would suggest.

The nature of the data make it difficult to quantify (in dollars) the total effect of war spending on post-war consumption. However, it is clear that the effects on both home and car purchases are economically significant.

Bibliography

Auerbach, Alan and Yuriy Gorodnichenko (2012) “Measuring the Output Responses to Fiscal Policy,” *American Economic Journal: Economic Policy*, Vol. 4, pp. 1–27.

——— (2013) “Output Spillovers from Fiscal Policy,” *American Economic Review*, Vol. 103, pp. 141–146.

Automobile Manufacturers Association (1947) *Automobile Facts and Figures*, 27th edition.

Barro, Robert J. and Charles J. Redlick (2011) “Macroeconomic Effects from Government Purchases and Taxes,” *Quarterly Journal of Economics*, Vol. 126, pp. 51–102.

Bureau of the Budget (1946) *The United States at War: Development and Administration of the War Program by the Federal Government*.

Bureau of Labor Statistics, United States Department of Labor (1947) *Activities of the Bureau of Labor Statistics in World War II: Development and Administration of the War Program by the Federal Government*.

Bureau of Economic Analysis (2005) *Government Transactions. Methodology Papers: U.S. National Income and Product Accounts*.

Bureau of Economic Analysis (accessed April 2017) *National Income Product Accounts (Various Tables)*.

The Chilton Co. (1941) *Automotive Industries*.

Chodorow-Reich, Gabriel, Laura Feiveson, Zachary Liscow, and William Gui Woolston (2012) “Does State Fiscal Relief During Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act,” *American Economic Journal: Economic Policy*, Vol. 4, pp. 118–145.

Civilian Production Administration (1946) *Alphabetic listing of major war supply contracts: cumulative June 1940 through September 1945*.

Collins, William J. (2001) “Race, Roosevelt, and Wartime Production: Fair Employment in World War II Labor Markets,” *American Economic Review*, Vol. 91, pp. 272–286.

Committee on Appropriations, U.S. House of Representatives (2010) *A Concise History of the House of Representatives Committee on Appropriations*.

Congressional Budget Office (2016) *CBO's Projections of Federal Receipts and Expenditures in the National Income and Product Accounts*.

Crain Automotive Group (1952) *Automotive News (Almanac Edition)*.

Duflo, Esther, William Gale, Jeffrey Liebman, Peter Orszag, and Emmanuel Saez (2006) "Saving Incentives for Low- and Middle-Income Families: Evidence from a Field Experiment with H&R Block," *Quarterly Journal of Economics*, Vol. 121, pp. 1311–46.

Economic Behavior Program, Survey Research Center, University of Michigan (1947–1950) *Surveys of Consumer Finances*, ICPSR version.

Fetter, Daniel K. (2013) "How Do Mortgage Subsidies Affect Home Ownership? Evidence from the Mid-Century GI Bills," *American Economic Journal: Economic Policy*, Vol. 5, pp. 111–147.

——— (2016) "The Home Front: Rent Control and the Rapid Wartime Increase in Home Ownership," *Journal of Economic History*, Vol. 76, pp. 1001–1043.

Feyrer, James and Bruce Sacerdote (2011) "Did the Stimulus Stimulate? Real Time Estimates of the Effects of the American Recovery and Reinvestment Act," *NBER Working Paper 16759*.

Fishback, Price and Joseph A. Cullen (2013) "Second World War spending and local economic activity in US counties, 1939–58," *Economic History Review*, Vol. 66, pp. 975–992.

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

Goldin, Claudia (1991) "The Role of World War II in the Rise of Women's Employment," *American Economic Review*, Vol. 81, pp. 741–756.

Goldin, Claudia and Claudia Olivetti (2013) "Shocking Labor Supply: A Reassessment of the Role of World War II on Women's Labor Supply," *American Economic Review*, Vol. 103, pp. 257–262.

Goldsmith, Raymond William (1955) *A Study of Saving in the United States*, Vol. I: Princeton University Press.

Hall, Robert E. (2009) "By How Much Does GDP Rise If the Government Buys More Output?," *Brookings Papers on Economic Activity*, Vol. fall, pp. 183–231.

Hausman, Joshua K. (2016) "Fiscal Policy and Economic Recovery: The Case of the 1936 Veterans' Bonus," *American Economic Review*, Vol. 106, pp. 1100–1143.

Higgs, Robert (1992) "Wartime Prosperity? A Reassessment of the U.S. Economy in the 1940s," *Journal of Economic History*, Vol. 52, pp. 41–60.

Hsieh, Chang-Tai (2003) "Do Consumers React to Anticipated Income Changes? Evidence from the Alaska Permanent Fund," *American Economic Review*, Vol. 93, pp. 397–405.

