ESSAYS ON STATE DEPENDENCE IN THE GOVERNMENT SPENDING MULTIPLIER

A Dissertation

Presented to

The Faculty of the Department

of Economics

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Doctor of Philosophy

By

Christopher Biolsi

May 2015

ESSAYS ON STATE DEPENDENCE IN THE GOVERNMENT SPENDING MULTIPLIER

Christopher Biolsi

APPROVED:

Bent E. Sørensen, Ph.D
Committee Co-Chair

David H. Papell, Ph.D
Committee Co-Chair

Steven G. Craig, Ph.D

Rauli Susmel, Ph.D
Department of Finance
University of Houston

Steven G. Craig, Ph.D.
Interim Dean, College of Liberal Arts and Social Sciences
Department of Economics

ESSAYS ON STATE DEPENDENCE IN THE GOVERNMENT SPENDING MULTIPLIER

An Abstract of a Dissertation

Presented to

The Faculty of the Department

of Economics

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Doctor of Philosophy

By

Christopher Biolsi

May 2015

Abstract

This dissertation is comprises three essays. The first attempts to answer the following question. Is fiscal policy more effective, as measured by the government spending multiplier, when the economy is "weak" relative to when it is "strong?" Results in the empirical literature have been mixed on this question. I use local projection techniques to estimate the impulse response functions of real output and real government spending to a shock to military spending. In addition, I attempt to endogenously estimate the level of the unemployment rate that distinguishes between states of the economy. I find that fiscal multipliers are near two at horizons of two to four years when unemployment is relatively high, compared to below 1 when unemployment is low. The second paper seeks to understand why disagreement in the emprical literature is so pervasive and if there are certain modeling choices that systematically lead to particular findings on the state dependence of the government spending multiplier. I identify eight dimensions along which many of the studies in the literature vary and determine if choices along these dimensions have a systematic impact on the results. I conclude that estimation of a state-dependent multiplier is, in general, not robust to various plausible specification assumptions. Finally, I estimate the effect of government spending at the county level using a previously little studied spending program, the Vinson-Trammell Act of 1934. Stimulated by fears about Japanese military expansion, this act aimed to build up the United States Navy to treaty allowances. I am able to identify local areas in the United States that hosted shipyards in 1934, and I estimate the effects of government spending on these areas. I find that manufacturing output, employment, and earnings all rise faster over the

course of the 1930s in counties hosting shipyards at the time of the bill's passage. Also, I see significantly faster growth in county level retail sales and a positive effect on household consumption. Attempting to scale these results to an aggregate government spending multiplier, however, leads to a wide range of estimates for the effect on overall output.

Acknowledgements

I am most especially indebted to my committee members, Bent Sørensen, David Papell, and Steven Craig, whose guidance and advice have been invaluable over the last five years. In addition, I am also grateful to Dietrich Vollrath, German Cubas, Christian Murray, Sarah Zubairy, and Aimee Chin, as well as the rest of the facuty and staff of the Department of Economics at the University of Houston. I also particularly thank my classmates (Mahboobeh Asghari, Randy Crouch, Amrita Dhar, Ashmita Gupta, Volodymyr Korsun, Wen Long, Ryan Ruddy) who helped make a long, difficult process much more enjoyable than it might have been. I need to acknowledge my family, especially my parents and all of my brothers and sisters (Theresa, Jeremy, and Colton Korth, Matthew, Andrew, Timothy, Nicole, Kim, Joseph, Bobby, and Adam[1]). Lastly, Tiffanie and Tania Tribble provided me with immeasurable emotional support throughout my graduate studies, and without them, they would not have concluded as successfully as they have. I am most grateful to them.

[1]To say nothing of Baxter, Lambeau, and Pooter

Contents

List of Figures

List of Tables

Chapter 1

Introduction to Essays on State Dependence in the Government Spending Multiplier

In this dissertation, I will address a question that has interested economists since at least the time of Keynes (1936), which is to what extent an increase in government purchases of goods and services impacts a nation's economy more generally. In particular, I will consider the notion of "state dependence" in the government spending multiplier.[1] Put simply, the state dependence that I refer to is the possibility that additional government purchases may stimulate economic activity more when the aggregate economy is in a certain condition (or "state") than they would if applied

[1]The government spending multiplier is, as I will touch upon again later, the amount of extra dollars of output that an economy produces for each additional dollar of output bought by the government.

under different circumstances. Indeed, although the empirical examination of this question has only become popular in recent years (notably since the advent of the Great Recession and the move by the Federal Reserve to hold nominal interest rates at zero), the idea that the effects of government purchases might be state dependent is actually addressed in Keynes (1936).

The focus of the essays in this dissertation is predominantly empirical. Many studies have been written trying to formalize a theory as to why government purchases may have differing effects in different states of the world, and I will briefly address these theoretical propagation mechanisms in the chapters to follow, but my primary interest is in the estimation of a pair of multipliers, one that describes the effects of increased (or decreased) government spending in a "good" state of the world and one that describes the analogous multiplier in a "bad" state of the world. I will also treat to some extent the methodological issues that attend such an estimation, both at an aggregate, or economy-wide, level and at the local, specifically the U.S. county, level.

The first substantive chapter of this dissertation, which entitled, "Asymmetric Effects of Government Purchases over the Business Cycle," goes about the estimation referred to above in a very straightforward way. Using a time series of unexpected shocks to expected government purchases (driven mainly by foreign wars) that was developed and introduced by Ramey (2011b), Owyang, Ramey, and Zubairy (2013), and Ramey and Zubairy (2014), I sort all of the post-World War II observations on United States economic outcomes into a high unemployment state and a low unemployment state by means of endogenous threshold estimation. The findings of this particular analysis are striking. The results suggest that when unemployment is relatively low (a "good" state of the economy), the effects of an increase in government purchases are fairly modest, with a multiplier well below one. This implies that the government's actions compel private actors in the economy to trim their consumption or investment activity. On the other hand, if unemployment is relatively high, the government spending multiplier is closer to two, which implies that for every dollar the government spends buying output, nearly another dollar of output is produced. The only way that this could occur is if private consumption or private investment were to rise with (and as a result of) the increase in government purchases.

The econometric approach taken in the essay described above is fairly conventional relative to much of the recent literature. A notable feature about the recent empirical literature estimating state dependence in the multiplier is that, especially when estimated on aggregate macroeconomic time series variables, there is still little consensus about the results. In fact, there are almost as many studies arguing plausibly that there is no evidence of different multipliers for purchases in different states of the world as there are those finding this state dependence. The second chapter of this dissertation, which is titled, "The Uncertain State Dependent Government Spending Multiplier," explores whether or not this is so because of the myriad of specification choices available to the econometrician. In a sense, this chapter is a very broad generalization of the first chapter, evaluating the sensitivity of the results to variations in the estimation setting. Specifically, this chapter allows variation along eight dimensions of specification choice, which results in nearly 2000 pairs of state dependent multiplier estimates. It will show that some choices are more likely to deliver incredibly high or low multiplier values or to systematically push the multiplier in either state of the world in one direction or another. In so doing, it illustrates the simple lack of robustness that plagues the estimation of the multiplier on aggregate data and provides part of the motivation for the third chapter of the dissertation.

In the third chapter, "Local Effects of a Military Spending Shock: Evidence from Shipbuilding in the 1930s and 1940s," I examine the effects of a specific government spending program, the Vinson-Trammell Act of 1934, which had, as its goal, the expansion and rehabilitation of the United States Navy's fleet. The setting of this study is local, as I am able to identify the counties of the United States that are subject to the spending authorized by this act. With a relatively underutilized panel data set that tracks local economies throughout the 1930s, I test to see if the counties that hosted shipyards at the time of the passage of this act experience better economic outcomes in the latter part of the decade than otherwise like counties that did not have a shipyard. Indeed, I will show that this was the case. In the midst of the Great Depression, counties that hosted shipyards experience manufacturing output growth six percentage points faster (at an annual rate) than their peers without shipyards. This faster growth extends even to retail sales and household consumption. The examination of this question at a local level has a number of advantages, including sharper identification of an exogenous shock to purchases, the ability to difference out potentially confounding effects of aggregate tax policy or monetary policy, and, simply, a much larger number of observations. The downside, however, is the difficulty of translating these local results to the aggregate government spending multiplier in good times and in bad times. This difficulty will be more fully explored in the essay below.

On balance, the evidence reported in this dissertation seems to argue that, in fact, government purchases are more likely to stimulate economic activity in bad times than in good times. That is, there does seem to be state dependence in the government spending multiplier. Each chapter, however, comes with its own caveats, and it still does not seem as though it is safe to conclude definitively that an increase in government purchases will boost activity enough to justify the welfare costs that may be associated with it, even at a time when unemployment is very high. Thus, more research will be necessary to continue building the body of evidence.

The next part of this dissertation is a brief survey of about ten studies in the literature that are most relevant to the analysis that follows. The three dissertation chapters alluded to in the preceding paragraphs can then be found, before a brief section that contains some concluding remarks.

Chapter 2

Review of the Literature on State Dependence in the Government Spending Multiplier

In this part of the dissertation, I will review a small number of the most influential papers to have informed the analysis that is to follow. Specifically, I consider ten sets of particularly insightful studies.

Baxter and King (1993) develop a neoclassical model in which they aim to answer a number of questions, among them being how much a permanent change in government spending changes the level of output, to what extent a temporary change in spending has different effects, how much the financing of the government spending matters, and whether productive government spending has different effects relative to unproductive spending. They find that a permanent change in government purchases induces a negative wealth effect on private households, leading to an increase in labor supply, a coincident decline in consumption, and an increase in private investment as the marginal product of capital rises. The upshot is a multiplier just above unity. In contrast, Baxter and King (1993) find that a transitory increase in government purchases has a multiplier below unity, because the investment boom that accompanies a permanent change does not materialize (private investment is crowded out along with private consumption). Output declines on impact in response to both kinds of government spending increase when taxes are raised concurrently to finance the higher spending. Although they do not consider state dependence in the multiplier they study, the results discussed by Baxter and King (1993) form the theoretical bedrock for the argument that an increase in government purchases need not be stimulative for the economy as a whole and help justify empirical estimates below unity.

On the other side of the spectrum is the work of Rotemberg and Woodford (1992), who consider a model in which oligopolistic price setting allows for an increase in government spending to lead to an increase in aggregate labor demand. This phenomenon does not occur in the work of Baxter and King (1993). An increase in labor demand can offset the effects on real wages from the increase in labor supply stimulated by the negative wealth effect. With higher wages, the response of private consumption may be positive potentially giving a multiplier greater than one. The model of Rotemberg and Woodford (1992) also relies on a countercyclical markup of price over marginal cost, a common mechanism by which New Keynesian models generate a role for aggregate demand in firms' behavior that can allow government purchases to have more positive effects on overall output. This paper is one of the earlier dynamic stochastic general equilibrium models whose influence led the way for models in which government purchases may have a multiplier above one.

The seminal work of Blanchard and Perotti (2002) has had a tremendous influence on the empirical estimation of government spending multipliers. They employ what was at the time a fairly novel technique to identify exogenous shocks to government spending through the use of vector autoregressions. Specifically, it was this paper that popularized the strategy of assuming that government spending does not react contemporaneously to innovations in output or other macro aggregates. In short, they order government purchases first and use Cholesky decompositions to identify structural shocks. With these shocks in hand, Blanchard and Perotti (2002) are then capable of estimating the impulse responses of output and government spending to these structural shocks. The identification scheme that they introduced has become standard in the empirical literature estimating multipliers, with the only other popularly employed identification being the narrative method that is discussed below. Their empirical results suggest that the multiplier on spending shocks was around one over a period of five years.

Of the papers mentioned above, none of them consider that there may be any nonlinearity in the effects of government purchases on output. Among the earliest and still most influential to do so is the work of Christiano, Eichenbaum, and Rebelo (2011), who examine how the economy may respond to a government spending shock when the nominal interest rate is held constant, such. A specific case would be when the monetary authorities is constrained by the so-called "Zero Lower Bound" from reducing interest rates when they might otherwise be inclined to. They construct a model in which agents are subject to shocks to their discount factors that may generate a suddenly drastically increased desire to save. There is no positive nominal rate that can restore equilibrium to the market for loanable funds. In this situation, a shock to government purchases raises inflation, which, when combined with the constant nominal rate, implies a sharply lower real rate of interest, spurring private consumption. In “normal” times, general equilibrium effects, in contrast, work against this consumption increase. The multiplier in their model may be as large as 3.7. Although other mechanisms have been proposed for generating non-linearity in the multiplier, such as financial accelerators and occasionally binding capacity constraints, it is the Zero Lower Bound that has generally garnered the most attention.

This theoretical breakthrough, along with the work of Eggertsson (2010) and Woodford (2011), was accompanied by new empirical methods that sought to estimate multipliers that had different values in different states of the economy. A leader in this new strand of the literature was Auerbach and Gorodnichenko (2012b). In that study, the authors combine the identification scheme of Blanchard and Perotti (2002) with a smooth transition vector autoregression to evaluate whether or not purchases were more impactful when growth was slowing or the economy was in recession. They use a centered moving average of GDP growth and a calibrated transition function to distinguish between boom states and recession states, and their result is a stark one. Auerbach and Gorodnichenko (2012b) find a multiplier well above two in low growth regimes, compared to a multiplier very close to zero when output is growing rapidly. In a similar work applied to cross country data (Auerbach and Gorodnichenko (2012a)), they again find a larger multiplier in bad times relative to good times. What is more, they find that both private consumption and investment rise in response to higher purchases in a recession, but they fall following a boost to spending in expansions. Auerbach and Gorodnichenko (2012b) remains the most influential study in the literature that finds evidence in support of a countercyclical government spending multiplier.

The next set of papers offers a new approach to the identification of government spending shocks using narrative methods and consequently comes to very different conclusions on the efficacy of government purchases as a means of stimulus, whether one is considering an overall multiplier or a state dependent one. Perhaps the headline work in this strand of the literature is Ramey (2011b). That paper develops a time series of the change in the expected present value of future military spending that results from some political or military event, based on close readings of contemporary periodicals. The motivation behind this exercise is the notion that identifying shocks to government purchases using structural VAR methods as in Blanchard and Perotti (2002) risks counting as "shocks" innovations to spending that private agents already expected and had reacted to. This approach builds on a simpler series introduced in Ramey and Shapiro (1998) that only considers military build-ups associated with wars and it is extended in later work to the Canadian context and as far back as 1890 in Owyang, Ramey, and Zubairy (2013) and Ramey and Zubairy (2014). All of these papers come to a similar conclusion as to the magnitude of the government spending multiplier, which is that it is below one, even when unemployment is fairly high (that is, even when there is a large amount of underutilized resources in the economy). These papers, then, through sharper identification of spending shocks, find evidence supporting the theoretical results of Baxter and King (1993). The mechanism at work is postulated to be the negative wealth effect that impedes consumption or costly reallocation of physical capital from the civilian to the military sector and back again.

To this point in the literature review, all of the studies under discussion deal with the aggregate effects on output of an overall increase in government spending. Of course, a substantial portion of the literature takes the tack of examining subnational economies, such as states or counties. The third chapter of this dissertation adopts this approach as well. A paper that considers a similar time period at a similar level of aggregation as my contribution is Fishback, Horrace, and Kantor (2005). The primary focus of that paper is whether grants and loans distributed during the 1930s as a part of the New Deal had a positive impact on retail sales at the county level. They adopt an instrumental variables approach, and they find that there is heterogeneity in the effects of various New Deal programs. In fact, public works projects and relief grants do provide a stimulus to local retail sales, but that funds distributed by the Agricultural Adjustment Administration (which, in fact, paid farmers not to produce) had a strongly negative effect on retail sales. One can interpret their results as at least providing evidence consistent with the notion that public spending can boost activity when the aggregate economy is in a state of severe slack.

Another recent paper that evaluates the effects of government spending shocks at a disaggregated level is Nakamura and Steinsson (2014). Their paper consists of an empirical section in which they estimate the multiplier on federal defense purchases in U.S. states (including an extension in which they interact purchases with the state's unemployment rate). They find an empirical multiplier on the order of about 1.5. In their terminology, the local multiplier is called an "Open Economy Relative Multiplier," while the aggregate multiplier is called the "Closed Economy Aggregate Multiplier." To interpret these empirical findings, the authors build a DSGE model with a number of possible variations that aims to scale the local values up to an aggregate government spending multiplier. They consider both flexible and sticky price models, utility specifications that are separable and nonseparable in labor and consumption, and different stances of the monetary policy authority. They find that their empirical results are best approximated by a sticky price model with nonseparable preferences. The aggregate multiplier is very small when monetary policy is described by a standard Taylor rule, but it can be extremely large in their model if monetary policy accommodates the increase in expenditure.

In the first two substantive chapters of this dissertation, I make use of an econometric technique developed by Hansen (2000) to estimate the threshold level of some variable across which the effects of a certain independent variable on a dependent variable differs. This procedure has a fairly straightforward intuition. Conditional on the existence of a threshold, I look for the value amongst all of the candidate values that minimizes the sum of squared errors. The particular contribution of Hansen (2000) is to develop an asymptotic distribution theory that facilitates statistical inference on the estimated threshold level. Thus, not only can I estimate a least squares threshold level in a given variable (unemployment, say) across which the effects of government purchases on output may differ, but I can also construct confidence intervals on that threshold level that help inform on the estimates of the state dependent government spending multiplier. This procedure is heavily used in the analysis to follow.

Another econometric procedure that I lean on is the local projections method for estimating impulse responses examined by Jordà (2005). Relative to the conventional means of estimating impulse response functions by iterating on the coefficient matrix estimated in a vector autoregression, local projections are much more flexible. They allow more easily for nonlinear specifications and do not necessitate that the left hand side variable be expressed in exactly the same form as the right hand side variables. These attributes make this approach particularly appealing in the context of estimating state dependent government spending multipliers, as observed by Auerbach and Gorodnichenko (2012a) and Ramey and Zubairy (2014), who were the first authors to estimate government spending multipliers in this fashion.

Chapter 3

Asymmetric Effects of Government Purchases over the Business Cycle

3.1 Introduction

Interest in the effects of fiscal policy, particularly those of government purchases, has risen in recent years as the zero lower bound on nominal interest rates has limited the ability of monetary policy to provide stimulus via conventional measures. The government spending multiplier is a statistic often used to summarize the effects of the government's purchase of goods and services, and is defined as the amount of extra output generated by an additional dollar of spending. A value above one is often considered evidence that the fiscal authorities are successfully encouraging more consumption or investment on the part of the private sector, while a figure below one implies that this private activity is being crowded out.

Although many papers have attempted to estimate an overall government spending multiplier for the economy of the United States, a growing literature, both theoretical and empirical, has questioned whether the multiplier has nonlinear properties, that is, whether it might have different values in times of economic strength and in times of economic weakness. This is important, since in "bad" times, policy makers may be more inclined to use spending measures to stimulate the economy than they would be in bad times. Indeed, this was the motivation behind the American Recovery and Reinvestment Act of 2009. This paper again seeks to answer the question of whether government purchases have asymmetric effects over the business cycle.

I will start from the framework of Owyang, Ramey, and Zubairy (2013) and Ramey and Zubairy (2014), using the identified military spending news shock of Ramey (2011b) in a local projections setting to construct impulse responses and, from these, government spending multipliers when there is relatively more or less "slack" in the economy. In contrast to these papers, I will only use data from the period after the end of World War II, and I will make other slight modifications to their specification. Most importantly, I will endogenously estimate the unemployment rate that divides "good" from "bad" times, rather than using an imposed threshold unemployment rate, as those papers do. I will show that when one allows for stochastic trends, controls for the monetary policy stance, and estimates the threshold level of the unemployment rate, the fiscal multiplier in times of economic weakness is on the order of 1.6 to 1.9, significantly higher than the corresponding multipliers in times of economic strength. These estimates fall in between the generally small multipliers (below 1 in both states of the world) found in Ramey and Zubairy (2014) and the very large recession multipliers (above 2) and negative expansion multipliers reported in Auerbach and Gorodnichenko (2012b).

As already hinted, there have been many contradictory findings in the academic literature about this issue. On one side of the debate, one can find the work of Auerbach and Gorodnichenko (2012b) and Auerbach and Gorodnichenko (2012a), which has become quite influential in a short amount of time. Using a couple of different econometric techniques, they find fiscal multipliers over 2.0 during recessionary periods and multipliers below 1.0 (or even negative, as mentioned above) during expansionary periods. They arrive at these multiplier estimates looking first at the United States only (Auerbach and Gorodnichenko (2012b)), and then in a panel setting using a large set of OECD countries (Auerbach and Gorodnichenko (2012a)). In further work (Auerbach and Gorodnichenko (2013)), they also find evidence of spillover effects of fiscal policy in one country on economic outcomes in major trading partners of that country, which effects are again stronger when the economies concerned are weak. Other papers using various estimation methods and different schemes for identifying exogenous shocks to government spending that find results in line with those of Auerbach and Gorodnichenko (2012b) and Auerbach and Gorodnichenko (2012a) include Bachmann and Sims (2012), Fazzari, Morley, and Panovska (2013), Gordon and Krenn (2014), Jordà and Taylor (2013), Candelon and Lieb (2013) and Tagkalakis (2008), among others.

On the other hand, there is a substantial literature questioning whether government spending is any more effective in times of economic weakness than in normal times. Perhaps the most prominent of these studies are those of Owyang, Ramey, and

Zubairy (2013) (ORZ) and Ramey and Zubairy (2014) (RZ). Using historical data for the U.S. and Canada dating back to 1890, a narrative series of exogenous innovations in military spending, and local projection estimation tools, they demonstrate that government spending multipliers are no higher in times of high unemployment relative to times of low unemployment. Nor are they any higher during periods where the nominal interest rate is constrained by the zero lower bound (ZLB). This result also features in the work of Hall (2009), Crafts and Mills (2013), Bognanni (2013) (who actually finds larger fiscal multipliers in expansionary periods), and Barro and Redlick (2011). This is to say nothing of the "expansionary austerity" strand of the literature, headlined by Giavazzi and Pagano (1990) and discussed more recently by Alesina and Ardagna (2013), wherein large fiscal retrenchments focused around spending cuts are found to boost output growth.

With so much disagreement, there is yet more room for contribution to this literature. I start by explaining why theoretically one might find larger government spending multipliers in times of slack than in times of strength in Section 3.2. There are several possible explanations, as I will discuss. In Section 3.3, I explain the data with which I will work, and the estimation methodology behind impulse response estimation and endogenous threshold estimation. I also will make a formal argument as to why I only consider the post-World War II period as my sample for estimating multipliers, especially given the existence of data that extends far past the Second World War that Ramey and Zubairy (2014) use in their study. Section 4.3 contains the results, while Section 3.5 contains a small number of extensions of the empirical work, specifically looking at the components of GDP and the effect of tax changes.

Finally, Section 5.6 concludes the paper.

3.2 Theoretical Framework

In recent years, there has been an explosion of theoretically oriented papers seeking to explain how government spending multipliers may be higher in "bad" economic times relative to "good" (with a handful of different ways of defining bad and good times). Many of these are of a New Keynesian bent, incorporating nominal rigidities in goods or labor markets (or both). These include studies that consider the zero lower bound on nominal interest rates as the condition distinguishing states of the world (i.e., whether the zero lower bound is binding or not). Other studies consider tightness in the labor market as the trait defining states, while still others focus on financial frictions that constrain economic activity more during recessions. My empirical analysis does not rely on the specific details of any of the models discussed below, but the purpose of this section is merely to touch upon the various mechanisms considered in the theoretical literature through which one can produce state-dependent fiscal multipliers. I will consider each of these in turn.