Industrial College of the Armed Forces (1947) *Applied Military Economics: A Case Study of the American War Economy (Preliminary Draft).*

Internal Revenue Service (1939–1946) *Statistics of Income Report, Part 1*.

Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles (2006) "Household Expenditure and the Income Tax Rebates of 2001," *A*merican Economic Review, Vol. 96, pp. 1589–1610.

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

Lacey, James G. (2011) *Keep From All Thoughtful Men: How U.S. Economists Won World War II*: Naval Institute Press.

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

Mertens, Karel and Morten O. Ravn (2012) "Empirical Evidence on the Aggregate Effects of Anticipated and Unanticipated US Tax Policy Shocks," *American Economic Review*, Vol. 4, pp. 145–181.

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

Nelson, Donald M. (1946) *Arsenal of Democracy: the Story of American War Production*: Harcourt, Brace and Co.

Office of Management and Budget (1923–2016) *Budgets of the U.S. Government*, digitized by the Federal Reserve Bank of St. Louis.

Office of the Comptroller of the Currency (1939–1948) *Annual Reports of the Comptroller of the Currency*.

Parker, Jonathan A., Nicholas S. Souleles, David S. Johnson, and Robert McClelland (2013) "Consumer Spending and the Economic Stimulus Payments of 2008," *American Economic Review*, Vol. 103, pp. 2530–2553.

Ramey, Valerie A. (2011) "Identifying Government Spending Shocks: It's all in the Timing," *Quarterly Journal of Economics*, Vol. 126, pp. 1–50.

Ramey, Valerie A. and Matthew D. Shapiro (1998) "Costly capital reallocation and the effects of government spending," *Carnegie-Rochester Conference Series on Public Policy*, Vol. 48, pp. 145–194.

Ramey, Valerie A. and Sarah Zubairy (forthcoming) "Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data," *Journal of Political Economy*.

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

Schick, Allen (1990) *The Capacity to Budget*: Urban Institute Press.

Shapiro, Matthew D. and Joel Slemrod (2003) "Consumer Response to Tax Rebates," *A*merican Economic Review, Vol. 93, pp. 381–396.

——— (2009) "Did the 2008 Tax Rebates Stimulate Spending?" *A*merican Economic Review, Vol. 99, pp. 374–79.

Shoag, Daniel (2013) "Using State Pension Shocks to Estimate Fiscal Multipliers since the Great Recession," *American Economic Review*, Vol. 103, pp. 121–124.

Smith, R. Elberton (1959) *United States Army in World War II: the Army and Economic Mobilization*: Center of Military History, United States Army.

Thompson, Peter (2001) "How Much Did the Liberty Shipbuilders Learn? New Evidence for an Old Case Study," *Journal of Political Economy*, Vol. 109, pp. 103–137.

United States Treasury Department, Office of the Secretary (1942–1947) *Treasury Bulletin*.

United States Department of Commerce. Bureau of the Census (1947–1977) *County and City Data Book [United States] Consolidated File: County Data*, Provided by ICPSR.

United States Treasury Department (1991) *A History of the United States Savings Bonds Program* (50^{th} Anniversary Ed.).

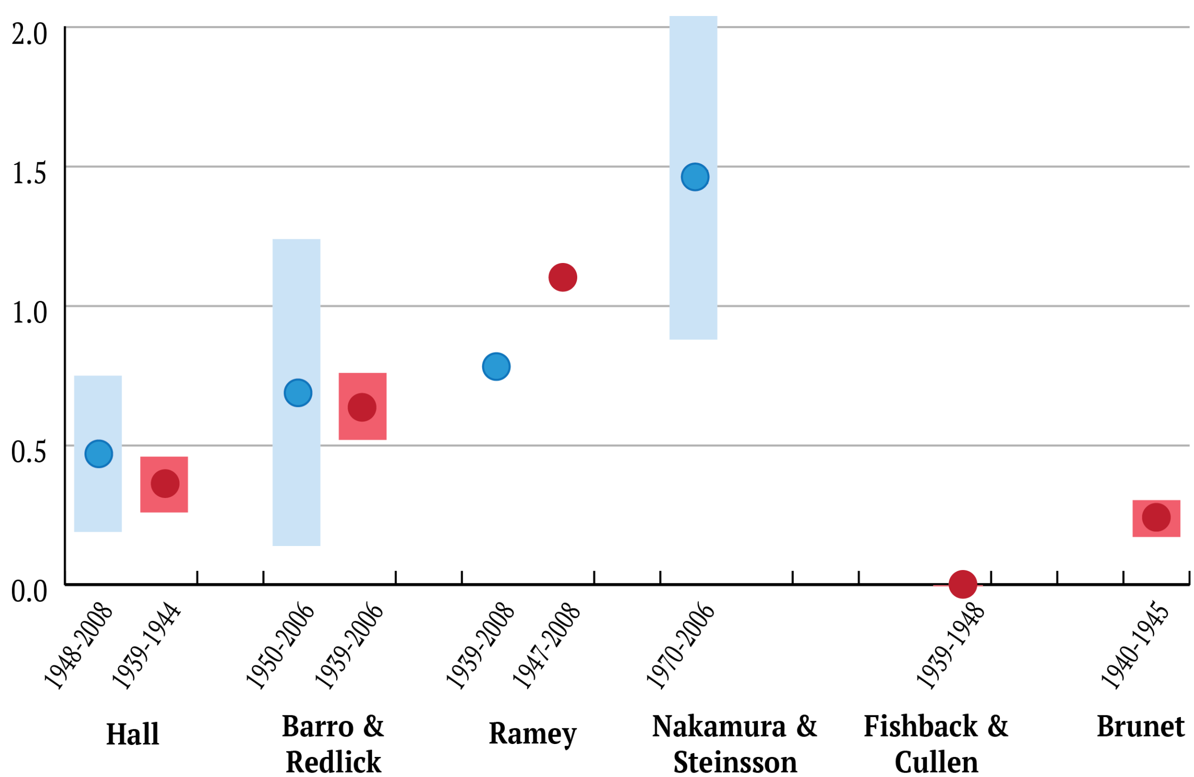
Wilson, Daniel J. (2012) "Fiscal Spending Jobs Multipliers: Evidence from the 2009 American Recovery and Reinvestment Act?," *American Economic Journal: Economic Policy*, Vol. 4, pp. 251–282.

Appendix A

Appendix to Chapter One

A.1 World War II in the Fiscal Multiplier Literature

Figure A.1: Fiscal Multiplier Estimates Including and Excluding World War II



How do estimates of the fiscal multiplier vary when World War II is included in or excluded from the sample? While Figure A.1 gives a useful summary of the literature, it is also useful to understand how differences in methodology and approach

Hall (2009) uses annual aggregate time series on military purchases and GDP to measure the fiscal multiplier. His specification uses one year differences and examine effects over a one-year time horizon. Hall finds a multiplier of 0.47 for 1948–2008 and a multiplier of 0.55 for 1930–2008, but a multiplier of 0.36 for 1939–1944, suggesting that the aggregate multiplier was smaller during World War II than during other periods.

Ramey (2011) constructs a narrative measure of anticipated defense spending (again at the national level), which she uses to construct a time series of shocks to the present discounted value of anticipated defense spending. This is the only paper to find substantially larger multipliers when World War II is included in the sample. There appear to be two factors driving this finding. First, Ramey's narrative measure shifts the balance of spending shocks significantly earlier than in either NIPA or the War Production Board contract data, with the first significant shock coming in the spring of 1940, before U.S. entry into the war (the defense effort was significantly picking up at this time, and Ramey measures changes to *expected future defense spending*, which certainly did increase in early 1940). Conversion was not yet significant during this period and unemployment was dropping quickly, so it is not surprising that this shift in timing would lead to larger estimates of the aggregate multiplier.

Second, Ramey's measure uses a present value approach and assumes that shocks to spending will last for varying periods of time, usually 3–5 years. However, for a few events for which shocks are long-lasting, Ramey uses longer time horizons. One such event is the start of the Korean War, signaling the start of the Cold War and a seemingly permanent increase in U.S. defense spending. Already a large shock even in immediate terms, the Korean War becomes very large because the shock is computed over a 10-year time horizon. This reflects a different conception of what about government spending triggers economic responses, but as a consequence the start of the Korean War is much larger relative to World War II in Ramey's measure than it is in any other measure, likely leading to relatively smaller estimates when World War II is excluded from the sample (and important because both wars are highly leveraged due to the scale of spending shocks).

Barro & Redlick (2011) similarly use annual data on military purchases to estimate the fiscal multiplier, but with several additional controls, including (lagged) marginal tax rates, unemployment rates, and Ramey's defense news measure. Their specifications use annual changes to both output and government spending, though they include the lagged change in government spending in addition to the contemporaneous change. Instead of interpreting the lagged change in government spending as a control (better isolating the shock to spending), they argue that the coefficients on the changes to contemporaneous and lagged government spending should be added. With this interpretation (as shown in Figure A.1), the fiscal multiplier for periods including World War II is marginally smaller than the fiscal multiplier for periods excluding World War II. However, when the data include serially correlated positive shocks, as is the case for defense spending in World War II, summing the coefficients on the current and lagged changes in government spending (rather than computing changes to both government spending and output over a longer window) may lead to an upward bias. When comparing the coefficients only on the contemporaneous change in government spending, the estimated multiplier for 1939–2006 is considerably lower

(.44) than the estimated multiplier for 1950–2006 (.68). Given the possibility of bias when summing coefficients in these circumstances, the best point estimate for the fiscal multiplier including World War II is likely somewhere between .44 and .64.