3.2.1 Zero Lower Bound

The effects of an increase in government purchases when the zero lower bound is binding[1] have been discussed at great length, for example in Christiano, Eichenbaum, and Rebelo (2011), Woodford (2011), and Nakamura and Steinsson (2014). The mechanism at work is actually quite simple. Essentially, when nominal interest rates are held constant (as they would be when the ZLB is binding), an increase in government purchases pushes up output, marginal cost, and, thus, inflation. Combining this increase in inflation with the constant nominal interest rate leads to a lower real interest rate, spurring an increase in private consumption and investment. The result can be some very large government spending multipliers, as seen in Table 3.1.[2]

3.2.2 Labor Market Weakness

As ORZ and RZ consider whether government spending multipliers are greater during periods when the unemployment rate is relatively high, it is important to consider some theoretical underpinning for why unemployment rates lead to especially effective fiscal policy. One recent such paper is that of Michaillat (2014). Focussing specifically on government employment (rather than government purchases as a whole), Michaillat (2014) uses a search-and-matching model of the labor market to evaluate

[1] It may perhaps be more precise to say, instead of nominal interest rates being constrained by the Zero Lower Bound, that nominal interest rates are held constant by the monetary authority, the most salient example of this in recent times being a binding ZLB.

[2] Studies that give large, if unespecified, multiplier estimates at the ZLB include Wieland (2012) and Farhi and Werning (2013).

when increases in public hiring have their least negative effects on private employment. In this model, public employment increases always crowd out private employment. The paper considers two distinct steady states, one in which real wages are high (which gives low aggregate labor demand, fairly flat quasi-labor supply, a low degree of labor market tightness, and high unemployment) and one in which real wages are low (which consequently leads to high aggregate labor demand, relatively steep quasi-labor supply, great labor market tightness, and low unemployment). An increase in public employment in the former state of the world leads to less crowding out of private employment than an increase in the latter state.

The mechanism is as follows. As discussed in the paper, the government posts vacancies to attract and hire unemployed workers, which necessarily raises labor market tightness. When unemployment is high, however, the increase in the number of public workers can be relatively large, while the increase in labor market tightness that results may be quite modest. The opposite would be the case when unemployment is low. Crowding out is subsequently less when unemployment is high, relative to when it is low. The interested reader should consult Figure 1 of Michaillat (2014) for more details. Although this study does not consider the classic government spending multiplier that I am going to estimate, the result that unemployment declines much more as a result of public hiring when it is initially high might imply that the unemployment rate is an important initial condition for evaluating the expected effectiveness of a given fiscal stimulus more generally.[3]

[3]Gordon and Krenn (2014) argue that the unemployment rate is a less attractive threshold variable for distinguishing states of the economy than the output gap, pointing to 1941, where they cite evidence that capacity constraints were being reached in several sectors of the economy, despite slack labor markets. More on this argument follows below.

3.2.3 Credit Spreads

A third strand of the literature hypothesizes that credit market frictions may also generate differential effects of spending. Canzoneri et al. (2012) develop a model in which government spending proves more effective during recessions by bringing down spreads between the interest rates on saving and borrowing. In this model, financial frictions (represented by the aforementioned spread) respond more to increases in output when they are relatively large. Thus, when the government increases purchases during a recession (when the spread is likely to be larger), the results is a more substantial reduction in the credit spread than if the same purchases took place during an economic expansion. This gives rise to a financial accelerator mechanism such that the large decrease in the credit spread leads to continued increases in output, which in turn give further reductions in the spread, and so on. This effect is muted when the spread is already low. Canzoneri, et al. (2013) compute a fiscal multiplier during recessions of 2.25, compared to a multipler value less than 1 during expansions.

There are other papers that seek channels through which government spending could boost output differentially across states. In a very interesting recent paper, Sims and Wolff (2013) develop a model in which government spending has different impacts on output and welfare depending not only on whether the economy is in recession at the time the spending hits, but also depending on whether that recession was precipitated by a negative aggregate demand shock or a negative aggregate supply shock. Galí, López-Salido, and Vallés (2007) generate large government spending multipliers in a New Keynesian model by introducing a large proportion of rule-of-thumb consumers, who consume all of their income in each period. If times of economic weakness are prone to induce liquidity constraints for a greater number of economic agents, then this model could also produce higher multipliers in those times.

3.3 Data and Empirical Methodology

This section discusses the data that will be employed in my study, and I also explain how I estimate fiscal multipliers in different states of the economy.

3.3.1 Data

Most of the data used in this study comes from the dataset made available by ORZ. It includes U.S. quarterly data spanning 1890 through 2010 on real gross domestic product, real government consumption and investment, the unemployment rate, and the GDP deflator. While the data for these variables for the post-World War II period come from the standard sources, ORZ need to construct the data for earlier years. The interested reader can learn how this data was constructed by consulting their data appendix, but since I only use postwar data, I do not discuss their extended data here. For all results presented below, I also considered estimating on the full historical sample considered by ORZ and RZ, and their findings hold throughout. While it is an interesting question as to why the results should differ so much depending on whether one starts the estimation before or after World War II, such a question is beyond the scope of this paper. For 1947 to 2010, the data on government current tax receipts come from the St. Louis Federal Reserve FRED database.[4]

The key variable used in ORZ and RZ, which I will be making use of as well, measures the change in the expected present value of military spending due to exogenous political and military events overseas. The intuition behind using this variable was first discussed in Ramey and Shapiro (1998), and it has since been expanded in both detail and length of the time series in Ramey (2011b) and ORZ. Full details of the variable's construction can be found in Ramey (2011b), but the general idea is that by combing through contemporary news accounts, Ramey (2011b) was able to put together a narrative variable that identified when events occurred that led to a contemporary change in expected future government spending. In that study, she shows that this variable Granger causes innovations to government spending identified via Cholesky decompositions in a vector autoregression (VAR), as is a common way of identifying government spending shocks in the literature (see, for example, Blanchard and Perotti (2002)). This suggests that the public had reason to expect these changes to spending and had possibly already begun reacting to them. If then, the objective is to measure the reaction of various macroeconomic variables to a shock to government spending, it is important to know when economic agents learned of the shock. Not doing so could lead to underestimating the government spending multiplier by failing to attribute movements in economic variables that occurred before the spending was actually implemented to the spending shock, or it could lead to overestimating the government spending multiplier by underestimating the size of

[4]Series ID: W054RC1Q027SBEA

the total spending package in the period in which it first shows up in the data (by not taking account of future related spending).

Another advantage of utilizing this variable as the key indicator of changes in spending is the fact that realizations of the series are likely exogenous to the contemporaneous state of the economy. Several key events marked by the variable include the fall of France to Nazi Germany (1940:2), the invasion of South Korea by Communist forces from the North (1950:3), the Soviet invasion of Afghanistan (1980:1), and the September 11, 2001 terrorist attacks (2001:3). It also includes periods of expected declines in military spending, such as the end of World War II and the Cold War. These, and the other events described by the narrative variable, were all plausibly unrelated to the state of the U.S. economy at the time that they occurred. In the empirical work to follow, the variable takes the form of the change in the expected present value of military spending as a proportion of lagged nominal GDP. The variable is plotted in Figure 3.1. Although the figure displays the entire time series of the military news shock, I use this opportunity to reiterate that my analysis will cover only the postwar period.

3.3.2 Empirical Methodology: Impulse Response Estimation via Local Projections

My estimation strategy is very similar to that employed in ORZ and RZ, with a few modifications that I will show have significant ramifications for the results. Following ORZ and RZ, as well as Auerbach and Gorodnichenko (2012a) and Auerbach and

Gorodnichenko (2013), I will use the Jordà (2005) local projection method for estimating impulse response functions of output and government spending in response to a shock to government spending, represented by the narrative military spending variable. The strengths of the Jordà (2005) local projection method are discussed at length in RZ, but among these are the ease with which it can be adapted to state dependent models, compared to, for example, structural vector autogressions, and its relative robustness to misspecification.

To compute the government spending multiplier, I will first estimate impulse response functions for the growth rate in output and government spending. The general forms for the regressions are as follows,

$$\begin{aligned}\frac{(y_{t+h} - y_{t-1})}{y_{t-1}} &= I_{t-1} * (\alpha_{y,good,h} + \beta_{y,good,h} Milnews_t + X'_{t-1}\theta_{y,good,h}) \\ &\quad + (1 - I_{t-1}) * (\alpha_{y,bad,h} + \beta_{y,bad,h} Milnews_t + X'_{t-1}\theta_{y,bad,h}) + \varepsilon_{y,t+h}\end{aligned} \tag{3.1}$$

$$\begin{aligned}\frac{(g_{t+h} - g_{t-1})}{y_{t-1}} &= I_{t-1} * (\alpha_{g,good,h} + \beta_{g,good,h} Milnews_t + X'_{t-1}\theta_{g,good,h}) \\ &\quad + (1 - I_{t-1}) * (\alpha_{g,bad,h} + \beta_{g,bad,h} Milnews_t + X'_{t-1}\theta_{g,bad,h}) + \varepsilon_{g,t+h},\end{aligned} \tag{3.2}$$

for $h = 1, 2, ..., 20$, which indexes the number of quarters after the spending shock that each regression considers. Note that a separate regression is run for each horizon, h, and the impulse response function is a plot of the coefficients on $Milnews_t$ over time. Here, y_t and g_t are the levels of output per capita and government spending per capita, respectively. I follow Hall (2009) and Barro and Redlick (2011) in specifying the dependent variables so that multipliers at each horizon of interest can be calculated directly from appropriate combinations of the coefficients on the military spending news shocks ($\beta_{y,good,h}, \beta_{y,bad,h}, \beta_{g,good,h}, \beta_{g,bad,h}$). That is, the accumulated government spending multiplier at horizon H can be calculated as $\frac{\sum_{h=1}^{H} \beta_{y,i,h}}{\sum_{h=1}^{H} \beta_{g,i,h}}, i \in good, bad.$[5] X_t is a vector of control variables. Here lies the first important distinction of my work from that of ORZ and RZ. The control vector in their work includes log levels of output and government spending per capita.[6] In addition, their specification has a quartic deterministic trend. In this study, the control vector comprises first differences of the logs of these variables and does away with the deterministic trend. The inclusion of the deterministic trend implicitly imposes that output and government spending per capita are trend stationary, and, since the question of whether or not there is a unit root in output is far from settled,[7] it seems as though allowing for a stochastic trend might be informative. In fact, in their seminal paper, Blanchard and Perotti (2002) run their model considering both deterministic and stochastic trend specifications.

I will also add some other control variables to the right hand side of my estimating equations. I will include lagged values of the shock, since, in some periods, it may be reasonable to believe that expectations-altering events may themselves cause agents to expect further such events in the future. Also, I will include lags of the three month

[5]Further discussion of the advantages of specifying the dependent variables this way can be found in RZ.

[6]An earlier version of RZ also includes log levels of tax revenues per capita. My model includes these as well.

[7]See for example the literature starting with Nelson and Plosser (1982) and continuing with Perron (1989), Zivot and Andrews (1992), Diebold and Senhadji (1996), Murray and Nelson (2000), Murray and Nelson (2002), Murray and Nelson (2004), and Papell and Prodan (2004), just to name a few.

T-bill rate, so as to take account of the critique of Rossi and Zubairy (2011), who argued that when estimating the effects of fiscal policy, it is essential to control for monetary policy, and vice versa. Note also that serial correlation of the error terms is necessarily induced due to the overlapping nature of the dependent variables, so I will apply the correction of Newey and West (1987) to address this.

The indicator variable I_t distinguishes between the good and bad states of the economy. It takes on a value of 1 when the economy is in its "good" state, and it takes on a value of 0 when the economy is in its "bad" state. In the benchmark analyses of ORZ and RZ, the economy is considered to be in a good state when the unemployment rate is less than 6.5 percent. In this study, I will take that distinction as my starting point, but I will also consider an alternative way of defining good and bad states of the economy according to the unemployment rate,[8] as I will discuss in the next subsection.

In principle, Equations 3.1 and 3.2 are estimated via ordinary least squares for each horizon, so that, if, as in this case, the longest horizon is twenty quarters, one would estimate a total of forty regressions (twenty each for output growth and government spending growth). Calculating government spending multipliers, however, requires combining the coefficients on the military news variable across some subset of these forty regression equations. In order to conduct inference on the multiplier

[8] Gordon and Krenn (2014) argue that the output gap is a better indicator variable to use to distinguish between good and bad states of the economy, as opposed to the unemployment rate. Other commenters have made this argument to me as well. I also used various measures of the output gap to distinguish between times of high and low slack in the economy. The results are mixed and rather imprecise, but a common pattern was that output gaps that tended to attribute more of the fluctuations in output to the permanent component of output, such as the decomposition of Beveridge and Nelson (1981), tended to produce higher multiplier estimates in the high slack state.

statistics themselves, including testing if they are higher in bad times relative to good times, I will need to get an idea of the covariance between the relevant coefficients across the forty equations. To do so, I will stack each of the forty regression equations in a seemingly unrelated regressions-like framework. Let $\boldsymbol{X_{i,h}}$ represent the matrix of right hand side variables for each regression equation $i = y, g$ and $h = 1, 2, ..., 20$, and $\boldsymbol{\Gamma_{i,h}}$ and $\boldsymbol{\epsilon_{i,h}}$ represent the associated coefficient vectors and error vectors, respectively. Also let $\boldsymbol{y_h}$ and $\boldsymbol{g_h}$ denote the vector of left hand side variables for each horizon as well. Then, I run the system regression

$$\begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \\ \vdots \\ \boldsymbol{y_{20}} \\ \boldsymbol{g_1} \\ \boldsymbol{g_2} \\ \vdots \\ \boldsymbol{g_{20}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X_{y,1}} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X_{y,2}} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{X_{g,20}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Gamma_{y,1}} \\ \boldsymbol{\Gamma_{y,2}} \\ \vdots \\ \boldsymbol{\Gamma_{g,20}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon_{y,1}} \\ \boldsymbol{\epsilon_{y,2}} \\ \vdots \\ \boldsymbol{\epsilon_{g,20}} \end{bmatrix}. \tag{3.3}$$

By estimating the regressions in this framework, I can not only construct impulse response functions by plotting the coefficients on the military news variable, but can conduct inference on calculated multiplier statistics as well.

3.3.3 Empirical Methodology: Endogenous Threshold Estimation

ORZ and RZ use the unemployment rate as their threshold variable for distinguishing between good and bad states of the economy, and they set 6.5 percent as their threshold level. Of course, there is no theoretical significance attached to 6.5 percent.[9] This raises the question of whether some other level of the unemployment rate might prove a more meaningful threshold. In other words, by arbitrarily setting the threshold level at 6.5 percent, are some "good" periods being inappropriately thrown in with the "bad" ones, or vice versa?

To address this question, I will make use of the sample splitting technique of Hansen (2000). This paper takes its cue from the literature examining estimation of structural breaks and their associated sampling distributions. The intuition behind this technique is essentially as follows. For each possible value of the threshold variable, which in this case is the unemployment rate, the algorithm estimates the regression as though that value were the true threshold value. It then chooses the threshold level that minimizes the sum of squared errors. Hansen (2000) also develops a distribution theory associated with this procedure, so that I can determine if there are significant threshold effects in each regression and construct a confidence interval around the estimated threshold in the unemployment rate.

[9] ORZ and RZ cite recent comments by former Federal Reserve Chairman Ben Bernanke to the effect that unemployment would have to fall below that level before monetary policy could be tightened. In fact, more recent comments by Bernanke's successor, Janet Yellen, indicate that the "threshold" below which policy may be tightened has shifted lower. See http://www.bloomberg.com/news/2014-03-19/fed-links-rate-outlook-to-range-of-data-drops-6-5-threshold.html

In a similar exercise, Jordà (2005) estimates thresholds in several candidate variables looking only at the one-period ahead regression ($h = 1$). Because understanding fiscal multipliers necessitates considering longer horizons than just one period ahead, I estimate separate thresholds for each regression horizon ($h = 1, 2, ..., 20$) separately and adopt as the threshold value of the unemployment rate the median of the twenty estimated values.[10]

3.3.4 The Post-WWII Period

One of the main contributions of ORZ and RZ is to construct historical quarterly data on GDP and government spending in the United States that dates back to the late nineteenth century. By doing this, they argue that they are better able to exploit considerable variation in a number of key variables, including government spending, by taking account of such episodes as the two world wars and the Great Depression. This method has the additional nontrivial advantage of nearly doubling the sample size. While these episodes carry with them some caveats (such as rationing during World War II or increased patriotism during the same time, which may have had offsetting effects on output), through robustness checks, RZ demonstrate that their results are not terribly sensitive to these issues.

I will argue that, while the longer data series is exceptionally useful and interesting in its own right, for the purposes of measuring the government spending multiplier during times of relatively high or low slack in the economy, it is not advisable to pool

[10] In practice, the estimated threshold values were the same for a substantial number of the horizons.

the prewar data with the postwar data. This is because it is very unlikely that the prewar data series and the postwar series are derived from the same data generating process. As Gorodnichenko (2014) shows in his discussion of RZ, the volatility of government spending growth is of a much greater magnitude before the war. This could be down to the time series process actually changing, or it could be because of the volatility of the interpolator series used by RZ.[11] Related to this point is the argument that mixing interpolated data with more homogeneously collected data after World War II can confound statistical tests, as shown byMurray and Nelson (2000). The specification put forward in their paper also does not account for the possibility of a structural break in the government's share of output, a point made by Gorodnichenko (2014) and Gordon and Krenn (2014).

A major question raised by RZ, however, is whether the military news instrument has enough explanatory power for government spending in the postwar period, in particular when unemployment is relatively high. This is especially salient, given the use of the military spending news variable as an instrument and the means by which multipliers are calculated. As stated above, the multiplier for a given state of the economy is calculated as $\frac{\sum_{h=1}^{H} \beta_{y,i,h}}{\sum_{h=1}^{H} \beta_{g,i,h}}, i \in good, bad$. To simplify notation and to fix ideas, suppose that the coefficients come from a univariate regression of output or government spending on the military news variable at any given horizon $h = 1$. In that case, the multiplier is

$$Mult = \frac{\beta_{y,i,1}}{\beta_{g,i,1},}\ i \in good, bad\ . \tag{3.4}$$

[11]In fact, because of the interpolator series' volatility, RZ warn against using a Cholesky decomposition to identify government spending shocks in a VAR framework with this prewar data.

Suppressing the notation indicating the state of the economy, this implies that

$$Mult = \frac{\sum_{t=1}^{T}(Y_t \times MilNews_t)}{\sum_{t=1}^{T}(Milnews_t^2)} \div \frac{\sum_{t=1}^{T}(G_t \times MilNews_t)}{\sum_{t=1}^{T}(Milnews_t^2)} \tag{3.5}$$

and, further, that

$$Mult = \frac{\sum_{t=1}^{T}(Y_t \times MilNews_t)}{\sum_{t=1}^{T}(G_t \times MilNews_t)} \ . \tag{3.6}$$

This expression shows clearly that the use of the military spending news variable and the local projections method in this way leads to an instrumental variables interpretation of the multiplier. Thus, the explanatory power of the news variable for future growth in government spending is a critical element of the analysis.

To evaluate the strength of the instrument, RZ present a series of tests showing the F-stats for regressions of the form

$$G_t = \alpha + \beta_1 Milnews_{t-1} + \beta_2 Milnews_{t-2} + \beta_3 G_{t-1} + \beta_4 G_{t-2} + \beta_5 Y_{t-1} + \beta_6 Y_{t-2} + trend + \varepsilon_t, \tag{3.7}$$

where G_t is the log of per capita government spending in year t, Y_t is the log of per capita output in year t, and $Milnews_t$ is the identified news shock to military spending. They find that the instrument relevance of the military spending variable (as measured by the F-stat) is very low in the postwar period with high unemployment. In the whole sample, the F-stat for the news variable in bad times clears ten.[12] RZ cite the relative weakness of the military spending news shock as an instrument as a key reason for putting less weight on the postwar results they present in their paper. Gorodnichenko (2014), however, points out that while the instrument relevance is

[12] See Table 1 of Ramey and Zubairy (2014).

weak in the very short term, it is fairly strong in the medium term. This is important, because the multipliers that RZ construct (and that I will construct in this paper) depend on the reaction of government spending not only within a year, but over the course of several years.

Table 3.2 gives the F-Statistics for the null hypothesis that the military spending instrument has no explanatory power for the growth in government spending over horizons of two and four years (the horizons over which I will construct multiplier estimates). RZ consider a cutoff of 10 for sufficient instrument relevance. It can be seen from the table that when unemployment is relatively low,[13] the military news instrument has fairly good explanatory power for government spending growth, with F-Stats of 13.75 at two years and 66.45 at four years. In contrast to the findings of RZ, however, even when unemployment is relatively high, the instrument has decent explanatory power, with a statistic of 15.34 at a four year horizon, greater than that for the low unemployment state at two years. With this evidence in hand, I continue considering only the postwar period.

3.4 Results

In this section, I will present my estimation results, first considering specifications in which I use the ORZ and RZ threshold value of the unemployment rate (6.5 percent), and then estimating the threshold endogenously.

[13]The definition of unemployment as relatively high or low is determined by the endogenously estimated unemployment rate which is discussed further in Section 3.4.2. Also, the first stage regression is exactly that of Equation refgbench.

3.4.1 Fixed Unemployment Threshold

Here, I will estimate four specifications in which I used the fixed unemployment rate threshold. First, I will employ exactly the same specification as in ORZ and RZ. The first modification will be to allow stochastic rather than deterministic trends. In the third specification, I augment the control variable set with lagged values of the three month T-Bill and the military news shock. In the fourth, I allow only the coefficients on the constant and the curent military news shock (which is the coefficient of interest) to switch across states.

3.4.1.1 Replicating Owyang, Ramey, and Zubairy (2013) and Ramey and Zubairy (2014)

The first specification is nearly identical to that used in ORZ and RZ.[14] Estimating the regression system on the dataset that begins in 1948, the results are rather erratic. Output rises in response to a news shock, but not significantly, when unemployment is relatively low. On the other hand, when unemployment is above the threshold, there is an insignificantly negative fall in GDP over the first four years, after which it begins to rise. The reaction of government spending is measured more precisely, but, even here, the results may raise eyebrows. In the high unemployment state, for example, government spending begins to fall after about two and a half years. When unemployment is low, spending lurches sharply higher before slowing

[14]More precisely, the differences from RZ are slight, because I include tax revenue in the conditioning set on the right-hand side, which is not done in ORZ but is the case in older drafts of RZ. Also, I do not have exactly the data they have, as their study extends the sample an extra couple of years. Also, RZ control for only two lags of the various right-hand side variables, whereas I control for four.

down almost as abruptly after six quarters. The estimated multiplier statistics also seem incredible, especially for the high unemployment state. These can be found in Table 3.3. Over two years, one gets a multiplier of -2.11, significantly negative at the 90 percent confidence level. There is no theoretical model of which I am aware that produces such large negative multipliers. The model that does come closest to producing such a figure may be that of Baxter and King (1993), who, in a neoclassical model, find a multiplier on government purchases of -1.1. There are some differences between their model and the circumstances that likely prevail in the sample that I study. The multiplier that arises from that paper is not assigned to periods of relatively weak activity, and the spending they consider is permanent and financed by contemporaneously higher distortionary labor taxes. These conditions are not likely to describe the military spending shocks recovered in times of relatively high unemployment. Adding to the confusion, the same multiplier measured over a four year period is extremely large at 32.886, which is similarly unjustified by any conventional theoretical model, and the confidence interval around this figure is so large as to be effectively uninformative.

As I will show below, imposing the trend stationarity assumption seems to be a source of the volatile impulse response functions found. When one allows stochastic trends, the coefficient estimates are much more well behaved, leading to multiplier estimates more in line with theoretical predictions. I discuss these in the next subsection.

3.4.1.2 Stochastic Trends

At this point, I will modify the regression specification slightly, so as to assume stochastic trends as opposed to deterministic trends. In effect, this means that rather than including four lags of log levels of real output per capita and real government spending per capita, as well as deterministic time trends, in the control vector (as is the strategy pursued in ORZ and RZ), I will include four lags of first differences of real output per capita, real government spending per capita, and real tax revenues per capita. A reason for taking this approach is that, if there is a unit root in these series, then including lagged log levels and a deterministic trend produces a system subject to the Nelson and Kang (1981) critique, in that the cyclical components of the series are incorrectly specified. This could potentially confound the estimation results if the resulting conditioning set includes spurious data.

To justify specifying the right hand side variables in differences as opposed to levels, Table 3.4 presents results from unit root tests in the main control variables. I use DF-GLS tests for a unit root, which Elliott, Rothenberg, and Stock (1996) demonstrate to have power that approaches the asymptotic power envelope. Table 3.4 gives the results of these tests. As can be seen from the table, the hypothesis of a unit root cannot be rejected in any of these series. This suggests that the appropriate specification for the regressions is one in which these control variables are specified in first differences without a deterministic time trend.

The impulse response functions for output and government spending in high- and low-unemployment states are found in Figure 3.3. Looking at the impulse responses for the period beginning from 1948, it appears that employing a difference-stationary, as opposed to a trend-stationary, approach produces results that make much more intuitive sense, especially in the high-unemployment state. That is, neither the output growth impulse response nor the government spending impulse response turns quickly negative in response to a positive shock to spending. The output response also appears to be more precisely estimated. Even in the low-unemployment state, the shape of the government spending impulse response function is somewhat smoother. It does not have the sharp upturn and steep decline seen when deterministic trends are used. It is interesting that the difference-stationary specification seems to make such a difference when considering only the relatively homogenously collected data of the postwar period, especially in light of the observation of Murray and Nelson (2000) about the difficulty of rejecting the unit root hypothesis when looking at quarterly postwar data.

Table 3.5 contains the multiplier statistics for the difference-stationary specification. For both time horizons considered, the low unemployment multiplier is significantly less than unity at the 90 percent confidence level. Low unemployment multipliers are 0.414 at a two-year horizon and 0.541 at a four-year horizon. In the high-unemployment state, the difference is marked. The multiplier statistics are much more well-behaved in the difference-stationary specification compared to the trend-stationary specification. Although the multiplier at the two-year horizon is still fairly low at 0.224, it is no longer significantly negative, a result that had very little theoretical justification. At the four-year horizon, one gets a value of 1.392, which is not statistically different from 1, but is at least significantly positive, and is a much more plausible value than a fiscal multiplier over 30 found in the deterministic trend regressions, as well as falling within the plausible range over which "reasonable people can argue" given by Ramey (2011a) and Ramey (2012). One cannot reject the equality of the two-year multiplier across high- and low-unemployment states, but there is more evidence of a larger high-unemployment multiplier at the four-year horizon, although it is still not significant at the 90 percent confidence level. The better behavior of the multiplier statistics, especially in the postwar sample, leads me to conclude that the difference-stationary specification is the more informative, more appropriate specification.

3.4.1.3 Expanded Control Sets

In this section, I augment the difference stationary specifications by including four lags of the military spending news variable and four lags of the three month T-bill rate as controls. I include lags of the military spending news variable only out of an abundance of caution, because although Ramey (2011b) and ORZ constructed it so as to only include changes in the expectations of future military spending, which should necessarily control for the information already in hand, it is clear from looking at a plot of the variable (See Figure 3.1) that some periods in history, such as the years leading up to and including World War II or the years during the Vietnam conflict, were more susceptible to these shocks than others. By including lags, I hope merely to control for the political environment in which the shocks hit. I include lags of the T-Bill so as to to control for the stance of monetary policy at the time of the spending shock. The importance of taking account of monetary policy when investigating the macroeconomic effects of fiscal shocks is emphasized in such papers as Rossi and Zubairy (2011), Davig and Leeper (2011), Ilzetzki, Mendoza, and Végh (2013), and Zubairy (2014). The other rationale for including the T-Bill rate has to do with concerns about omitted variable bias. In the construction of the military spending news variable, Ramey (2011b) uses contemporary Treasury bond rates to discount spending changes expected to take place in the future. These interest rates are highly likely to be correlated with the state of the economy at the time the spending shock hits, and are also likely to influence the paths of output and government spending going forward. Since a higher interest rate will produce a smaller spending shock and could be a dampening influence on output growth, I would expect a downward bias in the coefficient on military news.

The impulse response functions for this expanded specification are found in Figure 3.4. Especially at shorter horizons, the downward bias suspected in the coefficient on the military spending news variable seems to be evident. By controlling for the lagged T-Bill rate and lagged news variable[15], I find considerable differences in both the impulse response functions and the multiplier statistics over the high- and low-unemployment states of the world. In the low-unemployment state, output rises modestly, but its increase is not statistically significant at most horizons. Government spending rises strongly, moreso than does output, and is precisely estimated, leading to small multiplier statistics. At the two-year horizon, I get a multiplier of 0.456, and at four years, I get 0.488. See Table 3.6.

In the high-unemployment state, output responds strongly positively to a military

[15]Including lagged first differences of the T-Bill rate instead of levels does not affect the results (A unit root cannot be rejected in the T-Bill rate in the data).

news shock, especially at longer horizons, although even at two-year horizons, the reaction is stronger than that of government spending. The response trajectory is significant at the 90 percent confidence level for both variables. Again, Table 3.6 contains the multiplier statistics. At two years, the fiscal multiplier is 3.572, with a 90 percent confidence interval of 2.480 to 4.663. At four years, it is 3.094, with a confidence interval of 2.383 to 3.804. The differences between the government spending multipliers in the low- and the high-unemployment states are large and significant. Thus, the results I get for the postwar sample contradict those found in ORZ and RZ, and they are in line with many of the large fiscal multipliers found in New Keynesian theoretical models, like those referenced in Table 3.1.

3.4.1.4 Parsimonious Specification

In this section, I experiment with a more parsimonious specification, in which only the constant and the coefficient on the military spending news variable is allowed to switch across the states of the economy. The primary purpose of this exercise is as a robustness check, to test the sensitivity of the results to the particular specification, in which all of the variables' coefficients switch across the regime. In particular, this will be useful when I conduct endogenous threshold estimation, where I will be more interested in targeting threshold effects for only the military spending news variable. On one hand, conducting the estimation this way allows me to economize on degrees of freedom, although I lose the ability to condition the spending shock on state-dependent control sets. This may matter if mean reversion properties differ across the two states. The regression system now takes the form

$$\begin{aligned}\frac{(y_{t+h} - y_{t-1})}{y_{t-1}} &= I_{t-1} * (\alpha_{y,good,h} + \beta_{y,good,h} Milnews_t) \\ &\quad + (1 - I_{t-1}) * (\alpha_{y,bad,h} + \beta_{y,bad,h} Milnews_t) + X'_{t-1}\theta_{y,h} + \varepsilon_{y,t+h}\end{aligned} \tag{3.8}$$

$$\begin{aligned}\frac{(g_{t+h} - g_{t-1})}{y_{t-1}} &= I_{t-1} * (\alpha_{g,good,h} + \beta_{g,good,h} Milnews_t) \\ &\quad + (1 - I_{t-1}) * (\alpha_{g,bad,h} + \beta_{g,bad,h} Milnews_t) + X'_{t-1}\theta_{g,h} + \varepsilon_{g,t+h}.\end{aligned} \tag{3.9}$$

I continue to use the expanded control set introduced in Section 3.4.1.3. Figure 3.5 contains the impulse response functions. Here, it is the case that the results are not quite robust to restricting the number of coefficients that switch, at least at shorter horizons. First, looking at the impulse response functions, I see little change in the shape of the response, but it is less precisely estimated, especially in the high-unemployment state. I cannot even say that the output response is significantly positive at short to medium-term horizons. Multipliers in low-unemployment states rise slightly, but there is little qualitative difference compared to the specification allowing all coefficients to change. In the high-unemployment state, the results are more sensitive. The two-year integral multiplier drops from 3.572 to 1.278, and it is no longer statistically different from 1. In fact, the lower bound on the 90 percent conffidence interval is not much higher than zero. Thus, I also cannot reject that the multiplier at a two year horizon is the same in the high- and low-unemployment states of the world. At a four-year horizon, the results are more robust. By the four-year point, the output impulse response is significantly positive, and the integral multiplier is still a quite large 2.284, close to the theoretical multipliers displayed in Table 3.1. It is also the case that the hypothesis of equally sized multipliers in the high- and low-unemployment states can be rejected with 90 percent confidence.

In this section, then, my findings of larger fiscal multipliers in "bad" states of the economy relative to "good" states of the economy are somewhat muted, although, again, this particular specification imposes common mean reversion dynamics in both high- and low-unemployment states.

My next objective to apply the Hansen (2000) sample splitting method to endogenously estimate the threshold level of the unemployment rate distinguishing between good and bad states of the economy.

3.4.2 Multipliers with Endogenous Threshold Estimation

In this section, I turn to one of the main contributions of this paper, which is endogenously estimating the threshold level of the unemployment rate that distinguishes "good" times from "bad" times, using the method decribed in Hansen (2000). In so doing, I will employ three alternative specifications. The first specification will find the least squares estimate of the threshold level of the unemployment rate allowing the entire coefficient vector to switch between the regimes. The second, employing the more parsimonious specification seen in Section 3.4.1.4, estimates the threshold level allowing only the coefficient on the military spending shock to change. The third estimates the threshold level with only the military spending coefficient changing, but then conducts impulse response estimation allowing all of the coefficients to switch across the threshold level estimated using only the military spending shock. In other words, I take the following approach. In the notation of Hansen (2000) (See his equations (1)-(2)), I estimate

$$y_i = \theta_1' x_i + e_i, q_i \leq \gamma \tag{3.10}$$

$$y_i = \theta_2' x_i + e_i, q_i > \gamma, \tag{3.11}$$

and the procedure detailed in Hansen (2000) gives the least squares estimate of γ. In Specification (1), the vector x_i includes all of the independent variables, including the military spending news shock, when estimating γ. Then, the actual impulse response estimation follows Equations 3.1 and 3.2.[16] In my Specification (2), I estimate γ using only the military spending news variable in the vector x_i as well as a constant. Then, impulse response estimation follows the parsimonious model represented by Equations 3.8 and 3.9. Finally, in Specification (3), I estimate the threshold with an x_i vector containing only a constant and the military spending news shock, as in the second model, but I estimate impulse responses using Equations 3.1 and 3.2. I do this, because I may only be interested in finding the level of the unemployment rate across which the reactions of output and government spending to the military news shock differ (thus, I include only the military news shock in the threshold estimation). Conditional on this estimated threshold, however, I may still want to include the full control vector that contains lags of output growth and the nominal interest rate, among others. The potential problem that could arise in the first specification is that the threshold estimation may be confounded by different coefficients on, say, lagged output growth, which would be the case if there is a high-growth phase of

[16]It should be noted that threshold estimation is conducted with output growth as the dependent variable, and the government spending regressions use the threshold level of unemployment estimated from the output regression.

the business cycle following recessions (see, for example, Kim and Murray (2002) and Kim, Morley, and Piger (2005)). This third specification allows me to avoid this problem.

There may be some doubt about using the unemployment rate as the threshold variable in this procedure. In developing a distribution theory on the estimation of the threshold variable, Hansen (2000) assumes that the threshold variable is strictly stationary (See Assumption 1.1 in Hansen (2000)), ruling out trend-stationary and integrated processes. A unit root in the unemployment rate cannot be rejected with conventional tests, but, because it is bounded from above and below, it cannot technically be an integrated process. This is the argument advanced in King and Morley (2007). Also, Nelson and Plosser (1982) exploit the presumed stationarity of the unemployment rate as a benchmark against which to compare their results from unit root tests on other macroeconomic series. For this reason, I feel confident using it as the threshold variable in this context.

Table 3.8 gives the bootstrapped "asymptotic p-values" (using the terminology of Hansen (1996)) for the null of no threshold effect in a linear regression specified as in Equation 3.1 (Specification (1)), or as a simple regression $\frac{(y_{t+h}-y_{t-1})}{y_{t-1}} = I_{t-1} * (\alpha_{y,good,h} + \beta_{y,good,h} Milnews_t) + (1 - I_{t-1}) * (\alpha_{y,bad,h} + \varepsilon_{y,t+h}$ (for Specifications (2) and (3)). Following Hansen (1996), the purpose of conducting this procedure is to get some evidence that threshold effects exist in these processes, which would provide some support for estimating the threshold level of the unemployment rate. I run the threshold effects test at each horizon $h = 1, 2, ..., 20$ individually. There are certainly problems associated with this approach. Specifically, one of the assumptions made in Hansen (1996) is that the error terms are independent and identically distributed, or at the least form a martingale difference sequence (See Corollary 2 in Hansen (1996)). When the horizon is longer than one, this condition is clearly violated. Still, for Specifications (2) and (3), I can still reject the null hypothesis of linearity at conventional significance levels when the horizon is only one period, when the conditions laid out in Hansen (1996) are most likely to be satisfied. I consider this as justification for estimating the least squares threshold level.

Table 3.9 gives the least squares estimates of γ, the threshold level of the unemployment rate across which output growth responds differently to a military spending news shock, estimated following the procedure of Hansen (2000). As when testing for the existence of threshold effects, the threshold was estimated separately for each horizon $h = 1, 2, ..., 20$. In order to estimate impulse responses to a shock that hits in one coherent state of the world, I must choose one threshold level in order to estimate the benchmark equation system, so I will take the median estimated threshold across horizons for each specification. That is, for Specification (1), I will use an unemployment rate of 5.14, and for Specifications (2) and (3), I will use 5.71, which, as can be seen from Table 3.9 are these median values. The impulse response functions for the three specifications are found in Figures 3.6 to 3.8.

Specification (1) offers the strongest evidence against the notion that fiscal multipliers could be greater during periods of high unemployment, as the results of ORZ and RZ stand up to endogenous threshold estimation. When looking at the post-war sample only, in general, the responses are not very precisely estimated, but there seems to be a stronger response of output to a spending shock when unemployment is below the endogenously estimated threshold level of 5.14. Whether unemployment is above or below the threshold, government spending exhibits a greater response than output, leading to small government spending multipliers and no significant differences between the multipliers across states. The estimated multiplier statistics for this specification, as well as the next two, can be seen in Table 3.10.

As argued above, however, it is possible that Specification (1) is not capturing the dynamics that are at the heart of this question. For example, if the process governing output growth is asymmetric, responding differentially to its own lags depending on whether the economy is in recession, it is possible that the endogenous estimation of the threshold unemployment rate is picking up these effects, rather than differences in its reaction to the military spending shock. That is, this specification may be taking account of asymmetric mean reversion effects in output, as opposed to different reactions to the military news shock. Besides this, from Table 3.8, it is not clear that there are meaningful threshold effects in the regression when it is specified this way. This leads me to consider Specifications (2) and (3), which both estimate the threshold unemployment rate by only considering a switching coefficient on the spending shock (and which offer evidence of threshold effects). I find that the results from these specifications look more like the main results presented in Section 3.4.1.2, Section 3.4.1.3, and Section 3.4.1.4.

Looking at Specification (2), I recover large multipliers in the high-unemployment state of the economy. When I estimate the threshold level of the unemployment rate using only the military spending shock and the impulse responses using the framework represented by Equations 3.8 and 3.9, the government spending multiplier is 1.605 after two years and 1.678 after four years, both significantly greater than unity at the 90 percent confidence level. I cannot say that the high-unemployment state multiplier is significantly greater than the low-unemployment state multiplier at a two-year horizon, but I can say so at a four-year horizon.

If I consider instead Specification (3), which meshes the threshold estimation introduced in Specification (2) and the impulse response estimation of Specification (1), the fiscal multipliers are even larger.[17] At two years, when unemployment is above the estimated threshold level, the government spending multiplier is 1.808, and at four years, it is 1.918. These are both significantly greater than 1, and they are both significantly greater than the multipliers when unemployment is below the threshold at each respective horizon. What is more, these estimates are in line with the theoretical multiplier predictions introduced in Section 3.2. It is clear from the impulse response functions given in Figure 3.8 that output rises significantly when unemployment is above the threshold, while, for most horizons, it does not do so when unemployment is below the threshold. In this case, the multiplier is 0.274 at two years, and 0.427 at four years. Even as government spending rises more when unemployment is above, as opposed to below, the threshold, its response is dwarfed by that of output.

At this point, it must be acknowledged that the confidence intervals reported for each multiplier estimate are constructed via the delta method for nonlinear combinations of parameter estimates, and are conditioned on the estimated threshold

[17]Using this estimated threshold, there are 51 non-zero observations of the military news variable when unemployment is relatively low and there are 18 non-zero observations when unemployment is relatively high.

estimate being the "true" level of the unemployment rate that distinguishes between good and bad times in the economy. Of course, this "true" threshold is a latent variable (and potentially even nonexistent). It may be appropriate to construct confidence intervals taking into account this uncertainty on the threshold unemployment rate level. Table 3.11 replicates Table 3.10, but replaces delta method confidence intervals with Bonferroni-type bounds, as suggested by Hansen (2000). Essentially, I estimate multiplier values for every threshold level of the unemployment rate within the confidence intervals displayed in Table 3.9 and report the highest and lowest values. Clearly, especially for the bad state of the economy, inference becomes much more difficult, with the confidence intervals spanning a wide range of possible values. This is due to the fact that, as the candidate threshold level of the unemployment rate gets larger, the number of nonzero observations of the military spending news shock in the bad state of the economy becomes quite small. Recall that there are only 18 when the point estimate of the unemployment rate threshold is considered (for Specifications (2) and (3)).

Figures 3.9 and 3.10 show the responses of output and government spending to military spending shocks equal to the median that actually occurred in each of the high and low unemployment states. The purpose of this exercise is merely to demonstrate that the multiplier statistics estimated in this paper are not the product of finding the right specification. Rather, a relatively granular look at the data reveals that output responds more forcefully to a military spending shock in a period of high unemployment. Figure 3.9 displays the case of relatively high unemployment. The magnitude of the median (non-zero) shock is 0.23% of lagged GDP.[18] The plot shows that, after a short spike upwards, government spending declined, hitting a trough of 6% below the level prevailing the period before the shock, before stabilizing at around 3% lower. Output grew 4% in the year after the shock (although not as quickly as government spending), and it plateaued at a level about 6% higher than its level in the period before the news hit the economy. In this particular case, then, it seems as though the multiplier was negative.

Figure 3.10 illustrates the case of a shock that occurred when unemployment was above the threshold. Its magnitude is equal to the median size of all shocks that were realized when unemployment was relatively high, about 0.36% of GDP.[19] After the realization of this news, both government spending and output grew rapidly, each around 20% over five years, in contrast to the sequence that followed a shock during a period of low unemployment. What is more, the unemployment rate in the economy before this period was about 5.9%. This means that RZ would have counted this shock as having taken place during a relatively non-slack state, since the unemployment rate was below their cutoff. In my paper, with endogenous threshold estimation, it is included among periods of high slack. This provides further evidence of the usefulness of the endogenous threshold estimation technique.

With these results, I can place my study in between the findings of RZ and those

[18]Based on the narrative provided by Valerie Ramey on her website (Ramey (2014)), it appears that this spending shock (1952:Q3) refers to a newly established determination to maintain relatively high security spending after the end of the Korean War, which had not been expected.

[19]This shock occurred in 1962:Q1, and, from Valerie Ramey's narrative, was driven predominantly by tensions with Cuba (this period is sandwiched by the Bay of Pigs invasion the prior spring and the Cuban Missile Crisis, which occurred later in the year) and the Kennedy Administration's goal of putting a man on the moon.

of Auerbach and Gorodnichenko (2012b) and Auerbach and Gorodnichenko (2012a). While ORZ and RZ find multipliers below unity in all states of the world, I find significantly different multipliers in the two states I consider, where the fiscal multipliers in good times is positive but small (less than one), and the multiplier in bad times is above one and significantly so. Although Auerbach and Gorodnichenko (2012b) and Auerbach and Gorodnichenko (2012a) find larger "bad" state multipliers (above two), they find negative "good" state multipliers, which is also not very compelling. I would argue that the results presented here are consistent with the growing theoretical literature on state dependent effects of fiscal policy,unlike RZ, but that give multiplier values in times of low slack that do not seem implausibly low either. Table 3.12 shows the multiplier values from my preferred specification (endogenous threshold estimation (Specification 3) with stochastic trends, monetary policy, and expanded interactions) alongside those of RZ and Auerbach and Gorodnichenko (2012b) and Auerbach and Gorodnichenko (2012a).

3.5 Extensions

This section will attempt to extend the analysis somewhat, with a look at what components of private activity are driving multipliers greater than one when the unemployment rate is relatively high. Also, I consider the other side of fiscal stimulus, specifically whether tax changes have differential effects according to the state of the economy. To do this, I will exploit the narrative tax series constructed by Romer and Romer (2010).

3.5.1 Consumption and Investment Multipliers

Multipliers that reach above the level of unity imply that some component of private activity is also increasing in response to the government spending shock, in addition to government spending itself. That is, some measure of private activity is being "crowded in." In many empirical studies looking at aggregate fiscal multipliers, such as Blanchard and Perotti (2002), researchers tend to find a positive response of private consumption, with negligible or negative effects on private investment. In fact, several theoretical papers have emerged that seek to explain this positive consumption response, since the negative wealth effect of a government spending shock tends to depress private consumption in standard neoclassical and New Keynesian models. With this in mind, this subsection seeks to examine whether large multipliers in high-unemployment periods are also due to a large positive consumption response.

To do so, I employ Specification (3) used in Section 3.4.2, which takes the endogenously estimated threshold unemployment rate for output, letting only the military news shock enter the threshold estimation equation, but letting all right-hand side variables have switching coefficients in the main regression for consumption or investment. I consider the responses of consumption of nondurables and services, consumption of durable goods, and gross private domestic investment. The regression equations are exactly those of Equations 3.1 and 3.2, only with the macro variable of interest substituting for output on the left hand side of the regressions. The results are found in Table 3.13 and in Figure 3.11, Figure 3.12, and Figure 3.13.

When unemployment is relatively low, a government spending shock as identified by the military spending news variable causes a short-run significant decline in consumption of nondurables and services (hereafter referred to only as "consumption"), with the response becoming insignificantly different from zero after about two and a half years. When unemployment is relatively high, the impact response of consumption is significantly positive, but it wears off quickly. The estimated response is briefly significantly negative in the medium term, before rising in the long run (after about four years), but at this horizon, it is imprecisely measured. Consumption multiplier statistics are about -0.41 after two years and -0.22 after four years when spending hits an economy with relatively low unemployment, and about -0.10 after two years and -0.20 after four when the unemployment rate is fairly high when spending shocks the economy. The differences between good- and bad-state consumption multipliers are insignificant at both horizons. Thus, in contrast to much of the empirical literature, I find significantly negative government spending multipliers on private consumption in the medium term no matter the state of the economy. This finding actually accords with the hypothesized negative wealth effects of government spending put forth by a standard neoclassical model.