A.2 War Supply Contract Data

Digitization and Cleaning of Raw Contract Data

The war production contract data comes from the "Alphabetic listing of major war supply contracts: cumulative June 1940 through September 1945," published by the Civilian Production Administration, the successor agency to the War Production Board, in 1946. Digitizing and cleaning the raw contract data was a joint project with Dmitri Koustas, who is using the data for a separate project.

The original data consisted of four bound books, all of them essentially ledgers containing summary information for every war production contract between June 1940 and September 1945 with a value of at least $50,000. [1] Contracts are listed alphabetically by firm name, and within firms by establishment of main production (and then by award date). The complete data set includes 191,709 unique contracts.

The four volumes of contracts were digitized and put into CSV format by E-Records USA, a digitization firm in Richmond, CA which guarantees 99% accuracy in data extraction. We were impressed with their work; every typo and clear error we have found in the data set (e.g. contracts for scissors from "ACME Shear Co." with location listed as "BRIDGFPORT CONN" instead of "BRIDGEPORT CONN") were reproduced from the original.

Each contract listing includes *both* contract award date (month & year) and contract completion date (month & year). The complete award date (month & year) is not legible for roughly 50 contracts in the data, but for many of these we can use dates on similar contracts within the same firm, and other available date information (completion date, partial award date) to make an educated guess, which we do on a case by case basis (checking against the physical books, which generally did not help due to wear on the books). A similar number of contract listings lack full completion dates; we take the same approach with those. Another 215 contracts in the data set list completion dates that come *before* award dates. These observations are dropped.

Contracts are listed within firm-location pairings, with location reported as the primary location of manufacture for each contract. The location field is listed as "CITY STATE," with no punctuation other than spaces and using state abbreviations. Because the data was published before the introduction of standardized two letter state postal codes, state

[1] Some 241 contracts with values below $50,000 do appear in the data set, many of them for hardware, automobile parts, raw materials, or clothing items. These contracts appear to be for companies that had multiple contracts, many of them large, well-known firms. More than 25 of the 241 small contracts were with General Motors. Other firms with contracts under $50,000 include "U.S. Steel Export Co.," "Carnegie Ill. Steel Corp.," "Bethlehem Steel Co.," "General Electric Co.," and "Du Pont E. I. Co. Inc."

abbreviations vary in length, and some state abbreviations contain spaces, such as "S DAK" for South Dakota and "N MEX" for New Mexico.

In order to extract states from the data, we append " 123" to the end of each location field in the data. Then we search the location field for a string composed of "[state abbreviation] 123" for each state, and assign states accordingly. Because there are no numeric characters in city names, this ensures that we pull the state abbreviation from the end of the string. This prevents improper assignments of states, so avoids assigning contracts produced in New Orleans to Oregon rather than Louisiana, and other potential errors for cities whose names contain abbreviations used for other states.

Forty firm-location pairs lack clear information on state of production. For half of these pairs, the state can be determined with near-certainty from the city name: "BROOKLYN," "LOS ANGELES," "PHILADELPHIA," "POUGHKEEPSIE," and "NEW ORLEANS" fall into this category (assigned to NY, CA, PA, NY, and LA, respectively), as do "DENVER COIO," "PORTLAND ORFG," "ELKHART ING," and "PROVIDENCE R" (assigned to CO, OR, IN, and RI). For most other firm-location pairs, we used internet searches with the city and name of manufacturer to determine state. For example, "POTTS MFG CO" of Mechanicsburg can be located in Pennsylvania, and "ELMVALE DYE WORKS INC" of Pittsfield can be located in Massachusetts. Similarly, it is possible to verify that the "ACME CHAIR CO," whose location is listed as "READING HIGH," is in fact in Michigan. "HICKS BODY CO INC," whose location is listed as "LEBANON INS," can be traced to Indiana. There are also two firms whose locations are completely missing. One of those firms, "AUBURN BUTTON WORKS TNC," can be located in New York state. The final firm, "K4TECHNICAL SERVICES," we could not locate. There is at least one modern firm of that name, but it was founded in 2009, so cannot be a match. We drop the two contracts with "K4TECHNICAL SERVICES," which have a combined value of $142,000. Another 925 contracts with a total value of $1.8 billion are dropped from the sample because their location of manufacture is outside of the United States.

The resulting data set contains 190,599 contracts with a total value of $180.9 billion (in nominal dollars).

Constructing a State-Level Panel on Spending from Contract Data: Three Approaches

Because the contract data includes both award and completion dates for each contract, constructing a panel dataset on war production spending requires assumptions about the timing of spending.

For my baseline specification, I assume that spending is uniformly distributed over the length of each contract. This is the most neutral assumption about when contracts spur economic activity. To construct spending data under this assumption, I first calculate the length of each contract in months (inclusive of award and completion months). I then divide the total value of the contract by the contract length to get a value per month for the contract. This monthly value is then assigned to every month between the contract's award and completion dates (inclusive). Because the outcomes of interest are not available with high enough frequency below the state level, spending is then aggregated to states and the appropriate time period (quarters for employment effects, years for income effects).

I also construct spending data following two other assumptions about the correct timing of spending: assigning the entire value of each contract to its award date (motivated by the possibility of anticipation effects), and assuming that spending is normally distributed over the length of each contract.

The literature on anticipation effects (Ramey & Shapiro, 1998; Ramey, 2011; and Mertens & Ravn, 2012, among others) suggests that a firm's receiving a contract might be an economically important channel. This spending measure is constructed by assigning the full value of each contract to the contract's award date, and then aggregating to states and the appropriate time period (quarters for employment effects, years for income effects)

In contrast, the other two measures, in which spending is distributed uniformly or normally over the length of the contract, reflect a conceptual framework in which the economic effects of stimulus come directly from the economic activity created by the spending.

One potential concern about using a uniform distribution of spending over contracts is that it might take time for production to gear up, reducing economic activity at the start of the contract, and that some parts of the production process may necessarily finish before others, reducing economic activity at the end of the contract. To reflect this possibility, I construct an alternative spending measure in which spending is distributed normally over the length of each contract.

To construct data with spending distributed normally over contracts, I use $Binomial(n, .5)$ for all contracts of length n to obtain the correct discrete approximation of the normal distribution. More specifically, let m be the number of months from the starting month of the contract, with $m = 0$ for the award month. Then the value assigned to month m of a contract length n is the total value of the contract, scaled by $P(X = m|n, .5) = \binom{n}{m}0.5^m 0.5^{n-m}$. The resulting values are then aggregated to states and the appropriate time period (quarters for employment effects, years for income effects), just as for the other measures of spending.

Baseline results for all three measures of spending are shown in Tables A1 (employment) and A2 (personal income). Employment effects are slightly smaller for normally distributed spending than for uniformly distributed spending, and much smaller for contract awards than for either of the other two measures. The personal income results are statistically indistinguishable for normal and uniformly distributed spending, and not significantly different from zero for contract awards.

Given the differences in results from these three measures, it is natural to ask which measure of spending is most "correct." One way to answer this is to run horserace regressions to see which measure is preferred when more than one is present in the same regression. I run pairwise horserace regressions for each combination of two spending measures and both outcomes. The results of those regressions are summarized in Table A3.

The clear "loser" of the horserace is contract awards, which are never significant when either of the other spending measures are also included. Uniformly distributed spending

Table A1: Employment Effects by Sector for Three Measures of Spending

Spending Assumption	Total (SIC 1–7)	Total Private (SIC 1–6)	Manufacturing (SIC 3)	Private Non-Manufacturing (SIC 1–2, 4–6)	Government (SIC 7)
Awards	0.0354***	0.0507***	0.0563***	-0.00445	-0.0125
	(0.0131)	(0.0149)	(0.0112)	(0.00657)	(0.00964)
Normal	0.183***	0.225***	0.208***	0.000301	-0.0273
	(0.0224)	(0.0290)	(0.0231)	(0.0166)	(0.0245)
Uniform	0.230***	0.310***	0.274***	-0.00738	-0.0416
	(0.0318)	(0.0300)	(0.0296)	(0.0210)	(0.0364)

All effects estimated using two-year differences in both spending and employment. Standard errors clustered by state. Regressions weighted by states' 1939 populations. Time fixed effects and interaction between time fixed effects and pre-war agricultural employment shares estimated but not shown. *** p<0.01, ** p<0.05, * p<0.1

Table A2: Personal Income Effects for Three Measures of Spending

Spending Measure	Award Date		Normal		Uniform	
	0.00116	0.0730*	0.0728	0.245***	0.0415	0.248***
	(0.0399)	(0.0370)	(0.0517)	(0.0481)	(0.0586)	(0.0526)
1939 farmshare x Time FEs?	No	Yes	No	Yes	No	Yes
Observations	294	294	294	294	294	294
Within R-Squared	0.632	0.767	0.636	0.798	0.633	0.791