The situation is different for private investment and for durables consumption, however. In periods marked by both relatively low and relatively high unemployment, private investment responds positively to a government spending shock (at least in the short term), although it response has a much greater magnitude when unemployment is high. In relatively good times, the medium- and long-run responses are a bit choppy, whereas they are relatively steady in bad times. In the long run, the reaction of investment to a government spending shock at a time of relatively high unemployment is significantly positive at the ten percent level. A similar pattern can be seen in the response of durable goods consumption. It is significantly positive on impact in times of fairly low unemployment, though it turns negative within a couple of years. In times of high unemployment, durable goods consumption rises fairly steadily, until its response is significantly positive in the long run. When I compute government spending multipliers on investment, they are insignificantly different from zero at both the two- and four-year horizons when unemployment is relatively low, but the high unemployment investment multiplier is 0.60 after two years and significant. The multipliers on durables consumption are significantly negative at both two and four years in times of low unemployment, but positive (significantly so at four years) in times of high unemployment.

Although a positive investment response to government spending is not a uniform prediction, it could make sense in the context of a negative wealth effect, which lowers private consumption and raises labor supply, thus increasing the marginal productivity of capital and returns on investment. Considering that investment and durables consumption are more likely to be responsive to interest rate fluctuations, these results also support the notion that real interest rates do not rise as much in response to a government spending shock when unemployment is higher (that is, when there is greater spare capacity), although one cannot tell from these results whether that is a deliberate response of the monetary authorities or not. The main point is that the general equilibrium effects that tend to dampen the multiplier in standard neoclassical models seem to be mitigated when unemployment is relatively high.

3.5.2 The Response to Tax Changes

Romer and Romer (2010) introduce a narrative series of tax changes that is similar in spirit to the military spending news variable introduced by Ramey (2011b). By combing through various forms of the legislative record, they identify all legislated changes in the tax code since World War II. Their sample extends to 2007. Importantly, the authors identify the motivation behind each tax change and classify them into four groups, two of which they call endogenous (such as efforts to finance a spending change or to conduct countercyclical policy) and two of which they call exogenous (efforts to deal with an inherited budget deficit or to boost long run growth). Their analysis suggests that an exogenous increase in tax revenues of one percent lowers output by three percent over three years, a highly contractionary effect. They, however, look at the effect of tax changes on the economy averaged over all states. In this subsection, I will see whether their identified exogenous tax changes have differential effects according to the state of the economy in which they were enacted. [20]

In their paper, Romer and Romer (2010) specify their tax changes in three different ways, including and excluding retroactive changes and in present value terms. I will estimate multipliers using all three specifications. The equations that I estimate are very similar to the equations estimated using the military spending news shock,

[20] I am indebted to Joshua Hausman and Galina Hale, who each suggested this line of inquiry to me.

and are expressed as

$$\frac{(y_{t+h} - y_{t-1})}{y_{t-1}} = I_{t-1} * (\alpha_{y,good,h} + \beta_{y,good,h} TaxShock_t + X'_{t-1}\theta_{y,good,h})$$
$$+ (1 - I_{t-1}) * (\alpha_{y,bad,h} + \beta_{y,bad,h} TaxShock_t + X'_{t-1}\theta_{y,bad,h}) + \varepsilon_{y,t+h} \tag{3.12}$$

$$\frac{(T_{t+h} - T_{t-1})}{y_{t-1}} = I_{t-1} * (\alpha_{g,good,h} + \beta_{g,good,h} TaxShock_t + X'_{t-1}\theta_{g,good,h})$$
$$+ (1 - I_{t-1}) * (\alpha_{g,bad,h} + \beta_{g,bad,h} TaxShock_t + X'_{t-1}\theta_{g,bad,h}) + \varepsilon_{g,t+h}, \tag{3.13}$$

where $TaxShock_t$ is the exogenous legislated change in taxes identified by Romer and Romer (2010) in any of their three specifications, and T_t is real tax receipts per capita. The tax shock is scaled as a percentage of GDP, and I can construct tax multipliers in the same way as I construct government spending multipliers. The results can be found in Table 3.14 and Figure 3.14, Figure 3.15, and Figure 3.16.

The most precise results come from the specification involving the tax shock without retroactive tax changes, although the patterns are generally similar across all three specifications. A look at Figure 3.14 implies that tax changes are broadly neutral when the economy is in a state of low unemployment. Output declines initially, but over the longer term, its response is insignificant. Tax receipts themselves do not respond positively to the tax change. This gives tax multipliers insignificantly different from zero when unemployment is relatively low. When unemployment is relatively high, however, one sees a significant decline in output in response to an exogenous tax increase, especially in the long term. The multiplier on taxes is -2.23 at two years and -3.63 at four years when the tax change is imposed in a time of relatively high unemployment. This suggests that the dramatic contractionary response found by Romer and Romer (2010) is driven by the economy's reaction to tax changes in a state of relatively high unemployment. Tax changes are fairly benign for the economy when unemployment is relatively low.

3.6 Conclusion

In this paper, I have revisited the question of whether government purchases of goods and services are more effective in raising output (as evaluated by the government spending multiplier) when the economy is weak relative to when it is strong. Extending the framework developed by Owyang, Ramey, and Zubairy (2013) and Ramey and Zubairy (2014), I use the Hansen (2000) sample splitting technique to endogenously estimate the threshold level of the unemployment rate in postwar U.S. data. In addition, I change the estimating equation to include stochastic trends, as opposed to deterministic trends, and control for the monetary policy stance. In so doing, I find evidence for larger government spending multipliers in times of economic weakness, unlike Owyang, Ramey, and Zubairy (2013) and Ramey and Zubairy (2014), on the order of 1.6 to 1.9, which are significantly higher than one in most specifications and significantly higher than the associated multiplier in times of economic strength. Unlike Auerbach and Gorodnichenko (2012b) and Auerbach and Gorodnichenko (2012a), the multipliers in good times are positive, if small.

In addition, I find that consumption has a small, possibly negative response to government spending no matter the state of the economy, but that investment has a significantly positive multiplier on government spending when it is undertaken in a time of high unemployment. I also find a significantly negative tax multiplier in times of high slack, but an insignificant multiplier when the economy is operating closer to capacity.

It may be interesting in the future to further explore the possibility of more than two states relevant to the fiscal multiplier calculation. This notion is theoretically justified in the work of Sims and Wolff (2013), who construct a model in which output multipliers vary not only according to whether there is slack in the economy or not, but also according to the source of the slack. In their model, recessions caused by aggregate demand shock experience relatively small output multipliers, while those caused by aggregate supply shocks experience larger output multipliers. The opposite is true of what they call "welfare" multipliers, a metric that attempts to take into account whether economic agents are actually any better off as a result of the increase in government spending. That is, it tries to take into account the utility that agents extract from public spending or the disutility they might get from working more.[21] The possibility of a third regime to consider when thinking about fiscal multipliers is also hinted at in Fazzari, Morley, and Panovska (2013) and Hausman (2013). The sample size difficulties engendered by splitting the time series into a larger number of states do pose a barrier, however. One may also want to think about possible interactions with income inequality or government debt. A final future avenue that may be pursued is the possible extent to which the military news variable considered here (and arguably the best identified shock to spending in the

[21]The relative lack of attention to the effects on utility or welfare of increased government expenditures is also noted by Mankiw and Weinzierl (2011).

literature) might suffer from allocation bias, of the type considered in Angrist, Jordà, and Kuersteiner (2013) and Jordà and Taylor (2013). That is, are some periods in history more susceptible to receiving a military spending news shock? I leave all of this to future research.

Table 3.1: Theoretical Multipliers when the ZLB Binds

	Fiscal Multiplier
Christiano, Eichenbaum, and Rebelo (2011)	3.7
Eggertsson (2010)	2.3
Nakamura and Steinsson (2014)	1.7

These figures report government spending multipliers in benchmark versions of each paper's model. In the case of Nakamura and Steinsson (2014), the figure reported is from their model in the working paper where household preferences are separable in consumption and leisure. With nonseparable preferences, the model gives a multiplier of 8.73.

Table 3.2: Explanatory Power of Military Spending Shock for Future Government Spending

State	Horizon	F-Stat
Good	2 years	13.75
Bad	2 years	4.51
Good	4 years	64.44
Bad	4 years	15.34

The table gives F-Statistics for the null hypothesis that the coefficients on the military spending news shock of Ramey (2011b) and Owyang, Ramey, and Zubairy (2013) are jointly zero for government spending growth in the state of the economy and over the horizon indicated. Good states and bad states are determined by an unemployment rate of 5.71, which is the threshold level endogenously estimated in Section 3.4.2.

Table 3.3: Fiscal Multipliers: ORZ,RZ Specifications

Sample: 1948 to 2010	Horizon	Multiplier	Confidence Interval
Low Unemployment	2 years	0.176	[-0.224,0.575]
High Unemployment	2 years	-2.11	[-3.977, -0.244]
Low Unemployment	4 years	0.619	[0.183, 1.056]
High Unemployment	4 years	33.505	[-141.317, 208.327]
Difference	2 years	-2.289	[-5.022, 0.451]
Difference	4 years	32.886	[30.352, 35.420]

The table gives integral multipliers, calculated as $(\sum_{h=1}^{H} \beta_{y,j,h})/(\sum_{h=1}^{H} \beta_{g,j,h})$ for $j = good, bad$ and $H = 8, 16$. Confidence intervals are for the 90 percent significance level, calculated via the delta method for nonlinear combinations of coefficients. The parameter estimates come from the least squares regression of Equation 3.3, with Newey-West standard errors. The Newey-West lag parameter is 20.

Table 3.4: DF-GLS Tests of Main Time Series Variables

	Real GDP	Real Government Spending	Real Tax Revenues
DF-GLS Stat	-2.555	-1.223	-2.531
Lag Length	2	8	8

The table gives DF-GLS test statistics for the variables indicated in each column. Lag length is selected by according to Ng and Perron (1995). The maximum number of lags is 8. The null hypothesis for each regression is that the series has an autoregressive unit root, while the alternative hypothesis is trend stationarity. ***, **, and * indicate rejection of the null hypothesis at the 1, 5, and 10 percent levels.

Table 3.5: Fiscal Multipliers: Stochastic Trend Specifications

Sample: 1948 to 2010			
Low Unemployment	2 years	0.414	[0.236,0.592]
High Unemployment	2 years	0.224	[-0.718, 1.165]
Low Unemployment	4 years	0.541	[0.431, 0.652]
High Unemployment	4 years	1.392	[0.741, 2.043]
Difference	2 years	-0.190	[-1.873, 1.493]
Difference	4 years	0.851	[-0.020, 1.721]

The table gives integral multipliers, calculated as $(\sum_{h=1}^{H} \beta_{y,j,h})/(\sum_{h=1}^{H} \beta_{g,j,h})$ for $j = good, bad$ and $H = 8, 16$. Confidence intervals are for the 90 percent significance level, calculated via the delta method for nonlinear combinations of coefficients. Parameter estimates are from the least squares regression of Equation 3.3, with Newey-West standard errors. The Newey-West lag parameter is 20.

Table 3.6: Fiscal Multipliers: Expanded Control Set

Sample: 1948 to 2010			
Low Unemployment	2 years	0.456	[0.231,0.682]
High Unemployment	2 years	3.572	[2.480, 4.663]
Low Unemployment	4 years	0.488	[0.344, 0.632]
High Unemployment	4 years	3.094	[2.383, 3.804]
Difference	2 years	3.115	[1.566, 4.665]
Difference	4 years	2.606	[1.925, 3.286]

The table gives integral multipliers, calculated as $(\sum_{h=1}^{H} \beta_{y,j,h})/(\sum_{h=1}^{H} \beta_{g,j,h})$ for $j = good, bad$ and $H = 8, 16$. Confidence intervals are for the 90 percent significance level, calculated via the delta method for nonlinear combinations of coefficients. Parameter estimates are from the least squares regression of Equation 3.3, with Newey-West standard errors. The Newey-West lag parameter is 20.

Table 3.7: Fiscal Multipliers: Parsimonious Specification

Sample: 1948 to 2010			
Low Unemployment	2 years	0.528	[0.366,0.689]
High Unemployment	2 years	1.278	[0.254, 2.301]
Low Unemployment	4 years	0.562	[0.448, 0.677]
High Unemployment	4 years	2.284	[1.335, 3.233]
Difference	2 years	0.750	[-1.334, 2.834]
Difference	4 years	1.721	[0.605, 2.837]

The table gives integral multipliers, calculated as $(\sum_{h=1}^{H} \beta_{y,j,h})/(\sum_{h=1}^{H} \beta_{g,j,h})$ for $j = good, bad$ and $H = 8, 16$. Confidence intervals are for the 90 percent significance level, calculated via the delta method for nonlinear combinations of coefficients. Parameter estimates are from the least squares regression of Equation 3.3, with Newey-West standard errors. The Newey-West lag parameter is 20.

Table 3.8: Bootstrapped Asymptotic P-Values for Null of No Threshold Behavior

Horizon	Spec. (1)	Spec. (2),(3)
1	0.296	0.078
2	0.149	0.001
3	0.127	0.000
4	0.065	0.001
5	0.023	0.000
6	0.007	0.000
7	0.009	0.000
8	0.010	0.000
9	0.006	0.000
10	0.008	0.000
11	0.028	0.000
12	0.016	0.000
13	0.003	0.000
14	0.000	0.000
15	0.000	0.000
16	0.002	0.000
17	0.015	0.000
18	0.048	0.000
19	0.067	0.000
20	0.140	0.000

The table gives P-Values for a test of the null that there is no threshold in a linear regression under homoskedasticity, as in Hansen (1996). Specification (1) tests for a threshold in an equation system as in Equations 3.10 and 3.11 with all independent variables included in the x_i vector. Specifications (2) and (3) include only the military spending news variable in the x_i vector.

Table 3.9: Least Squares Estimate of Threshold Unemployment Rate

h (quarters)	Spec. (1)	Spec. (2),(3)	h (quarters)	Spec. (1)	Spec. (2),(3)
1	5.76	5.71	11	5.03	5.16
	[5.71, 6.62]	[4.74, 7.03]		[4.67, 5.27]	[4.96, 6.12]
2	5.71	5.71	12	5.03	5.16
	[5.47, 6.03]	[5.04, 6.15]		[4.67, 5.16]	[4.97, 6.38]
3	5.48	5.71	13	5.03	5.88
	[4.97, 6.03]	[4.99, 6.15]		[4.96, 5.16]	[5.13, 6.38]
4	5.49	5.16	14	5.03	5.88
	[3.95, 5.94]	[4.96, 6.15]		[4.97, 5.26]	[5.16, 6.38]
5	5.49	5.16	15	5.11	5.88
	[4.97, 5.62]	[4.97, 5.71]		[4.97, 5.26]	[5.16, 6.44]
6	5.16	5.16	16	5.11	5.88
	[4.97, 5.29]	[4.97, 5.48]		[4.97, 5.26]	[5.76, 6.50]
7	5.16	5.16	17	5.12	5.88
	[5.13, 5.29]	[4.96, 5.35]		[4.97, 5.16]	[5.76, 6.50]
8	5.16	5.16	18	5.03	5.88
	[5.13, 5.27]	[4.97, 5.35]		[4.96, 5.26]	[5.76, 6.50]
9	5.16	5.16	19	5.11	5.88
	[5.13, 5.29]	[5.00, 5.35]		[4.96, 5.26]	[5.76, 6.50]
10	5.16	5.16	20	5.11	5.88
	[4.96, 5.29]	[4.97, 5.35]		[4.96, 5.26]	[5.76, 6.44]

The table gives least squares estimates of γ in equation system given by Equations 3.10 and 3.11, estimated using Hansen (2000) procedure under homoskedasticity. Results allowing heteroskedasticity were almost identical. 95% confidence regions for the threshold level are in brackets. Details on the three specifications are found in the text.

Table 3.10: Fiscal Multipliers: Endogenously Estimated Unemployment Thresholds

		Spec. (1)	Spec. (2)	Spec. (3)
Low Unemployment	2 years	0.281	0.557	0.274
		[-0.573, 1.135]	[0.380, 0.733]	[-0.024, 0.573]
High Unemployment	2 years	0.201	1.605	1.808
		[0.093, 0.308]	[1.065, 2.146]	[1.413, 2.202]
Low Unemployment	4 years	0.791	0.587	0.427
		[0.234, 1.349]	[0.465, 0.710]	[0.251, 0.603]
High Unemployment	4 years	0.350	1.678	1.918
		[0.263, 0.438]	[1.301, 2.055]	[1.603, 2.232]
Difference	2 years	-0.081	1.049	1.534
		[-0.928, 0.767]	[-0.740, 2.837]	[0.223, 2.844]
Difference	4 years	-0.441	1.091	1.491
		[-1.023, 0.141]	[0.174, 2.008]	[0.769, 2.213]

The table gives integral multipliers, calculated as $(\sum_{h=1}^{H} \beta_{y,j,h})/(\sum_{h=1}^{H} \beta_{g,j,h})$ for $j = good, bad$ and $H = 8, 16$. Confidence intervals (in brackets) are for the 90 percent significance level, calculated via the delta method for nonlinear combinations of coefficients. Parameter estimates are from the least squares regression of Equation 3.3, with Newey-West standard errors. The Newey-West lag parameter is 20. Specifications (1), (2), and (3) are detailed in the text.

Table 3.11: Fiscal Multipliers: Endogenously Estimated Unemployment Thresholds

		Spec. (1)	Spec. (2)	Spec. (3)
Low Unemployment	2 years	0.281	0.557	0.274
		[-1.172, 0.510]	[0.268, 0.845]	[-1.088, 0.510]
High Unemployment	2 years	0.201	1.605	1.808
		[-5.596, 3.572]	[-5.739, 93.150]	[-130.407, 99.344]
Low Unemployment	4 years	0.791	0.587	0.427
		[-1.087, 0.959]	[0.035, 0.710]	[-0.922, 0.959]
High Unemployment	4 years	0.350	1.678	1.918
		[-0.036, 4.424]	[-14.590, 2.884]	[-0.976, 4.424]
Difference	2 years	-0.081	1.049	1.534
		[-6.106, 3.115]	[-6.273, 92.641]	[-130.833, 98.918]
Difference	4 years	-0.441	1.091	1.491
		[-0.630, 3.914]	[-15.137, 2.299]	[-1.493, 3.914]

The table gives integral multipliers, calculated as $(\sum_{h=1}^{H} \beta_{y,j,h})/(\sum_{h=1}^{H} \beta_{g,j,h})$ for $j = good, bad$ and $H = 8, 16$. Confidence intervals (in brackets) are Bonferroni bounds taking into account uncertainty in the true threshold unemployment rate, constructed using the 95% confidence intervals on the OLS estimate of the threshold. Parameter estimates are from the least squares regression of Equation 3.3, with Newey-West standard errors. The Newey-West lag parameter is 20. Specifications (1), (2), and (3) are detailed in the text.

Table 3.12: Comparison of State Dependent Multiplier Estimates

Paper	Horizon	Good State Multiplier	Bad State Multiplier
Current Study	2 years	0.274	1.808
Current Study	4 years	0.427	1.918
Ramey and Zubairy (2014)	2 years	0.79	0.69
Ramey and Zubairy (2014)	4 years	0.96	0.76
Auerbach and Gorodnichenko (2012b)	5 years	-0.33	2.24
Auerbach and Gorodnichenko (2012a)	5 years	-0.20	0.46

The table gives the preferred multiplier estimates at the horizons indicated for the given papers. The multipliers from this paper are those estimated from Specification (3) with endogenously estimated threshold unemployment, stochastic trends, and monetary policy. See Table 3.10. The multipliers for Ramey and Zubairy (2014) come from their Table 2. The multipliers for Auerbach and Gorodnichenko (2012b) come from their Table 1. The multipliers for Auerbach and Gorodnichenko (2012a) come from their Table 1.

Table 3.13: Government Spending Multipliers on Consumption and Investment

		Nondurables and Services	Investment	Durables
Low Unemployment	2 years	-0.411	-0.165	-0.162
		[-0.519, -0.302]	[-0.347, 0.017]	[-0.234, -0.090]
High Unemployment	2 years	-0.103	0.595	0.144
		[-0.270, 0.064]	[0.167, 1.023]	[-0.050, 0.337]
Low Unemployment	4 years	-0.216	-0.125	-0.116
		[-0.283, -0.148]	[-0.214, -0.036]	[-0.161, -0.071]
High Unemployment	4 years	-0.209	0.183	0.269
		[-0.375, -0.043]	[-0.194, 0.559]	[0.074, 0.465]
Difference	2 years	0.308	0.760	0.305
		[-0.298, 0.912]	[-0.635, 2.155]	[-0.403, 1.014]
Difference	4 years	0.006	0.308	0.385
		[-0.401, 0.414]	[-0.633, 1.248]	[-0.176, 0.946]

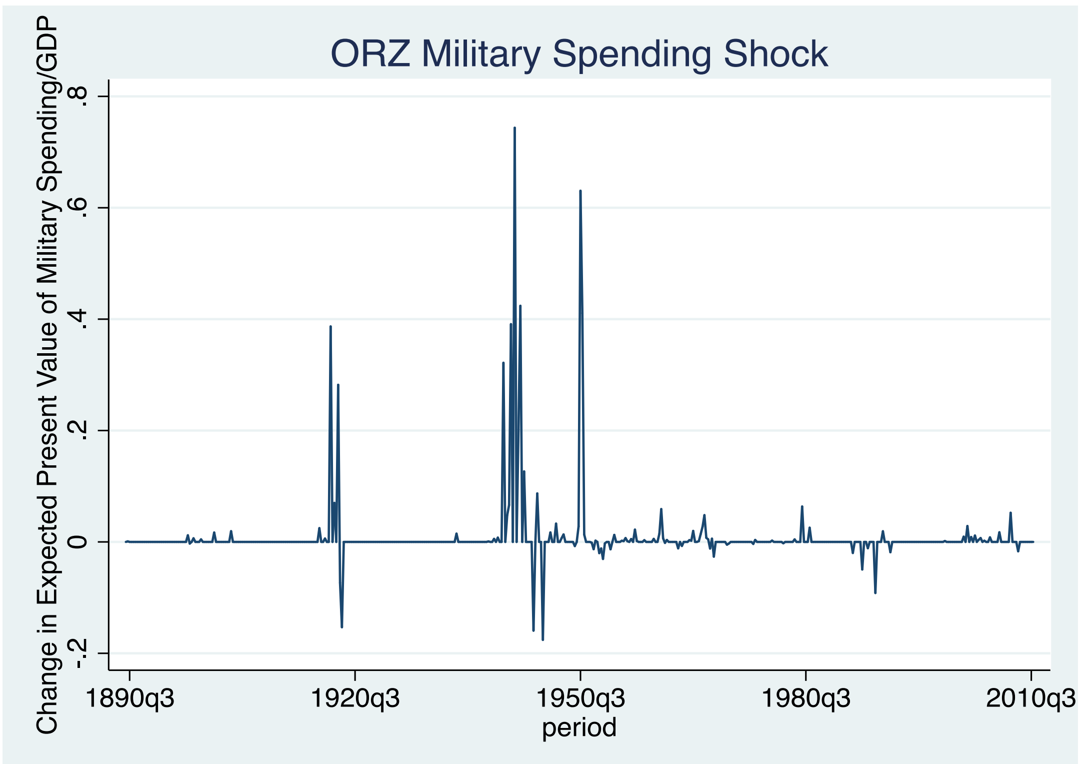
The table gives integral multipliers, calculated as $(\sum_{h=1}^{H} \beta_{y,j,h})/(\sum_{h=1}^{H} \beta_{g,j,h})$ for $j = good, bad$ and $H = 8, 16$. Confidence intervals (in brackets) are for the 90 percent significance level, calculated via the delta method for nonlinear combinations of coefficients. Parameter estimates are from the least squares regression of Equation 3.3, with Newey-West standard errors. The dependent variables now are the change in real personal consumption expenditures on nondurables and services per capita (Column 1), the change in real gross private domestic investment per capita (Column 2), and the change in real personal consumption expenditures on durable goods (Column 3). The regression equations are as expressed in Equation 3.1 and Equation 3.2 The Newey-West lag parameter is 20.