All effects estimated using two-year differences in both spending and employment. Standard errors clustered by state. Regressions weighted by states' 1939 populations. Time fixed effects and interaction between time fixed effects and pre-war agricultural employment shares estimated but not shown. *** p<0.01, ** p<0.05, * p<0.1

comes out a hairbreadth ahead of normally distributed spending for employment effects (with a t-stat of 3.44 vs. 3.14), while normally distributed spending is the clear "winner" for the personal income results. However, the baseline personal income results for normally and uniformly distributed spending are so close that this distinction is not economically meaningful. More generally, it is not surprising that normal and uniform assumptions about the distribution of spending produce such similar results: the two measures aggregate to broadly similar data series (reflecting the law of large numbers).

Table A3: Horserace Regressions: Employment

	Employment			Personal Income		
Awards	-0.00365	-0.0367***		-0.0161	-0.0684	
	(0.0126)	(0.0133)		(0.0412)	(0.0498)	
Normal	0.186***		0.0864***	0.258***		0.316**
	(0.0204)		(0.0275)	(0.0576)		(0.129)
Uniform		0.273***	0.138***		0.332***	-0.0846
		(0.0312)	(0.0402)		(0.0796)	(0.142)
Observations	980	980	980	294	294	294
Within R-Squared	0.876	0.880	0.879	0.798	0.794	0.798

All effects estimated using two-year differences in both spending and employment. Standard errors clustered by state. Regressions weighted by states' 1939 populations. Time fixed effects and interaction between time fixed effects and pre-war agricultural employment shares estimated but not shown. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

The relative lack of predictive power from contract awards should not be interpreted as evidence against anticipation effects. Rather, they seem to suggest that contract awards are a bad measure of anticipation. The historical narrative seems to suggest that firms often knew in advance that defense orders would be placed,[2] and of course the placement of contracts lags major news events. If the time lag between news and contract placement varies significantly (perhaps by industry), contract awards would not measure anticipation well.

The best evidence of the literature suggests that both anticipation and production channels are important for fiscal stimulus. These results suggest that contract data measures the production channel rather than the anticipation channel.

A.3 Timing of War Production

Before U.S. entry into the war (what is often called the "defense period"), war production (at comparatively modest levels) was driven by the military needs of U.S. allies, namely the United Kingdom and France. After Pearl Harbor, war production was driven largely by American military requirements.

Planning for war production during the defense period was inhibited by a military hierarchy that chronically underestimated supply needs.[3] Commencing and ramping up war production was contingent upon determination and communication of military needs; economic

[2] *Automotive Industries*, issue published December 1, 1941, p. 52.

[3] Before World War II, military needs did not have the primacy in the budget process that more contemporary scholars of public policy generally take for granted. Military appropriations did not take precedence, and between the Great Depression and the priority placed on balancing the budget during the interwar years, the military hierarchy was strongly incentivized to minimize appropriations requests. This is a stark contrast to the institutional incentives that have prevailed since 1950. For a broad discussion of the pre-war budget process and how it differed from post-war budgeting, see Schick ch. 2, pp. 15–50. For a discussion

policymakers had to beg and maneuver the military into providing the necessary information for planning during the defense period.[4] The front end timing of war production appears to result from these bureaucratic battles between military and civilian planners—and hence not from economic conditions. On the back end of war production, it is hard to overstate the extent to which the explicit goal was to produce as much war matériel as possible, as quickly as possible, while still producing necessary goods for the civilian economy. Wartime productivity increases and improvements in capacity utilization are extremely well documented.[5] This single-minded focus on speeding production is also evidenced by the many studies on worker absenteeism at war production facilities conducted by the BLS during the war.[6]

It takes significantly longer to produce and mobilize the supplies needed to prosecute a modern war than it takes to mobilize soldiers; the speed of war production was often a binding constraint on military plans.[7] In fact, it has been argued that the pace of war production was a major factor in the Allied decision to wait until the summer of 1944 for the cross-channel invasion of northern Europe[8] rather than attempting it in the summer of 1943. It is undisputed that British resistance to a cross-channel invasion—British command argued for and got the invasion of mainland Italy instead—also played a significant role; but it seems likely that the Americans acquiesced in large part due to awareness the logistical constraints.[9] It should be noted that American military delegation at the decisive strategic conference was headed by Marshall and not the more conciliatory Eisenhower.

of the specific problems relating to military estimates of supply needs, see Lacey, p. 17, and Nelson, pp. 34, 125–138.

[4]See Lacey, pp. 22-31.

[5]For a detailed case study on the shipbuilding industry, see Thompson (2001).

[6]See *Activities of the Bureau of Labor Statistics in World War II* (1947), pp. 53, 61–62.

[7]A 1959 army history emphasizes "the importance of the matériel side of modern war and the inevitable time lag between the mobilization of troops and the mobilization of their all-important supporting equipment and supplies." Smith (1959), p. 39.

[8]See Lacey, p. 4.

[9]See Atkinson (2007), pp. 9–20.

Appendix B

Appendix to Chapter Two

B.1 Data Construction and Methodology

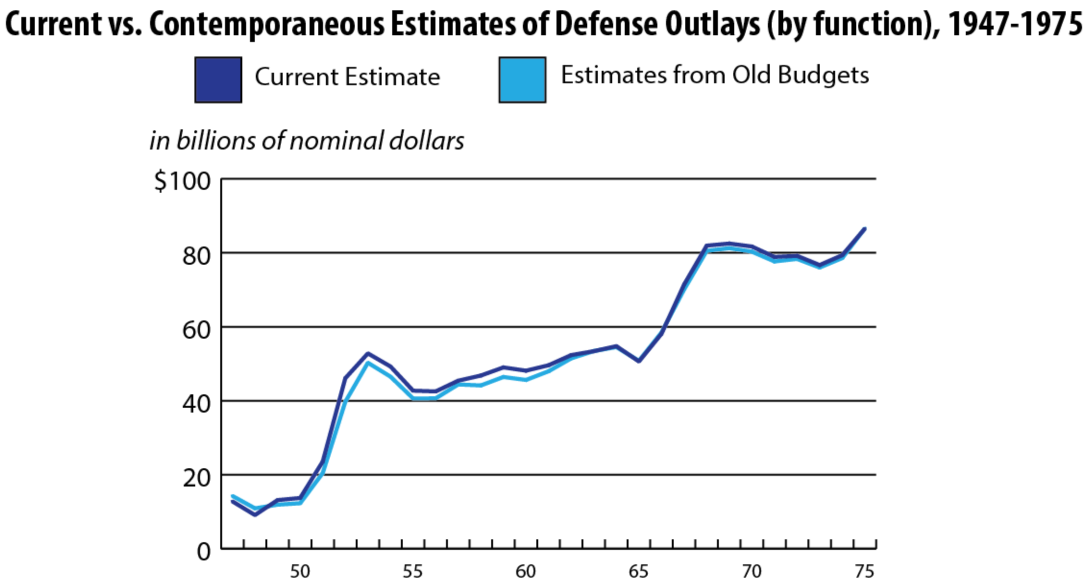
Annual budget authority totals are available from OMB for 1976 onward. These totals are broken down by category of spending, including national defense. To construct a historical series of budget authority for 1938–1975, I use United States Federal Budgets, which have been issued annually since 1923. I begin my series in 1938 because the budget was understood as placing upper bounds on spending rather than as creating spending targets prior to the New Deal; by starting in 1938, I ensure that appropriations have the same meaning throughout my series.

The budget put out each year contains figures for actual spending and budget authority in the previous year, estimates for the present year, and the President's requested amounts for the coming year.[1] I use the actual totals for the previous year from each budget, as those reflect the legal authority to obligate funds in each year.[2] As a check on the accuracy

[1]The term "budget authority" dates to the late 1960s. In the 1950s–1960s it was called "obligation authority," which is conceptually similar for our purposes. Prior to the 1952 budget, appropriations and other authorizations were reported separately. The differences between these concepts are relatively unimportant for federal spending on defense. However, while there have not been significant changes in the way funding for defense has been recorded since WWII, there have been changes to budgetary accounting that have significantly affected the ways in which funding for other areas of spending have been recorded.
The main issues are: how to account for revolving loan funds, where budget authority might have been recorded as zero while outlays representing net cash flows fluctuated; how to deal with programs such as highways, mass transit, and long-term subsidized housing for which appropriations were recorded to liquidate obligations made under multi-year "contract authority" in prior years while contract authority was not necessarily considered budget authority; and with the practice of counting appropriations from the general fund to the social insurance trust funds in the amount of dedicated tax receipts as budget authority rather than recording the amount of obligations for benefit payments in that year as budget authority. Thus it would be extremely difficult to construct a consistent historical series for many areas of non-defense spending.