Table 3.14: Tax Multipliers: Romer and Romer (2010) Narrative Tax Series

Sample: 1948 to 2007		Tax Shock (1)	Tax Shock (2)	Tax Shock (3)
Low Unemployment	2 years	0.506	0.730	1.232
		[-0.054, 1.065]	[0.425, 1.036]	[0.963, 1.501]
High Unemployment	2 years	-2.229	-6.028	-11.603
		[-3.591, -0.866]	[-10.215, -1.841]	[-31.866, 8.661]
Low Unemployment	4 years	0.028	0.737	1.215
		[-0.353, 0.410]	[0.568, 0.907]	[0.991, 1.439]
High Unemployment	4 years	-3.628	-18.161	14.667
		[-5.667, -1.590]	[-42.865, 6.544]	[0.458, 28.877]
Difference	2 years	-2.734	-6.758	-12.834
		[-3.307, -2.162]	[-7.088, -6.428]	[-13.095, -12.574]
Difference	4 years	-3.657	-18.898	13.452
		[-4.058, -3.255]	[-19.086, -18.709]	[13.214, 13.691]

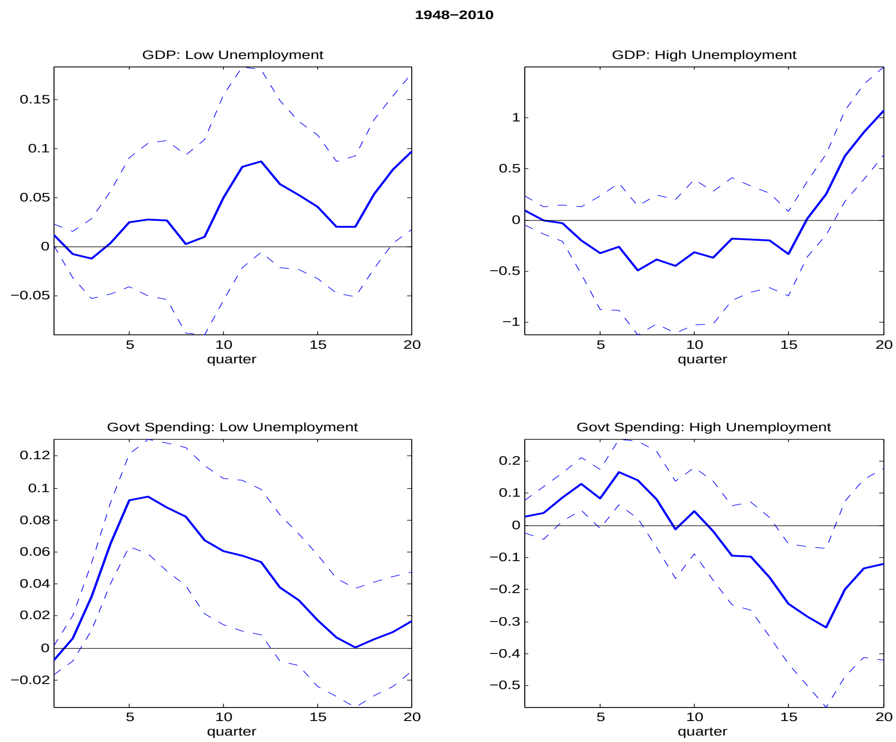
The table gives integral multipliers, calculated as $(\sum_{h=1}^{H} \beta_{y,j,h})/(\sum_{h=1}^{H} \beta_{g,j,h})$ for $j = good, bad$ and $H = 8, 16$. Confidence intervals (in brackets) are for the 90 percent significance level, calculated via the delta method for nonlinear combinations of coefficients. Parameter estimates are from the least squares regression of Equation 3.3, with Newey-West standard errors. The Newey-West lag parameter is 20. Tax Shock (1) refers to the tax shock of Romer and Romer (2010) excluding retroactive changes. Tax Shock (2) refers to the tax shock of Romer and Romer (2010) including retroactive changes. Tax Shock (3) refers to the tax shock of Romer and Romer (2010) in present discounted value terms.

Figure 3.1: An Exogenous Shock to Government Spending



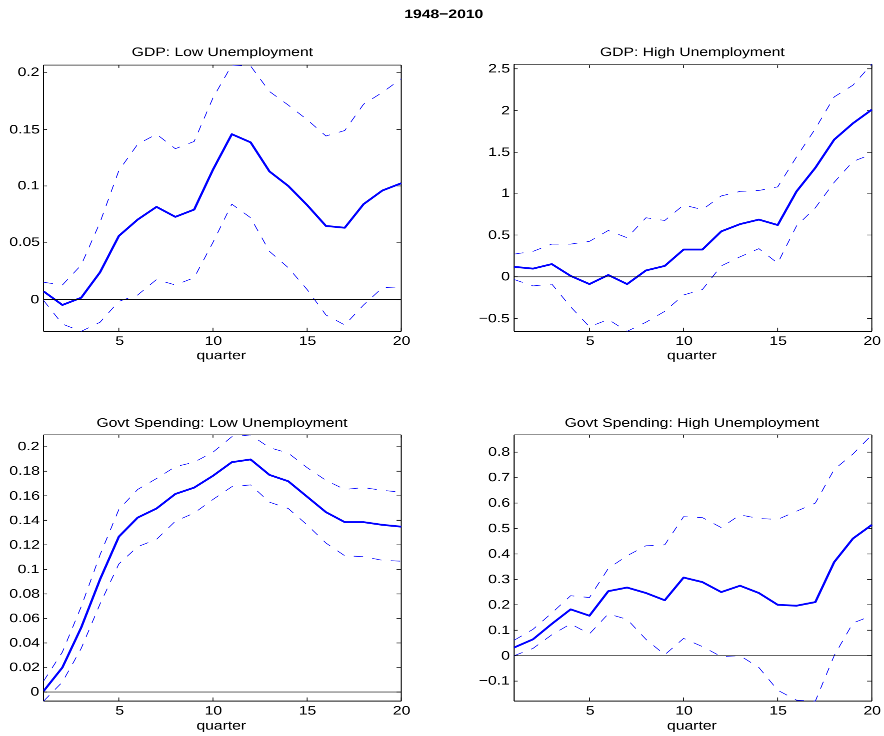
The figure gives the change in expected present value of military spending as a share of lagged nominal gross domestic product, as constructed first in Ramey (2011b) and then extended in Owyang, Ramey, and Zubairy (2013).

Figure 3.2: Impulse Response Functions, ORZ Specification



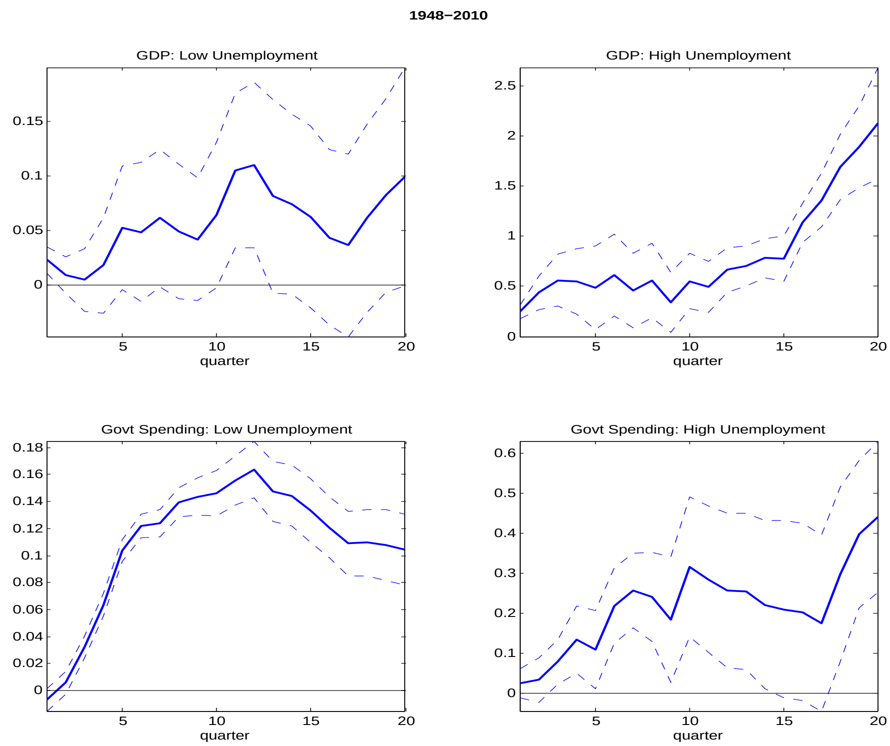
The figure shows impulse response functions for real output and real government spending per capita in response to a military spending news shock of Ramey (2011b) and Owyang, Ramey, and Zubairy (2013) equal to one percent of GDP. The sample is 1948-2010, and the specification is nearly identical to that employed in Ramey and Zubairy (2014). The dashed lines represent 90 percent confidence intervals. The number of observations in the low-unemployment state are 187 when $h = 1$ and 176 when $h = 20$. In the high-unemployment state, they are 61 when $h = 1$ and 53 when $h = 20$. The number of observations decline due to the particular specification of the dependent variable, which necessitates dropping one additional observation as the horizon increases.

Figure 3.3: Impulse Response Functions, Stochastic Trends



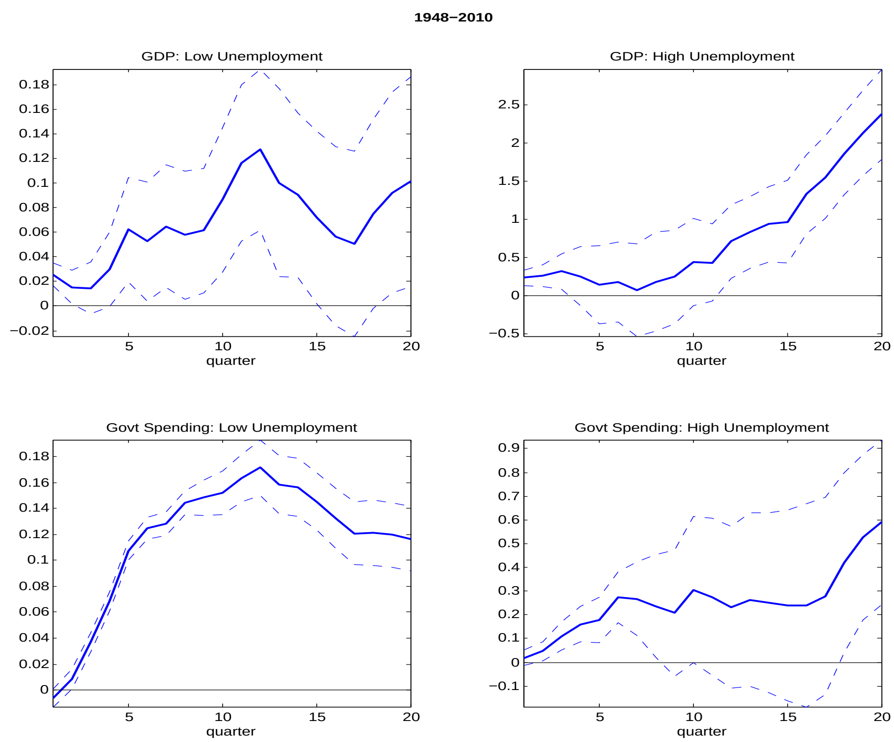
The figure shows impulse response functions for real output and real government spending per capita in response to a military spending news shock of Ramey (2011b) and Owyang, Ramey, and Zubairy (2013) equal to one percent of GDP. The sample is 1948-2010, and the specification allows stochastic rather than deterministic trends. The dashed lines represent 90 percent confidence intervals. The number of observations in the low-unemployment state are 187 when $h = 1$ and 176 when $h = 20$. In the high-unemployment state, they are 61 when $h = 1$ and 53 when $h = 20$. The number of observations decline due to the particular specification of the dependent variable, which necessitates dropping one additional observation as the horizon increases.

Figure 3.4: Impulse Response Functions, Expanded Control Set



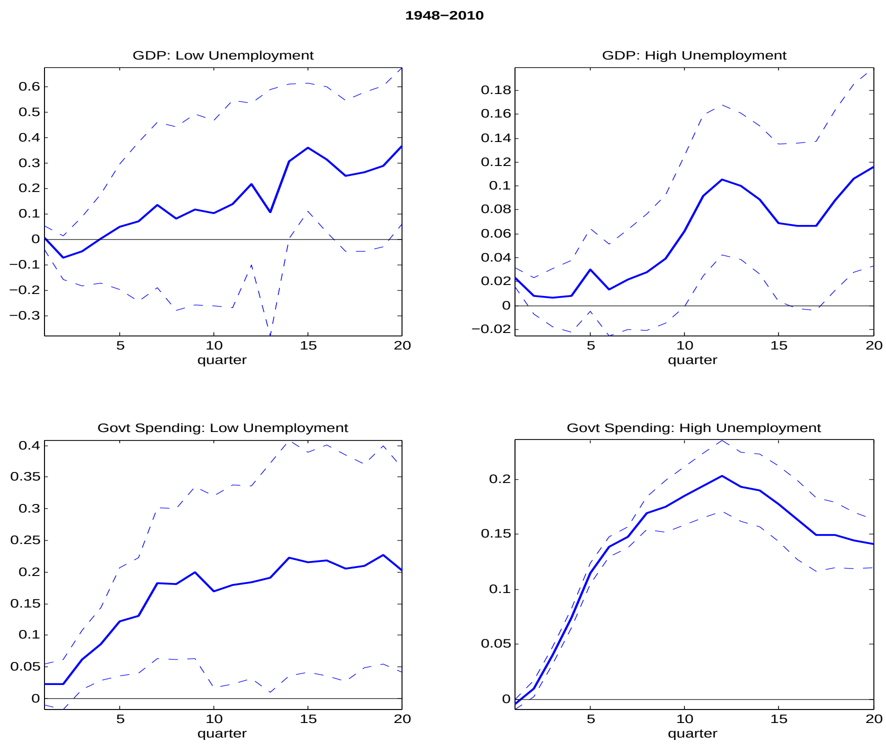
The figure shows impulse response functions for real output and real government spending per capita in response to a military spending news shock of Ramey (2011b) and Owyang, Ramey, and Zubairy (2013) equal to one percent of GDP. The sample is 1948-2010, and the specification allows stochastic rather than deterministic trends. It augments the control set with four lags of the military spending news variable and the three month T-Bill rate. The dashed lines represent 90 percent confidence intervals. The number of observations in the low-unemployment state are 187 when $h = 1$ and 176 when $h = 20$. In the high-unemployment state, they are 61 when $h = 1$ and 53 when $h = 20$. The number of observations decline due to the particular specification of the dependent variable, which necessitates dropping one additional observation as the horizon increases.

Figure 3.5: Impulse Response Functions, Parsimonious Specification



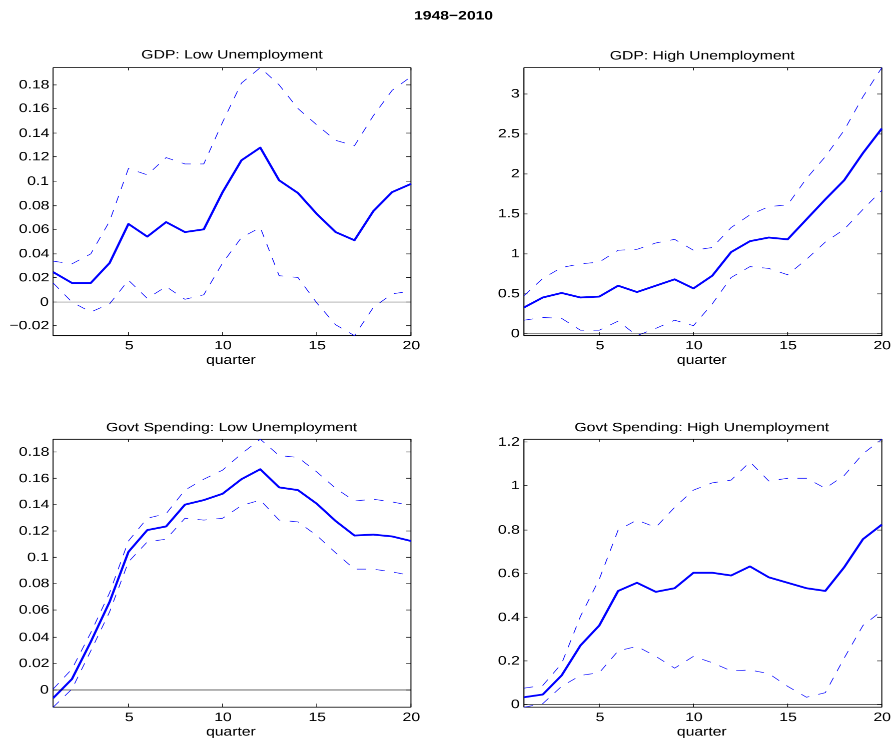
The figure shows impulse response functions for real output and real government spending per capita in response to a military spending news shock of Ramey (2011b) and Owyang, Ramey, and Zubairy (2013) equal to one percent of GDP. The sample is 1948-2010, and the specification allows stochastic rather than deterministic trends. It augments the control set with four lags of the military spending news variable and the three month T-Bill rate. Only the constant and the coefficient on the military news variable switch between regimes. The dashed lines represent 90 percent confidence intervals. The number of observations in the low-unemployment state are 187 when $h = 1$ and 176 when $h = 20$. In the high-unemployment state, they are 61 when $h = 1$ and 53 when $h = 20$. The number of observations decline due to the particular specification of the dependent variable, which necessitates dropping one additional observation as the horizon increases.

Figure 3.6: Impulse Response Functions, Endogenously Estimated Threshold: Specification (1)



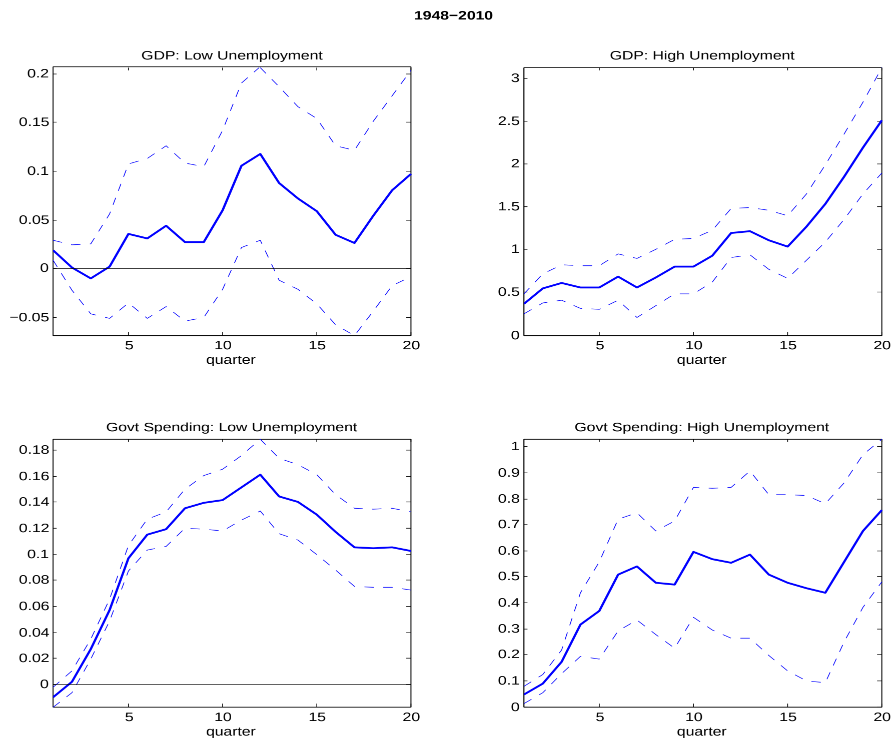
The figure shows impulse response functions for real output and real government spending per capita in response to a military spending news shock of Ramey (2011b) and Owyang, Ramey, and Zubairy (2013) equal to one percent of GDP. The sample is 1948-2010, and the specification allows stochastic rather than deterministic trends. It augments the control set with four lags of the military spending news variable and the three month T-Bill rate. The dashed lines represent 90 percent confidence intervals. Specification (1) is detailed in the text. The threshold level of the unemployment rate is endogenously estimated at 5.14. The number of observations in the low-unemployment state are 105 when $h = 1$ and 96 when $h = 20$. In the high-unemployment state, they are 143 when $h = 1$ and 133 when $h = 20$. The number of observations decline due to the particular specification of the dependent variable, which necessitates dropping one additional observation as the horizon increases.

Figure 3.7: Impulse Response Functions, Endogenously Estimated Threshold: Specification (2)



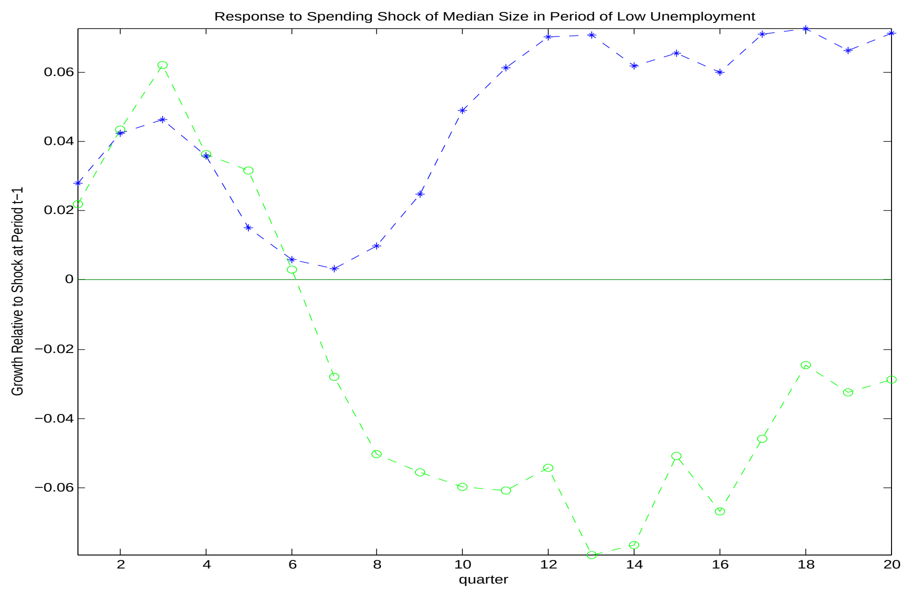
The figure shows impulse response functions for real output and real government spending per capita in response to a military spending news shock of Ramey (2011b) and Owyang, Ramey, and Zubairy (2013) equal to one percent of GDP. The sample is 1948-2010, and the specification allows stochastic rather than deterministic trends. It augments the control set with four lags of the military spending news variable and the three month T-Bill rate. The dashed lines represent 90 percent confidence intervals. Specification (2) is detailed in the text. The threshold level of the unemployment rate is endogenously estimated at 5.71. The number of observations in the low-unemployment state are 157 when $h = 1$ and 147 when $h = 20$. In the high-unemployment state, they are 91 when $h = 1$ and 82 when $h = 20$. The number of observations decline due to the particular specification of the dependent variable, which necessitates dropping one additional observation as the horizon increases.