[2]While I also collected annual defense spending totals by agency, I chose to focus on defense spending by category, or formally the "budget function" for national defense. This is particularly useful for the 1940s, when the agencies responsible for military spending were reorganized several times, and when a portion of

Figure B.1: Placebo Test for Construction of Historical Budget Authority



Source: Current estimates from OMB Historical Table 3.1, contemporaneous estimates taken from Budgets of the US Government, 1949 to 1977

of old budgets, I reconstruct the historical series for defense outlays (actual spending) for 1938–1975 using the same method and definition as for budget authority. I compare this historical series built from contemporaneous estimates to OMB's (current) historical series for defense outlays, which goes back to 1940. This comparison can be seen in Figure B.1. The two series do not match perfectly, but they are quite close.

It is also worth noting that the definition of the federal fiscal year changed between FY 1976 and FY 1977. Through FY 1976, the federal fiscal year ran from July 1 to June 30 (i.e. quarter 3 to quarter 2), just as most state fiscal years still do today. In the mid-1970s it was thought that moving the fiscal year back by one quarter might encourage Congress to finish its budget on time in future years, so the federal fiscal year was changed to run from October 1 to September 30 (i.e. quarter 4 to quarter 3). For purposes of comparison, I therefore use quarterly data and aggregate by fiscal year throughout the analysis.[3] For simplicity, I omit

defense spending was directly controlled by the Executive Office of the President. However, neither budget authority nor its predecessors was aggregated by category of spending in the data for 1938–1940, 1953–1956 or 1961. In 1938–1940, defense spending was such a small share of the budget that I simply used the agency-level totals instead. For 1953–1956 and 1961, I used the total by agency for that year, scaled by the ratio of spending by function to the ratio of spending by agency for the same year. This procedure avoids averaging data from other years.

[3]I aggregate data by federal fiscal year for GDP, CPI, actual defense spending as measured by NIPA, Ramey's narrative measure, and the tax shock measure derived from Romer & Romer's narrative series. For years prior to 1947, GDP, CPI, and NIPA are only available on a calendar year basis, so calendar-year figures are used. This is not a good solution, but it seems less problematic than inducing large moving

the transition quarter, which conveniently does not contain any significant changes to taxes or defense spending.

B.2 Conceptual Differences between NIPA and Budget Measures

Table 3.18B of the NIPA shows the differences between NIPA and the budget definition of spending (as measured by *budget outlays*, which are timed to reflect the disbursement of funds from the Treasury, while *budget authority* is timed to reflect authorizations but reflects the same concepts and definition of spending). BEA publishes historical data detailing these adjustments from 1968–2015. On an annual basis, total federal spending measured by NIPA varies between 96 and 113 percent of total federal spending measured using budget methodology. The year-to-year change in the ratio of the two estimates of total government expenditures is never more than 5 percent of total expenditures (as measured by NIPA), except for 2009 and 2010 (when large revisions to estimated costs of loan guarantees and net loan costs caused larger deviations between the two measures), so it seems unlikely that the definitional differences between NIPA and budget measures could account for an outsized share of the differences in year-to-year fluctuations between the two measures.

The vast majority of non-timing differences between NIPA and budget estimates of total expenditures come from differences in the treatment of pension funds, so for the most part do not affect defense spending. In the years around the Great Recession, differences in accounting for the costs of government guarantees on loans and deposit insurance also accounted for significant differences between the measures, but these differences should not affect defense spending either. The NIPA table does not distinguish between adjustments to defense versus non-defense expenditures, but it is clear from the items listed in the bridge tables published by both BEA and the Congressional Budget Office (CBO) that the vast majority of adjustments between the two measures involve non-defense spending.

There appear to be two areas of defense spending for which NIPA and budget authority may have conceptual differences. The first is in their treatment of the Military Retirement Trust Fund, which, while small compared to both the other pension programs in the budget and total defense spending, still will drive small differences between NIPA and budget measures due to differences in the measures' accounting basis for military pension funds. The second significant difference is in their treatment of receipts from sales of surplus equipment, which can sometimes appear in the budget as negative spending (since it may reduce the total dollars drawn from the Treasury by the Department of Defense) but may instead appear as receipts in the NIPA (though such payments sometimes appear as receipts in the budget rather than as "offsetting receipts"). Compared to total defense spending, both of these

average problems by averaging over calendar years. However, this may explain some of the data mismatches in the mid-1940s.

items are fairly small, and neither seems a likely candidate for driving outsized year-to-year fluctuations in the relationship between NIPA and budget authority.

Outside of these accounting differences in the Military Retirement Trust Fund and receipts for sales of surplus equipment, a dollar of defense budget authority should translate directly into a dollar of defense spending in NIPA. As explained previously, the government cannot spend without prior authorization, and budget authority is adjusted to exclude funding authorizations that never translate into spending.