Figure 3.8: Impulse Response Functions, Endogenously Estimated Threshold: Specification (3)



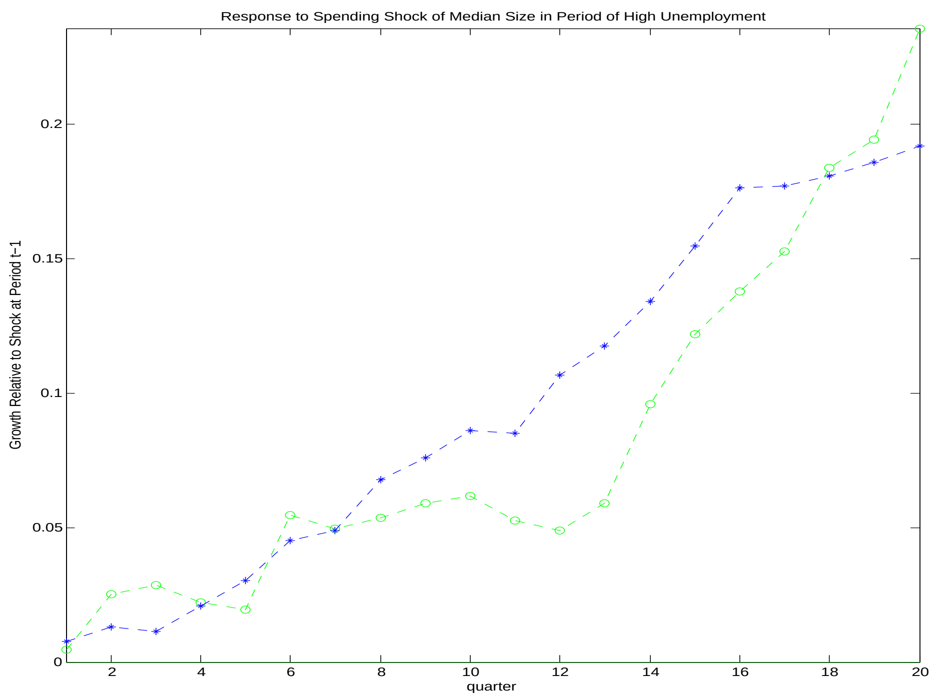
The figure shows impulse response functions for real output and real government spending per capita in response to a military spending news shock of Ramey (2011b) and Owyang, Ramey, and Zubairy (2013) equal to one percent of GDP. The sample is 1948-2010, and the specification allows stochastic rather than deterministic trends. It augments the control set with four lags of the military spending news variable and the three month T-Bill rate. The dashed lines represent 90 percent confidence intervals. Specification (3) is detailed in the text. The threshold level of the unemployment rate is endogenously estimated at 5.71. The number of observations in the low-unemployment state are 157 when $h = 1$ and 147 when $h = 20$. In the high-unemployment state, they are 91 when $h = 1$ and 82 when $h = 20$. The number of observations decline due to the particular specification of the dependent variable, which necessitates dropping one additional observation as the horizon increases.

Figure 3.9: Example for a Military Spending Shock in a Low Unemployment State



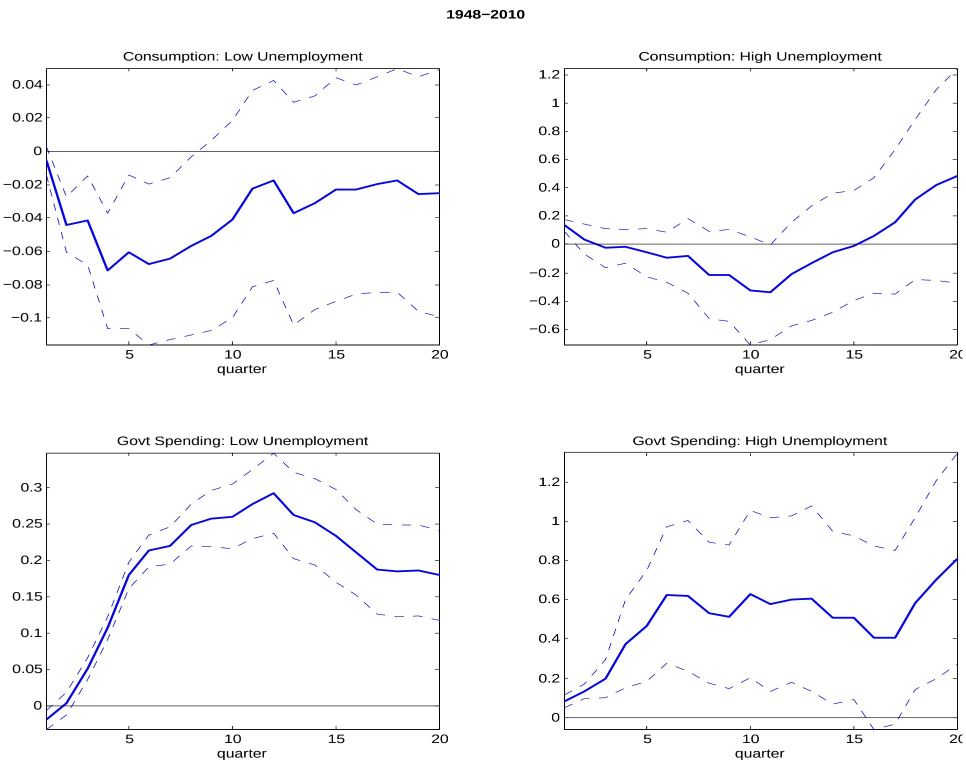
The figure shows the time paths of output (crossed line) and government spending (circled line) after the realization of a military spending news shock of median size when unemployment is relatively low. The magnitude of the shock is 0.23% of GDP.

Figure 3.10: Example for a Military Spending Shock in a High Unemployment State



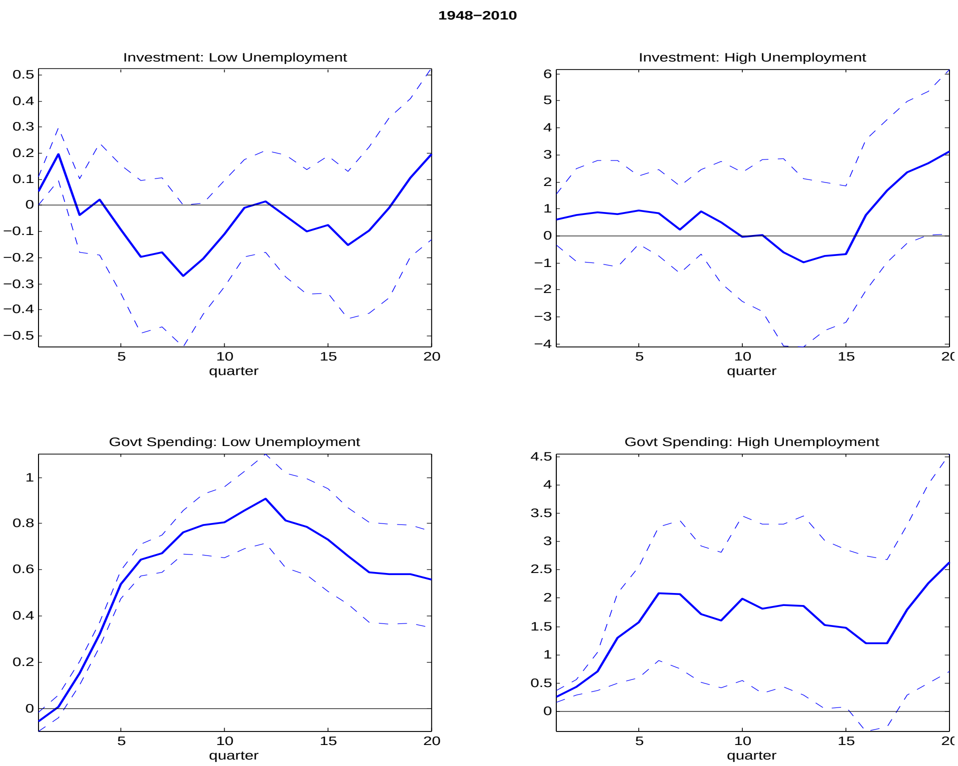
The figure shows the time paths of output (crossed line) and government spending (circled line) after the realization of a military spending news shock of median size when unemployment is relatively high. The magnitude of the shock is 0.36% of GDP.

Figure 3.11: Impulse Response Functions: Consumption



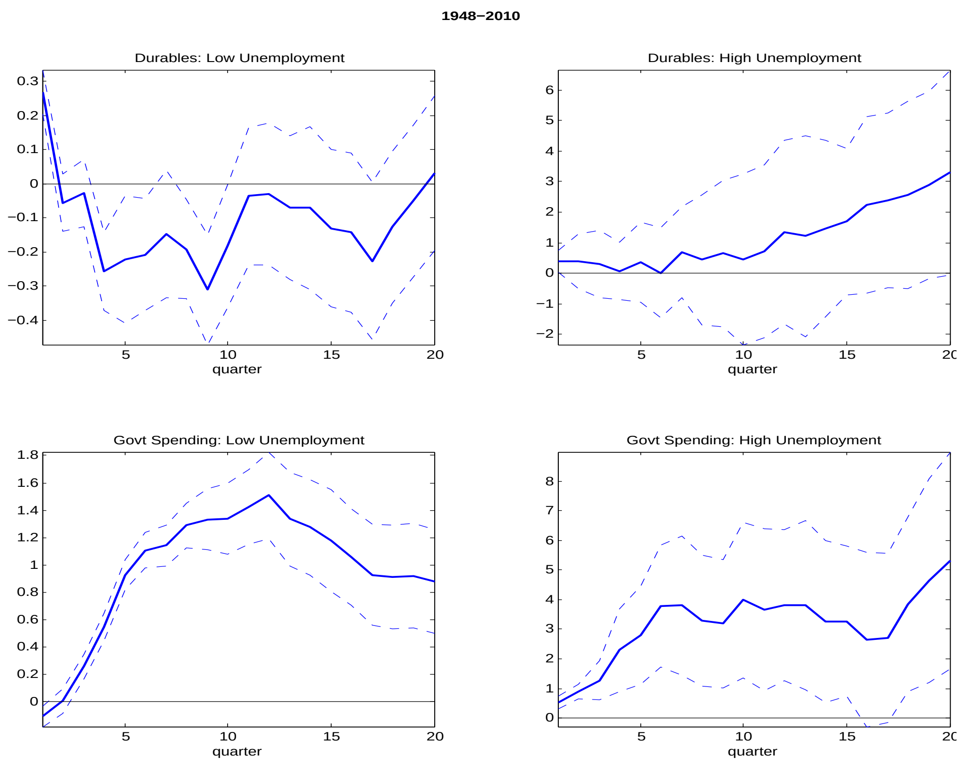
The figure shows impulse response functions for real personal consumption expenditures of nondurables and services and real government spending per capita in response to a military spending news shock of Ramey (2011b) and Owyang, Ramey, and Zubairy (2013) equal to one percent of GDP. The sample is 1948-2010, and the specification allows stochastic rather than deterministic trends. It augments the control set with four lags of the military spending news variable and the three month T-Bill rate. The dashed lines represent 90 percent confidence intervals. The threshold level of the unemployment rate is that estimated for Specification (3) for the output regression.

Figure 3.12: Impulse Response Functions: Investment



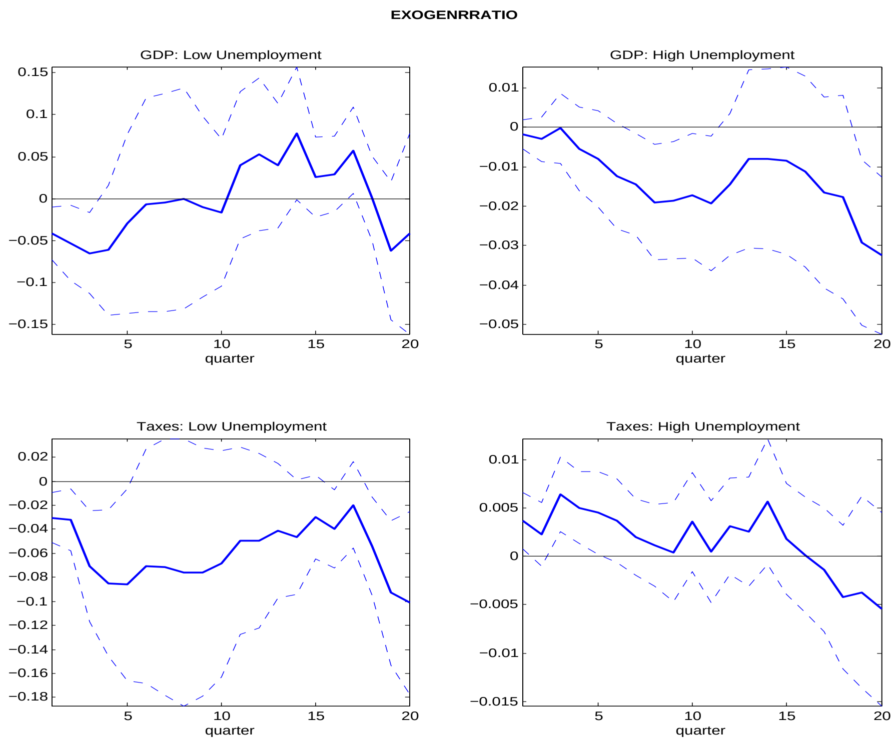
The figure shows impulse response functions for real gross private domestic investment and real government spending per capita in response to a military spending news shock of Ramey (2011b) and Owyang, Ramey, and Zubairy (2013) equal to one percent of GDP. The sample is 1948-2010, and the specification allows stochastic rather than deterministic trends. It augments the control set with four lags of the military spending news variable and the three month T-Bill rate. The dashed lines represent 90 percent confidence intervals. The threshold level of the unemployment rate is that estimated for Specification (3) for the output regression.

Figure 3.13: Impulse Response Functions: Durable Goods Consumption



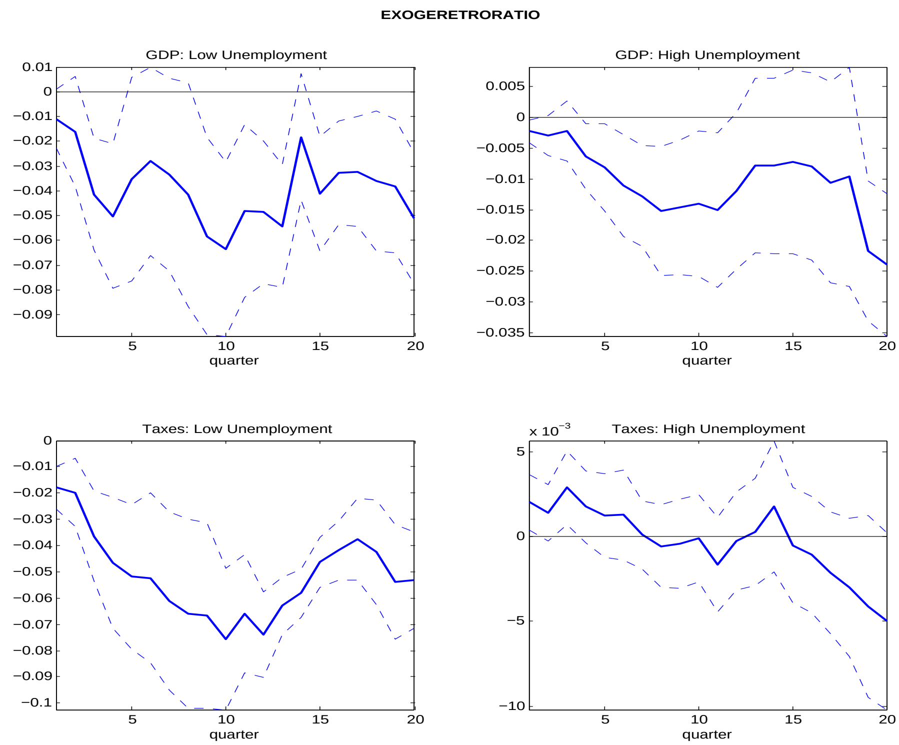
The figure shows impulse response functions for real durable goods consumption and real government spending per capita in response to a military spending news shock of Ramey (2011b) and Owyang, Ramey, and Zubairy (2013) equal to one percent of GDP. The sample is 1948-2010, and the specification allows stochastic rather than deterministic trends. It augments the control set with four lags of the military spending news variable and the three month T-Bill rate. The dashed lines represent 90 percent confidence intervals. The threshold level of the unemployment rate is that estimated for Specification (3) for the output regression.

Figure 3.14: Impulse Response Functions: Romer and Romer (2010) Tax Shock, Excluding Retroactive Changes



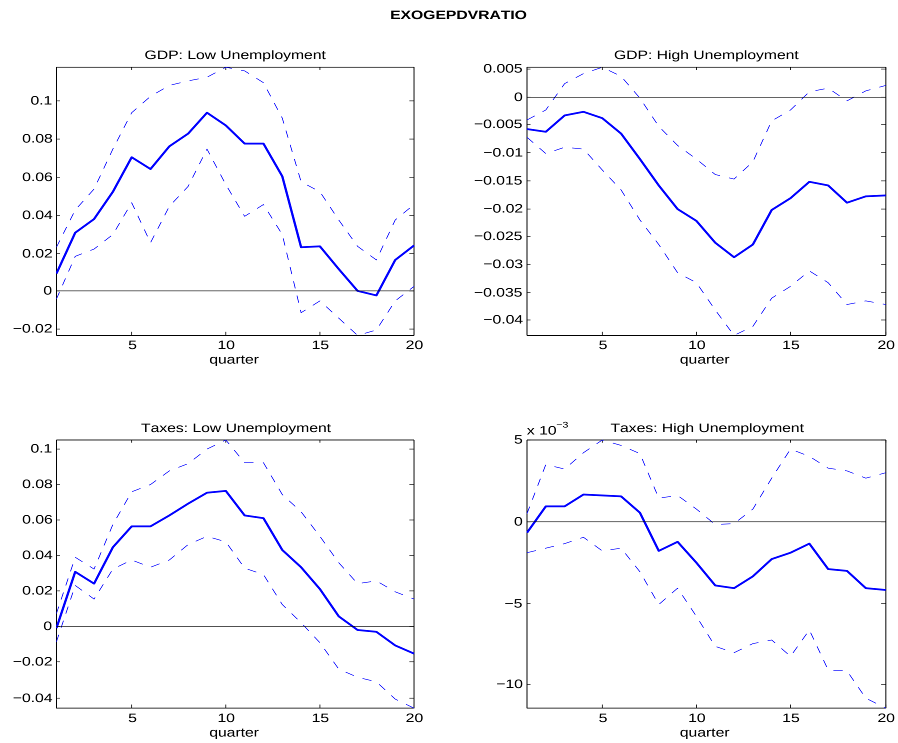
The figure shows impulse response functions for real output and real tax receipts per capita in response to a tax shock of Romer and Romer (2010) (excluding retroactive changes) equal to one percent of GDP. The sample is 1948-2007, and the specification allows stochastic rather than deterministic trends. It augments the control set with four lags of the military spending news variable and the three month T-Bill rate. The dashed lines represent 90 percent confidence intervals. The threshold level of the unemployment rate is endogenously estimated at 4.69, according to Specification (3) discussed in Section 3.4.2.

Figure 3.15: Impulse Response Functions: Romer and Romer (2010) Tax Shock, Including Retroactive Changes



The figure shows impulse response functions for real output and real tax receipts per capita in response to a tax shock of Romer and Romer (2010) (including retroactive changes) equal to one percent of GDP. The sample is 1948-2007, and the specification allows stochastic rather than deterministic trends. It augments the control set with four lags of the military spending news variable and the three month T-Bill rate. The dashed lines represent 90 percent confidence intervals. The threshold level of the unemployment rate is endogenously estimated at 4.46, according to Specification (3) discussed in Section 3.4.2.

Figure 3.16: Impulse Response Functions: Romer and Romer (2010) Tax Shock, Present Value Terms



The figure shows impulse response functions for real output and real tax receipts per capita in response to a tax shock of Romer and Romer (2010) (present value terms) equal to one percent of GDP. The sample is 1948-2007, and the specification allows stochastic rather than deterministic trends. It augments the control set with four lags of the military spending news variable and the three month T-Bill rate. The dashed lines represent 90 percent confidence intervals. The threshold level of the unemployment rate is endogenously estimated at 4.45, according to Specification (3) discussed in Section 3.4.2.

Chapter 4

The Uncertain State-Dependent Government Spending Multiplier

4.1 Introduction

As the economy in the United States has suffered from weak growth in recent years, conventional monetary policy has been unable to provide much in the way of stimulus, with nominal interest rates pinned to the "Zero Lower Bound." This has led to an explosion of research into whether fiscal policy can "pick up the slack," as it were, with a special focus on whether government spending multipliers are state-dependent. The empirical strand of literature extends back at least to Perotti (1999), although the bulk of the studies are more recent.

Theoretically, much research can be classified according to the neoclassical tradition, with one of the more famous papers being Baxter and King (1993), or the new Keynesian variety, such as Rotemberg and Woodford (1992). These studies do not consider any state-dependence in the government spending multiplier, that is whether the increase in output for a one dollar increase in government purchases of goods and services varies along with economic conditions. After all, with the Federal Reserve unable to lower interest rates any further and resorting to methods of uncertain efficacy, this is the question that policy makers would like to know the answer to. If fiscal stimulus is undertaken in a time of recession or a time of relatively low capacity utilization, might it not be more powerful in boosting output? Particularly in the new Keynesian class, however, a number of new models have appeared that attempt to provide some support for this notion, using a number of different propagation mechanisms.[1] Although the prediction of most of these models is that government purchases have a much higher multiplier when economic times are bad, empirically, the evidence for this is less certain.

Even without taking possible state-dependence into account, estimating the effect of government purchases on output is fraught with difficulty. Ramey (2011a) discusses a large number of empirical studies, conducted using a wide range of methodologies, and concludes that the average government spending multiplier over all time periods and states of the economy is likely to fall somewhere between 0.8 and 1.5. When state-dependence is taken into account, the econometrician must decide among a bevy of options as to how to best estimate this statistic. This paper will address

[1] See, for example, Christiano, Eichenbaum, and Rebelo (2011), Sims and Wolff (2013), Galí, López-Salido, and Vallés (2007), and Denes, Eggertsson, and Gilbukh (2013) among many others.

whether, among these choices, some decisions systematically lead to higher or lower multiplier estimates in a state-dependent context.

To this end, I have identified eight dimensions along which any attempt to measure the government spending multiplier in good times and in bad may vary. These include whether to adopt deterministic or stochastic time trends, the strategy for identifying fiscal shocks, whether or not to account for monetary policy, the choice of variable to demarcate good times and bad times, whether the level of that variable demarcating good and bad times is estimated or imposed *ex ante*, the particular sample period, how to estimate the impulse response functions, and whether to allow the coefficients on all variables in the regression equation to switch with the state of the economy, or only a select number. By no means is this an exhaustive list of the possible empirical methodologies to choose from, of course, and I will not examine every one. The goal of this paper is primarily to find out the general extent to which specification matters in answering this question. As a large number of studies have come to different conclusions (with some finding evidence for large differences in the multiplier between good and bad times and some finding no evidence that the multiplier is ever above unity, no matter the economy's condition), it seems appropriate to investigate if certain specification assumptions seem to bias multiplier estimates, in either state of the economy, in one direction or another.

A major inspiration for this study is Sala-I-Martin (1997), who estimates four million regressions of economic growth to find if any certain variables are significant more often than others.[2] Like in that paper, I will estimate a large number of pairs

[2] In the context of estimating the effects of fiscal policy, Engemann, Owyang, and Zubairy (2008) also do something similar.

of government spending multipliers,[3] 2112 to be exact, with each estimate differing from the one before only by one specification choice. With these in hand, I can learn whether some specifications are prone to over- or under-stating the multipliers relative to other choices via regression analysis. This will not shed any light on the "true" multipliers, but it can at least inform how the results of any one paper would look if minor adjustments to the specification were made. From a theoretical perspective, Coenen et al. (2012) and Kormilitsina and Zubairy (2013) perform exercises in a similar spirit by easy examining the predictions of a large number of models for the effect of fiscal shocks.

I find that, on average, fiscal multipliers in the "good" state are lower than multipliers in the "bad" state overall. I also find that the choices for trend specification, identification strategy, sample period, and impulse response estimation systematically nudge the government spending multiplier in one direction or another, no matter the state of the economy. The choice of threshold variable that distinguishes good and bad times matters in some instances, while the inclusion of the monetary policy stance and the decision to let all coefficients switch or not do not seem to have consistent effects.

The rest of this paper proceeds as follows. Section 4.2 discusses the various dimensions along which I will vary my multiplier specifications, as well as the data to be considered. Section 4.3 presents the results, and Section 5.6 offers a brief discussion and conclusion.

[3] A pair comprises a good state multiplier and a bad state multiplier.

4.2 Specification Choices

The majority of the data for this study comes from the historical data set assembled by Ramey and Zubairy (2014). It includes quarterly data on real GDP and real government spending, as well as the military spending news variable developed by Ramey (2011b) and Owyang, Ramey, and Zubairy (2013), in the United States from 1890 to 2010. I augment the data set with a series for real quarterly tax revenues, which spans 1929 to 2010.[4] Because I always include tax revenues in my regressions, the earliest start date in any analysis is 1929. Also, I obtain from the St. Louis Federal Reserve Economic Database as well as the NBER Macrohistory Database time series on the three-month Treasury bill rate, which I will use to proxy for monetary policy.[5] Similarly, data on the Ten-Year Treasury interest rate comes from both of these sources as well.[6] Data on the Moody's AAA and BAA corporate bond yields are retrieved from the St. Louis Federal Reserve. The rest of this section concerns the dimensions along which I will make specification choices.

[4]To generate a quarterly real tax revenues series for the period from 1929 to 1946 (before the government began recording tax receipts on a quarterly basis, I first seasonally adjusted the monthly federal surplus series from the NBER Macrohistory Database, and use this seasonally adjusted data to interpolate the annual NIPA series (W054RC1Q027SBEA) via the Denton method. This follows Owyang, Ramey, and Zubairy (2013) and Ramey and Zubairy (2014). Please note that the series is actually tax receipts, although I may refer to it as "tax revenues" in the course of the paper.

[5]The Treasury bill data 1929:I to 1934:I comes from the Macrohistory database, and all of the subsequent observations come from the St. Louis database.

[6]Observations from 1929 to 1943 are from the Macrohistory database series m13033a, and those from 1944 to 1953:I are from the Macrohistory database series m13033b. Subsequent observations come from the St. Louis Federal Reserve.

4.2.1 Time Trends

The first specification choice to be made is whether to include determinisitic time trends, such as a linear, quadratic, or even quartic time trend, or to somehow account for differences over time with some kind of stochastic trend. For example, Ramey (2011b), Ramey and Zubairy (2014), Alloza (2014) and Fisher and Peters (2010) estimate their regressions or vector autoregressions with time trends corresponding to the lengths of their respective samples. Other studies, such as Fazzari, Morley, and Panovska (2013) and Jordà and Taylor (2013) rather put all of their dependent and independent variables in first differences. In some cases, both approaches are employed, either separately (Blanchard and Perotti (2002)) or within the same equation system (Riera-Crichton, Végh, and Vuletin (2014)). In some cases, stochastic trends are employed, but expressed not as first differences, but as deviations of the series from the Hodrick-Prescott filtered trend, such as in Jordà and Taylor (2013) or Auerbach and Gorodnichenko (2014).

It is worthwhile to consider why this might make a difference in the estimation of government spending multipliers. Ramey (2011b) and Ramey and Zubairy (2014) make the argument that demographic changes over time have stimulated dynamics in hours per capita, which a linear or quadratic trend might well control for. There are, however, reasons to worry about the use of these trends. Nelson and Kang (1981) point out that detrending a time series that contains an autoregresive unit root with a linear trend generates spurious cycles. Thus, in a regression context, the detrended right hand side variables do not control for any cyclical fluctuations, and, in fact, the information contained in them is likely to be inaccurate, leading to imprecise estimates. A very similar problem afflicts other popular time series filters, such as the Hodrick-Prescott filter and the Baxter-King filter, as discussed in Cogley and Nason (1995), Murray (2003), and Nelson (2006). For these reasons, it seems perilous to use deterministic time trends or various "atheoretical" time series filters to control for the passage of time, and the safest route may be to use first differencing. On the other hand, Gospodinov, Herrera, and Pesavento (2013) show that, in the case of VARs, small deviations from an exact unit root in the relevant time series can produce spurious results when difference-stationary restrictions are imposed.

In this paper, I consider specifying trends using the deterministic method and the first differenced method, as these are the most popular in the literature. Half, then, of the pairs of multipliers I will present will be estimated with quadratic trends (for the shorter samples) or quartic trends (for the longer samples), and the other half will measure all nonstationary variables in first differences.

4.2.2 Identification Strategy

The difficulty of identifying truly exogenous fiscal shocks has been considered at length in the literature. To the extent that governments run countercyclical fiscal policy, one might imagine that there is not much in the way of government purchases that is actually unrelated to the state of the economy and is not predictable by private agents. Thus, there have been a number of proposals put forward to facilitate identification. In this study, I will discuss three of the more popular ways, although there have been other strategies as well.

Perhaps the most venerable procedure is that adopted first in Blanchard and Perotti (2002), who specify a vector autoregression that includes output, government spending, and tax revenues. While a number of VARs have been utilized that include all manner of different variables in addition to these three (or, in some cases, instead of tax revenues), the vast majority rely on the same basic logic. By ordering government spending first in the VAR, ahead of output, one implicitly makes the assumption that, while output might respond contemporaneously to a change in spending, government spending cannot respond to output without at least a lag of one period. On its face, this seems quite reasonable, as, in most advanced economies, and in the United States in particular, the legislative process required for authorizing new government purchases often takes a nontrivial amount of time. With this assumption in hand, the fiscal shock is identified as the structural innovation to government spending in the VAR (which is identical to the reduced form innovation when government spending is ordered first). Among the studies identifying fiscal shocks in this way are Auerbach and Gorodnichenko (2012b), Ilzetzki, Mendoza, and Végh (2013), and Gordon and Krenn (2014).

Ramey (2011b) puts forth a forceful and convincing argument that this may not be the best scheme for identifying shocks. She shows that the innovations in government spending from a vector autoregression are likely to be anticipated, as the hypothesis that they cannot be predicted by the war dates identified in Ramey and Shapiro (1998) is strongly rejected. If these spending innovations are expected by the public at large, it is likely that their behavior has already begun to adjust to them, throwing off the timing of the effects of interest. This is also a problem with empirical studies that take an annual frequency, such as Barro and Redlick (2011) and Hall (2009), and why Owyang, Ramey, and Zubairy (2013) and Ramey and Zubairy (2014) go to such pains to construct a historical data set that converts pre-World War II data to a quarterly frequency. With this in mind, Ramey (2011b) introduces a variable that pinpoints as well as possible, using contemporary news sources, when the public's expectations on government spending changed. This series only takes into account spending changes that are a result of some political or military event exogenous (as much as possible) to the economy of the United States, so as to avoid entanglements with endogenous counter- or pro-cyclical spending policy that might bias results. While Ramey (2011b) constructs such a series for the United States, Owyang, Ramey, and Zubairy (2013) do so for Canada, and Crafts and Mills (2013) do so for the interwar United Kingdom. The advantage of using this variable in identifying government spending shocks is hinted at above. It captures changes in expectations of spending, to which economic agents will react, and it is plausibly exogenous to current economic conditions. In many of the studies in this literature, if the Ramey (2011b) military news variable is not used as the primary government spending shock, it is often used at least as a robustness check.

The embrace of this method of identifying shocks is not universal, however. Fisher and Peters (2010) point out that, necessarily, the military news variable can cover only a limited number of observations, that it tends to feature mostly spending increases and only a few spending decreases, and that it is subject to what Ramey (2011b) readily admits are "judgment calls." Yang, Fidrmuc, and Ghosh (2014) even raise doubts about the series' exogeneity and argue that the kind of government spending spurred by such events is not typically the kind of spending that a government may want to pursue to stimulate the economy, a point echoed by Clemens and Miran (2012). Perhaps the most frustrating aspect of this particular variable is that it is, again almost necessarily by its very nature, limited to the United States. The United States is in a unique position in that it has been an active and preeminent military power over the last one hundred years, but has seen relatively little damage to its territorial productive capacity. These are the attributes that make the military news variable feasible, but also that constrain it. The Canadian and British series mentioned above, bu contrast, do not see the same variation over time as the American series. Cross-country studies, therefore, like Auerbach and Gorodnichenko (2012a), Ilzetzki, Mendoza, and Végh (2013), and Jordà and Taylor (2013) cannot easily take advantage of its benefits.

To get around this potential problem and to make use of the controls for anticipatory effects promised by this variable, other papers have inserted the Ramey (2011b) news variable into a VAR and used it to uncover orthogonalized government spending innovations not predicted by it. This method appears in Fazzari, Morley, and Panovska (2013) and Rossi and Zubairy (2011). In this paper, I will examine the relative output of using each of these three candidate identification schemes, evaluating the extent to which any of them might produce higher or lower multiplier estimates in either state of the world. Perhaps moreso than in any other dimension, however, the set of identification strategies I am examining is not exhaustive. Less prominent methods include the use of policy propensity scores in Jordà and Taylor (2013), sign restrictions in Mountford and Uhlig (2009) and Canova and Pappa (2007), error correction models in Candelon and Lieb (2013), forecast errors in Blanchard and Leigh (2013), or excess returns on defense contractors in Fisher and Peters (2010). Each of these present their own distinct trade-offs, and, while it is beyond the scope of this paper to consider them each in more detail, all may warrant further attention.

4.2.3 Monetary Policy

Despite the fact that many studies have emphasize the importance of the monetary policy stance for the effects of fiscal policy, such as in Davig and Leeper (2011) or Zubairy (2014), inclusion of monetary variables in estimation systems is nowhere near uniform. Biolsi (2015) shows, however, that controlling for the prevailing short-term interest rate makes a major difference for multiplier estimates in the postwar United States, especially for "bad times" multipliers. This may be because the monetary authority may be making an explicit choice to accomodate the supposed stimulating effects of government purchases or to offset their possible inflationary effects. In fact, Romer and Romer (2014) show that the likely cause of differences in the effects of increases in transfer payments and tax cuts is the Federal Reserve's differing responses to them. Of course, given proper identification of the fiscal shock, it is likely to be contemporaneously orthogonal to monetary policy. Although its importance has been emphasized, but its inclusion in empirical work is still relatively rare, this paper will directly test whether monetary policy variables (specifically, the short-term interest rate) have a systematic impact on estimates of the state-dependent government spending multiplier. Half of the multiplier pairs presented will control for the interest rate on three-month Treasury bills, and half will not.

4.2.4 Threshold Variable

Conceptually, the difference between "good" economic times and "bad" economic times is fairly clear, especially in the context of general equilibrium models, where the modeller can introduce negative technology shocks or negative aggregate demand shocks that jolt a system from its steady state. In practice, however, there is plenty of debate as to how to separate times of low slack from times of high slack. Auerbach and Gorodnichenko (2012b), Auerbach and Gorodnichenko (2012a), Auerbach and Gorodnichenko (2013), and Auerbach and Gorodnichenko (2014) use a seven quarter moving average of GDP growth to sort periods into times of "expansion" and times of "recession," but Ramey and Zubairy (2014) offer a critique of this decision, arguing that the relevant variable that would determine whether fiscal policy might be effective or not is how much spare capacity is available, not whether the amount of spare capacity is increasing or decreasing. This is because it is the amount of underutilized resources that will produce crowding out of private activity or not. To this end, they adopt the unemployment rate as their threshold variable. Gordon and Krenn (2014) agree with Ramey and Zubairy (2014) in principle, but assert that the more correct measure of slack in the economy is the output gap, noting that at the start of World War II, constraints on productive capacity started to bind long before the unemployment rate sank to a level that one might associate with tight labor markets.[7] The "output gap," however, can be an elusive concept, and it relies on a proper measurement of long term, or potential, output. Relative to the unemployment rate, this particular variable can be subject to a large degree of measurement error and disagreement. This can be seen in the proliferation of output gap candidates, such as the deviation of output from the Congressional Budget Office's measure of potential GDP, the deviation of output from its trend as measured by the filter of Hodrick and Prescott (1997), the deviation of output from its long run value as measured by the decomposition of Beveridge and Nelson (1981), or any of the other output gap measurements considered by Morley and Piger (2012). The CBO measure is limited by the fact that it does not extend past 1949, while the potential problems with the Hodrick-Prescott filter are noted above. The Beveridge-Nelson decomposition tends to produce very small cyclical components by attributing most fluctuations in output to permanent movements. Of those considered by Morley and Piger (2012), some are symmetric and some are asymmetric, and the possibility of structural breaks in the time series adds an additional measure of uncertainty.

Other researchers have pointed to a related, though distinct, indicator that may drive differences in the effects of government spending, which is the degree of credit tightness. Canzoneri et al. (2012) develop a model that features a financial intermediation mechanism that generates credit frictions and a financial accelerator. The idea is that increased government purchases boost output in the first order, and this output gain lowers credit spreads. Reductions in credit spreads stimulate a greater

[7]Kuhn and George (2014) develop a DSGE model in which the extent to which capacity constraints bind on firms affects the value of the government spending multiplier, thus formalizing the notion of a multiplier dependent on the output gap.

output response when the spreads are initially very large than when they are small. This implies that in addition to the unemployment rate and the output gap, credit spreads might also determine the magnitude of the government spending multiplier. To capture this, I will employ the difference between the Moody's AAA corporate bond yield and the ten-year Treasury yield, as well as the difference between the Moody's BAA bond yield and the Treasury yield and the difference between the AAA yield and the BAA yield. Gilchrist, Yankov, and Zakrajšek (2009) also use these differences between corporate and government bond yields to characterize the state of credit frictions.

In this paper, then, I consider eleven different variables to distinguish between the "good" state of the economy and the "bad" state of the economy. They are the unemployment rate, the CBO output gap, the Beveridge-Nelson output gap, an output gap computed via an unobserved components (UC) model, an output gap computed via the model of Hamilton (1989), a "bounceback" model along the lines of Kim, Morley, and Piger (2005), an output gap computed using the "plucking" model of Kim and Nelson (1999), the model-averaged output gap of Morley and Piger (2012), and the three credit spreads mentioned above.

4.2.5 Estimated vs. Imposed Thresholds

Given a certain economic indicator that the econometrician will use to distinguish between good and bad times, he or she will then be faced with the choice of determining what levels of that indicator might point to a particular time period being in the good or bad state. It could be that there is no one level of the threshold variable that separates good times from bad times, but rather the level is time-varying. This is the motivation behind using the cyclical component of Hodrick-Prescott filtered data as the threshold variable. Even if a single, time-invariant level is chosen, one must decide whether to impose a threshold level *ex ante* or estimate the appropriate level. In this paper, I will compare multipliers from systems with an imposed threshold level to those that come from an estimated level.

Ramey and Zubairy (2014), in their analysis of state-dependent multipliers using the unemployment rate, set 6.5% as the threshold for their benchmark analysis, with this value motivated by former Federal Reserve Chairman Ben Bernanke's comments that unemployment below this level would signal that the central bank could again start tightening monetary policy. Other papers that rely on detrending, such as Jordà and Taylor (2013), may distinguish good times from bad times to according to whether the threshold variable is above or below trend, or, as in Mittnik and Semmler (2012), whether output growth is above or below average. There are compelling reasons in all of these cases for the values chosen, but it is not clear that this chosen value gives the best evidence for possible differences in the effects on output of a change in government purchases. This is the motivation behind estimating the threshold level endogenously. Of the previous literature, one of the more prominent studies to take this route is Fazzari, Morley, and Panovska (2013), who use Bayesian techniques in a threshold vector autogression to compare models using different threshold variables and different levels of each variable, ultimately choosing the model with the best fit. In this paper, I will use the least squares method of

Hansen (2000) to identify the level of each threshold variable that offers the best evidence for state-dependence.

I do not take account of smooth transition models here, even though they have been employed heavily in Auerbach and Gorodnichenko (2012b), Auerbach and Gorodnichenko (2012a), Auerbach and Gorodnichenko (2013), Auerbach and Gorodnichenko (2014), Bachmann and Sims (2012), and Riera-Crichton, Végh, and Vuletin (2014). Although smooth transition models have several advantages, such as allowing the effect of purchases to vary with the degree of slack or recession in a way that discrete regimes cannot and potentially exploiting more information for the regime with fewer observations, they may suffer from the fact that they deliver parameter estimates that only apply to the two extreme cases of the world, which, in practice, may have delivered only a small number of observations. Also, they rely on a transition function that may have an arbitrary degree of smoothness. In this way, interpreting the results of smooth transition models may not be as straightforward as interpreting those of discrete regime models, which I will study in this paper. In any event, smooth transition models are another popular way to estimate state dependence in the multiplier that only add to the uncertainty produced by the models considered here.

4.2.6 Sample

One of the main contributions of Ramey and Zubairy (2014) was the introduction of newly constructed quarterly series on real GDP and real government spending in the

United States that dates back to 1889. The primary benefit is that it allows these authors to exploit much more variation in government spending (by including the periods of World War I and World War II) and more periods of deep recession and of ultra-loose monetary policy. In fact, they argue that the instrument relevance of their military spending news series in times of high unemployment depends vitally depends on the inclusion of this older data. Other papers that consider military spending as their primary instrument for government purchases, such as Barro and Redlick (2011) and Hall (2009), also are certain to make use of observations that cover some of the biggest fluctuations in spending and output in U.S. history, which provide helpful identifying variation.

There are drawbacks to this approach, however. Several authors have pointed out concerns associated with using heavily interpolated data, as that in Ramey and Zubairy (2014). Murray and Nelson (2000), for example, assert that interpolated data on output can bias the results of unit root tests. Gorodnichenko (2014) also questions whether it is appropriate to pool the data before 1947 (when most of the quarterly NIPA series begin) with that after 1947, considering the much higher volatility of measured spending before World War II, possible regime changes, and the possibility that interpolation of the data has attenuated the differences between good times and bad times. Therefore, any study of this question must navigate the trade-off presented by obtaining more information when using all of the data against the more homogenous character of the data after 1947.

In this study, I will consider two starting dates for all models, one being 1947 (the shorter sample) and one being 1929 (the longer sample). I start my longer sample in 1929, because this is when the annual NIPA series on tax receipts begin, and all of the models I consider will control for tax receipts on the right hand side. In this way, I lose the variation provided by the First World War, but maintain the much bigger fluctuations represented by the Great Depression and World War II.

4.2.7 Impulse Response Estimation

In recent years, the use of the Jordà (2005) local projection technique for estimating impulse response functions, especially in this particular context, has become very prevalent. This method has been applied in, among others, Auerbach and Gorodnichenko (2012a), Auerbach and Gorodnichenko (2013), Ramey and Zubairy (2014), and Riera-Crichton, Végh, and Vuletin (2014). Essentially, utilizing this method amounts to running a series of regressions,

$$y_{t+1} = \alpha_1 + \beta_1 Shock_t + \varepsilon_{t+1} \tag{4.1}$$

$$y_{t+2} = \alpha_2 + \beta_2 Shock_t + \varepsilon_{t+2} \tag{4.2}$$

$$\vdots \tag{4.3}$$

$$y_{t+H} = \alpha_H + \beta_H Shock_t + \varepsilon_{t+H} \ . \tag{4.4}$$

The impulse response function is then the series β_1, β_2, ..., β_H, which can then be plotted. Two series of regressions are run with first output growth and then growth in government spending as the dependent variables. There are several advantages associated with constructing impulse responses in this manner. First, there is no shape imposed on the response as there is when iterating recursively on the coefficients of an estimated vector autoregression. It is also robust to misspecification of the VAR, and it is well suited to nonlinear estimation environments. In particular, as noted by Ramey and Zubairy (2014), local projections do not compel the econometrician to have variables specified in exactly the same way on the left hand and right hand sides of the regression equation.

Doing so allows calculating multipliers directly from coefficients by specifying the dependent variables as $\frac{Y_{t+h}-Y_{t-1}}{Y_{t-1}}$ (for output growth) and $\frac{G_{t+h}-G_{t-1}}{Y_{t-1}}$ (for spending growth). This is because both regression equations are specified in the same units (a proportion of GDP). The integral government spending multiplier at the two year horizon, for example, is then merely $\frac{\sum_{h=1}^{8}\beta_{y,h}}{\sum_{h=1}^{8}\beta_{g,h}}$. Contrast this with the conventional method of computing multipliers in a VAR, in which one calculates elasticities of output growth and government spending with respect to the shock and then scale the elasticity ratio by the sample average of Y/G. Ramey and Zubairy (2014) demonstrate that this method can give biased results if the Y_t/G_t ratio changes much over time.

There are, however, concerns associated with this estimation technique, however. For example, especially at longer horizons, the estimated responses can become erratic with large standard errors complicating inference. Also, the specific nature of the dependent variables necessarily induces serial correlation in the error terms, which must be corrected in some manner. This estimation procedure also requires the use of long horizon regressions, which Sizova (2015) (2013) warns could be misleading. Finally, there need not be any smooth transition from the estimate of β_h to

β_{h+1}, leading to unintuitive jumps in the impulse response function. These concerns argue for the use of some sort of vector autoregression.

Since a standard linear VAR is incapable of producing state dependent multiplier estimates, I need to introduce some kind of nonlinearity. I do so by specifying a threshold vector autoregression, so as to keep within the spirit of the rest of the multiplier pairs, which rely on a threshold variable to distinguish between good times and bad times.[8] Threshold VARs are used in Fazzari, Morley, and Panovska (2013), with the impulse responses modelled using Bayesian techniques. In this paper, I will simulate Generalized Impulse Response Functions (GIRFs), as in Koop, Pesaran, and Potter (1996). GIRFs allow for impulse responses to differ according to the sign and size of the identified shock and provide a convenient way of allowing for feedback between states of the economy. They also control for dependence on the history of shocks to the system, which might influence the subsequent dynamics. A downside to employing GIRFs is that the threshold VAR will deliver only ratios of elasticities of output and spending to the shock and a (possibly bias-inducing) scaling factor will have to be applied to compute multipliers.

In this paper, half of the multiplier pairs I estimate will be calculated with the Jordà (2005) local projection technique, and the other half will use threshold VARs and GIRFs.[9]

[8]Although Smooth Transition Vector Autogressions (STVARs) have been used heavily in recent years Auerbach and Gorodnichenko (2012b), Auerbach and Gorodnichenko (2012a), Auerbach and Gorodnichenko (2013), Bachmann and Sims (2012), Caggiano et al. (2014)), they are ill-suited to the use of threshold variables, and so I do not consider them in this analysis. It is worth noting, however, that STVARs represent another possible way to arrive at very different estimated multiplier pairs.

[9]The code for calculating GIRFs is the same as that used in Schmidt (2013) (2013) and was generously provided by the author of that paper.

4.2.8 Expanded vs. Parsimonious Specifications

Each empirical specification will sort all observations into a "good" state and a "bad" state. Given that there is a question of which macro variables to include on the right hand side of each estimating equation (as discussed in Section 4.2.3, for example), a related issue concerns whether all or only a subset of these control variables ought to have switching coefficients. On the one hand, it may be desirable to allow only the coefficient on the estimated fiscal shock to switch, or, in the case of identifying fiscal shocks via orthogonalizing reduced form residuals to uncover structural shocks, only the coefficients on the government spending variable to switch. To that extent that the difference in the economy's response to an exogenous shock to spending is uncorrelated with the other control variables, this more parsimonious specification will allow more precise estimates of the effects of the spending shocks and conserve on degrees of freedom. On the other hand, restricting the coefficients on control variables such as lagged GDP log levels or growth to have identical coefficients no matter the initial state of the economy risks missing out on important mean reversion dynamics that may be state specific. This will be a problem if GDP has a high-growth "recovery" phase following recessions, as suggested by the work of Beaudry and Koop (1993), Kim, Morley, and Piger (2005), and Morley and Piger (2012). To that end, half of the multiplier pairs that I compute will restrict all control variables to have constant coefficients so as to take advantage of improved parsimony, and the other half will allow for regime-specific mean reversion dynamics.

4.3 Results

In this section, I consider the results of the analysis, starting by taking an overview of the effects of all of the specification choices, predominantly via the use of regression analysis. After that, I will conduct a Monte Carlo exercise to evaluate more closely the arguments behind specification of the time trend.

4.3.1 Regression Analysis

After estimating the state-dependent multipliers across all of the dimensions listed above, it seems appropriate to first take stock of their basic statistical characteristics. Figure 4.1 shows the density plots of the two-year good-state multiplier, the two-year bad-state multiplier, the four-year good-state multiplier, and the four-year bad-state multiplier. Table 4.1 gives the associated summary statistics. The picture painted by these two reports is that of a wildly variable distribution for each of the types of multipliers considered. In all cases, the mean multiplier estimate is negative (and at the two year horizon, it is negative with a very large absolute value), though the median estimate is positive and below unity. The standard deviation is in the range of about 350 in each case. The maximum and minimum estimates suggest that some methods of calculating the state-dependent multiplier can deliver figures of nearly 10000 or as low as -4500. This suggests that an extra dollar of government spending can produce $10000 in extra ouput or reduce output by as much as $4500. It is difficult to regard these notions as anything but absurd. The skewness and kurtosis calculations strongly reject that the distributions of estimates are normal, which is supported by just a quick glance at the histograms in Figure 4.1. Here, one can see a stacking of estimates at around zero, with very long tails stretching to both the left and the right.

With so many estimates well outside conventional theoretical (and even empirical) bounds, I start my analysis by considering if there are certain specification choices that are more likely to deliver extreme values. A multiplier estimate is considered to be "extreme" if it is higher than the 90th percentile of the relevant distribution or lower than the 10th percentile of the same. By the eighth and ninth columns of Table 4.1, any two year multiplier above 5.34 for the good state or 5.87 for the bad state or below -2.24 for the good state or -4.03 for the bad state is considered "extreme." Simiarly, at the four year horizon, this term applies to good state multipliers above 7.88 or below -0.54 and to bad state multipliers above 10.70 or below -2.84. I then run regressions of the following equations,

$$Extreme_i = \alpha + \beta \ Choice_i + \varepsilon_i \tag{4.5}$$

$$ExtremeHigh_i = \alpha + \beta \ Choice_i + \varepsilon_i \tag{4.6}$$

$$ExtremeLow_i = \alpha + \beta \ Choice_i + \varepsilon_i \ , \tag{4.7}$$

where $Extreme_i$ is a dummy indicator that takes on a value of 1 if the multiplier estimate is extreme, as defined in the preceding paragraph, and 0 otherwise. $Choice_i$ is a dummy indicator that takes on a value of 1 if the estimation that resulted in a given multiplier estimate employed a given specification choice. For example, in the regressions that consider the effect of using the unemployment rate as the threshold variable on the likelihood of getting an extreme estimate, $Choice_i$ switches on when the multiplier estimate used an unemployment rate as the threshold and 0 when any other variable was used. $ExtremeHigh_i$ and $ExtremeLow_i$ are dummy indicators that mark when a particular extreme estimate was especially high (above the 90th percentile) or especially low (below the 10th percentile).

Tables 4.2 and 4.3 give the results of these regressions. The tables report only the results on extreme values for the "good" state multiplier estimated at the two year horizon, but the results for the "bad" state multiplier at the same horizon and for both states at the four year horizon are quantitatively and qualitatively similar. A number of things stand out from these regressions. The first is the observation that the use of a deterministic time trend increases the probability of obtaining an extreme estimate by 32 percentage points, a dramatic increase, compared to the use of a stochastic trend. This is split about evenly between extremely positive and extremely negative estimates. In the case of the two year good state multiplier, deterministic trends are able to account for both the highest and lowest values reported in Table 4.1, and the standard deviation is 494.58, 41% higher than the standard deviation for all of the multiplier estimates as a whole. This suggests that the employment of a deterministic time trend requires careful specification of the other elements of the regression if one is not to wind up with absurd multiplier estimates.

Table 4.3 shows that the use of generalized impulse response functions (as opposed to the Jordà (2005) local projection technique) brings with it similar dangers. It increases the likelihood of an extreme estimate by 39 percentage points, again split evenly between extremely high and extremely low figures. As Ramey and Zubairy

(2014) show, this could in part be due to the need to scale ratios of elasticities up by an ex post scaling factor, a problem not encountered when one uses local projections to estimate the impulse responses. The incorporation of the military news variable introduced by Ramey (2011b) and extended by Ramey and Zubairy (2014), whether it is used as the fiscal shock itself or in a vector autoregression so as to purify standard orthogonalized government spending shocks of the likely anticipation of future spending, also seems to increase the probability of an extreme estimate somewhat, relative to the conventional structural shock popularized by Blanchard and Perotti (2002).

Other specification choices have relatively modest impacts on the probability of an extreme result. Accounting for monetary policy has almost no effect, as does the use of a more parsimonious specification. Some threshold variables make an extremely high estimate a bit less likely. Finally, although using the longer data sample does not make an extreme estimate in general very much more likely, it does raise the probability of an extremely high estimate, while reducing the probability of a very low one. This implies that employing data from before World War II, on average, may lead to higher multiplier estimates relative to using only postwar data.

When the extreme multiplier values are excluded, the distributions become somewhat more comprehensible, as can be seen in Figure 4.2. Still, they do not seem to approximate a normal distribution. Whether one considers the two year or four year horizon or the good or bad state, there is an evident rightward skew to these distributions. That is, no matter how one chooses to estimate the state-dependent multiplier, large values (greater than unity) are much more likely than negative values, although it is still clear that the majority of estimates cluster just above zero. Table 4.4 gives the summary statistics for the truncated distributions. From the table, it is clear that the hypothesis of a normal distribution is clearly rejected. Some patterns do emerge, however. The mean and median "bad" state multipliers at both time horizons are greater than the associated "good" state multipliers. This accords with the majority of the empirical literature on the subject (Auerbach and Gorodnichenko (2012b), Bachmann and Sims (2012), Nakamura and Steinsson (2014), and others). Bad state multipliers are also more dispersed than good state multipliers, as evidenced by their higher standard deviations. Multipliers tend to be larger at the four year horizon than at the two year horizon, suggesting that the effects of government spending shocks are fairly persistent. The rightward skew is also more apparent at the four year horizon.

Before examining the effect of various specification choices on these more moderate multiplier estimates, it may be useful to consider to what extent the trimming of the distribution changed the relative frequencies over each dimension.[10] Table 4.5 shows the new allocations across specification choices for each time horizon and state of the economy, across which dimensions the breakdowns are fairly consistent. In most cases, there is still a roughly equal number of multiplier estimates for each specification choice, the exceptions being the choice of time trend and impulse response estimator. There are more multiplier estimates with stochastic trends and

[10] Before truncating the distribution, all of the pairs of multiplier estimates were allocated equally across each dimension. The only exceptions are that estimations conducted using the CBO output gap and the model averaged output gap were not run using the longer sample period, due to a lack of data (the former) and computational issues (the latter).

Jordà (2005) impulse response estimation, as it was shown above that deterministic trends and generalized impulse responses were more likely to lead to extremely high or low multiplier values. Even across these dimensions, however, there are enough observations left to get a reasonable idea as to the overall effect on the multiplier estimate within the bounds of the more theoretically plausible results.

As the next step in the analysis, I will look to see if any particular specification choice systematically leads to higher or lower multiplier estimates. I will run regressions of the following equation,

$$Multiplier_i = \alpha + \beta \; Choice_i + \varepsilon_i \; , \tag{4.8}$$

where $Multiplier_i$ refers to the actual multiplier estimate arrived at, and $Choice_i$ refers to a specification choice as in the regressions above. These regressions are run for each time horizon and state of the economy. I will run univariate regressions of this sort as well as multivariate regressions in which all dimension choices are included in the same regression.[11] Since $Choice_i$ is a dummy variable taking on a value of one if a given specification choice is employed, one can read the coefficient β as the average increase or decrease in the multiplier engendered by that choice, holding all other specification decisions constant. The results are in Tables 4.6 and 4.7.

When considering the tables, perhaps the most eye-catching feature is the fact that in univariate regressions, using a deterministic time trend has no significant effect on the magnitude of any version of the multiplier, but, when one includes the

[11] As the extreme multipliers are dropped, it is no longer necessarily the case that all of the specification choices are orthogonal to each other, creating scope for possibly different coefficient estimates in the univariate and multivariate regressions.

other specification choices in the regression as well, it significantly magnifies every kind of multiplier. I will discuss this issue more below.

It is also evident that the use of the military news variable used in Ramey (2011b) and other studies also tends to amplify the multiplier in all states and at all horizons. The military news variable primarily aims to capture the public's anticipation of future government spending changes. Therefore, identifying a spending shock via the structural residuals from a VAR that includes government spending, output, tax revenues, and potentially other variables, but not the military news variable leaves one exposed to the possibility of spurious inference on the multiplier. In principle, this bias could be either upwards or downwards. Failing to take account of expectations of future spending could lead one to ignore behavioral changes before the spending actually hits the economy but that nonetheless is caused by the spending change. If these behavioral changes lead to gains in output, the multiplier estimate will be biased downward, whereas if they cause output to decline, it will be biased upward. The results in the third and fourth lines of Table 4.6 suggest that the multiplier obtained via identification of structural shocks to government spending in a VAR that excludes the military news variable is consistently understated.

Ignoring prewar data also seems to understate the multiplier at all horizons and in all states of the world, relative to only looking at data from the postwar period. The fifth line of Table 4.7 suggests that including observations from the Great Depression and World War II adds anywhere from \$0.18 to \$0.80 to the estimate. This effect is highly significant.

The final very salient point to note is that estimating impulse responses via the local projections method of Jordà (2005) tends to reduce multipliers of every type by a very large amount, with the smallest reduction being a full $1.73, relative to estimation via generalized impulse response functions. This is even after excluding extreme multiplier estimates, which local projections reduced the probability of by nearly forty percentage points (see Table 4.3). Since this approach to impulse response estimation is much less likely to give implausibly high or low estimated values, and even among the trimmed observations, gives more conservative values on average, using this relatively flexible method (as in Auerbach and Gorodnichenko (2012a) and Ramey and Zubairy (2014)) seems the more robust approach to finding the state-dependent fiscal multiplier.[12]

Other specification choices seem not to matter as much. Inclusion of a monetary policy variable or endogenous estimation of the threshold generally do not make a significant difference for any type of multiplier. The same can be said for whether or not one allows all regression coefficients to switch along with the state of the economy.

One may very well be interested not so much in whether or not some specification choices amplify or diminish both the good and bad states of the multiplier as in whether they accentuate the differences between them. The only choice that seems to do this is the choice of the threshold variable itself. In particular, the use of the CBO output gap appears to drive down the multiplier in the bad state of the world by about $0.35 (while having no impact on the good state multiplier). Similarly, when

[12] On average, when extreme values are dropped, the Jordà (2005) local projection technique deliver two year multipliers of 0.371 when the state is good and 0.438 when the state is bad, whereas GIRFs deliver analagous two year multipliers of 2.103 and 2.321. These are very large values, even theoretically.

the unobserved components output gap (such as that in Clark (1987)) is employed as the threshold variable, the good state multiplier tends to be boosted by between $0.20 and $0.50. At the four year horizon, a similar effect can be seen by use of the nonlinear unobserved components output gap and the model-averaged output gap of Morley and Piger (2012). All of these differences are relative to the omitted category, which in this case is the unemployment rate.

From the other side of the argument, variables that attempt to measure the degree of credit tightness seem to push the bad state multiplier up and the good state multiplier down, thus increasing the likelihood of concluding that the multiplier is countercyclical. This is true for the AAA-10 Year spread and the BAA-10 Year spread, but especially so for the BAA-AAA spread. The apparent reliance of the difference between good and bad state multipliers on the choice of threshold variable will be treated in greater detail below.

Next, I will consider if there are certain specification choices that are more likely to result in multiplier estimates that are significantly greater than unity at the ten percent level or significantly less than zero at the same level of confidence. This will be done via regressions of the sort of Equation 4.8 only with dummy indicators for whether or not a multiplier is greater than one at ten percent significance or negative at ten percent significance taking the place of the dependent variable in the regression. The results are tabulated in Tables 4.8 and 4.9.

First, I consider the probability of pushing multipliers significantly above one. It is clear that the use of deterministic trends is anywhere from 25 to 29 percentage points more likely to give a multiplier of any state or horizon above one, which is not very surprising given that these sorts of trends resulted in more extreme multipliers generally. Compared to not using the military news variable, using it makes multipliers of any state or horizon much less likely to rise above one significantly.

The choice of threshold variable also influences the chance of getting multipliers greater than one, particularly when evaluating the good state of the economy. One is more likely to get a large, positive multiplier when times are good when the threshold variable is the CBO output gap, the unobserved components output gap (linear or nonlinear), the model averaged output gap, or the AAA or BAA spreads. Bad state multipliers are less likely to be large and positive for the CBO output gap, the Hamilton output gap, and the model averaged output gap, while interest rate spreads tend to push them above unity.

The good state multiplier also tends to be larger when the longer historical sample is used. Finally, estimating impulse responses via the Jordà (2005) method puts downward pressure on the likelihood of a multiplier above one no matter what kind of multiplier is being considered.

As far as significantly negative government spending multipliers (for which there is relatively little theoretical justification), again deterministic time trends makes them statistically more likely in most cases. This accords with the notion that multiplier estimates are generally just more likely to be extreme when the data is detrended this way. For the most part, identifying one's government spending shocks using the narrative military news shock makes a negative multiplier less likely.

Although the inclusion of a monetary policy variable has not had a great impact in any of the experiments conducted to this point, it apparently does generate a statistically more likely chance of getting a negative multiplier estimate, for any kind of multiplier. Similarly, fixing the threshold level *ex ante*, although seemingly not crucial decision for the most part, does apparently give a greater probability of getting significantly negative bad state multipliers.

When thinking about the threshold variable itself, the bad state multiplier, especially at the four year horizon, is more likely to wind up significantly negative if one uses the CBO output gap, the Beveridge-Nelson output gap, the Hamilton output gap, the Bounceback output gap, the model averaged output gap, or the difference between the AAA interest rate and the ten year Treasury interest rate to distinguish between good and bad states of the economy. Use of the other two interest rate spreads being considered is more likely to result in good state multipliere below zero for any horizon.

Consistent with most of the results so far, the local projections IRF method has a dampening effect, producing a greater probability of significantly negative multipliers. Making use of pre-World War II data has the opposite effect. Finally, at the two year horizon, allowing all regression coefficients to switch between states makes the bad state multiplier more likely to be below zero and the good state multiplier less likely to be so.

Of course, the objective that motivates any paper in this literature is to determine whether the government spending multiplier is of greater magnitude when times are "bad," for any given specification. Thus, my next exercise is to evaluate if the difference between good and bad state multipliers depends on the empirical specification. Table 4.10 reports summary statistics on the difference between the multipliers for any given specification choice. The difference is specified as $Mult_{Bad} - Mult_{Good}$, so that positive numbers indicate a higher multiplier in bad times. The table demonstrates that, generally speaking, at either horizon, the difference between the multipliers in the two states tends to be fairly small. The median differences cluster around zero, and, when extreme values are excluded, the bulk of them range between -0.63 and about 1.5. There is a positive skew to the distribution when extreme values are dropped, suggesting that multipliers estimated for bad states are generally larger than those estimated for good states. This is consistent with the summary statistics shown above for the good and bad state multipliers separately.

Table 4.11 displays the results from a regression very similar to that in Equation 4.8, in which the dependent variable is the difference between the bad state multiplier for a given specification and its good state counterpart. As before, positive values indicate a higher multiplier in the bad state of the world and extreme values have been excluded. The table shows that using a deterministic trend dampens the difference between each multiplier in a pair by about \$0.085 at the two year horizon, significant at the ten percent level, although its effect at the four year horizon is not statistically significant. Relative to the use of a standard Cholesky decomposition identification scheme as in Blanchard and Perotti (2002), including the military news variable, whether on its own or as part of a VAR, amplifies the difference by between \$0.23 and \$0.38. Also, extending one's sample back to 1929 tends to reduce the difference at any horizon by a statistically significant amount. This seems to accord with the observation that papers that make use of a greater amount of historical data (see, for example, Ramey and Zubairy (2014), Barro and Redlick (2011), or Crafts and Mills (2013)) all find a lack of state dependence in their multiplier estimates. Other choices, such as controlling for monetary policy or allowing all regression coefficients to switch or even the use of local projections to calculate the impulse responses, do not seem to influence the difference between any specification's multiplier pair. Indeed, considering impulse responses as calculated by the Jordà 2005 method, multipliers seemed to be driven lower for either state of the world, so it is not surprising that using this technique would have an impact on their difference.

Choice of the threshold variable, on the other hand, has a considerable influence on the likelihood of finding big differences between the good and bad state multipliers. In particular, if one were to conduct the empirical analysis assuming that the CBO output gap or that described by a linear or nonlinear unobserved components model were the best variable to delineate good and bad times in the economy, then the result would be finding a relatively small difference between the bad state multiplier and the good state multiplier. One would therefore be likely to conclude that the state of the economy does not matter much for the output effects of fiscal policy. The opposite conclusion would be more likely reached if the chosen threshold variable were the Beveridge-Nelson output gap or any of the interest rate spreads, as proposed by Canzoneri et al. (2012). The interest rate spreads, in fact, can add close to a dollar to the difference between the two multipliers.

The analysis in this section is not meant to suggest that any specification choice is "right" or "wrong," or that any of them lead to a bias of any kind. To make such a statement would require knowledge of the true state dependent multiplier, which, of course, is not available. It is just meant to show that selection along any of these dimensions is likely to have a material impact on the results of the analysis. To make the problem that much more thorny, as can be seen from Section 4.2, there are quite valid conceptual and practical reasons for any of the choices available to the econometrician. The next subsection will deal more closely with a dimension that seems to have an especially large effect, the time trend specification.

4.3.2 Monte Carlo Study of the Time Trend Assumption

In order to get a better handle on the consequences of assuming that the relevant time series (specifically log GDP per capita and log government purchases per capita) are stationary around a deterministic time trend or stationary only in first differences, I undertake a Monte Carlo study similar in spirit to that of Christiano (1992). I start by estimating the following two systems of equations for the period 1948 to 2010:

$$\boldsymbol{y}_{1,t} = A_1 \boldsymbol{y}_{1,t-1} + B_1 \boldsymbol{Milnews}_t + C_1 \boldsymbol{trend} + \boldsymbol{\epsilon}_t \tag{4.9}$$

$$\boldsymbol{y}_{2,t} = A_2 \boldsymbol{y}_{2,t-1} + B_2 \boldsymbol{Milnews}_t + \boldsymbol{\varepsilon}_t \ . \tag{4.10}$$

Here, $\boldsymbol{y}_{1,t}$ is a vector containing the log of real per capita output and government spending, while $\boldsymbol{y}_{2,t}$ is a vector containing the first differenced specifications of the same two variables. $\boldsymbol{Milnews}_t = [Milnews_t, Milnews_{t-1}, Milnews_{t-2}, Milnews_{t-3}, Milnews_{t-4}]'$, where $Milnews_t$ denotes the military news variable of Ramey (2011). The vector $\boldsymbol{trend} = [t, t^2]'$ is a quadratic time trend, and the matrices A_1, B_1, C_1, A_2, B_2 are coefficient matrices. The vectors $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\varepsilon}_t$ are error vectors for their respective systems. Note that the first system imposes a deterministic time trend, while the second imposes a stochastic time trend. Table 4.12 includes the coefficient estimates for these regressions. As is apparent from the inclusion of the military news variable, this Monte Carlo simulation will take this as the identified shock to government purchases. From System 4.9, the estimated errors have the following covariance structure:

$$Covar_{Det} = \begin{bmatrix} 0.0000803 & 0.0000208 \\ 0.0000208 & 0.0001512 \end{bmatrix} . \tag{4.11}$$

From System 4.10, the estimated errors have this covariance structure:

$$Covar_{Stoc} = \begin{bmatrix} 0.0000817 & 0.0000239 \\ 0.0000392 & 0.0001436 \end{bmatrix} . \tag{4.12}$$

I then simulate 500 different series of output and government spending assuming a deterministic time trend with the coefficients in the first two columns of Table 4.12 and the covariance structure given by the matrix in Equation 4.11 as well as 500 different series of the two variables assuming a stochastic trend using the coefficients in the second two columns of Table 4.12 and the covariance matrix in Equation 4.12. I use the first four observations on output and government spending from the actual data as initial conditions for each simulation. For the military news variable, I sample with replacement from the observed series for each simulation. The deterministic series simulations then have 248 observations each, while the stochastic series have 247 observations each, with one observation lost to differencing.

For each simulation, once the data has been generated, I estimate a government spending multiplier using the Jordà (2005) estimation technique (as in Section 4.2.7. First, I estimate a linear multiplier. That is, I do not allow for any state dependence. For the deterministic processes, I estimate the multiplier by estimating the Jordà (2005) regressions in log levels and including a deterministic time trend and in log differences. I do the same for the difference-stationary processes. Thus, I will produce four different multipliers in each simulation: a multiplier estimated assuming trend-stationarity on actual trend-stationary data, one assuming difference-stationarity on trend-stationary data, one estimated assuming trend-stationarity on difference-stationary data, and one assuming difference-stationarity on actual difference-stationary data.[13]

Histograms of the linear multiplier estimates are contained in Figure 4.3 and a tabulation of the distributions is in Table 4.13. When examining the linear multiplier estimates, a couple of different points stand out. First, conditional on the data being generated by a particular process, there is not much difference in the central tendency of the multiplier between estimations that impose trend-stationarity of the data or difference-stationarity. For example, the median multiplier estimate when the true processes have a deterministic time trend are 0.46 when one imposes a quadratic trend and 0.50 when one does not. Similarly, if the true processes are nonstationary, the median multipliers are between 0.74 and 0.80 no matter how they are estimated. Secondly, when the data has a deterministic trend, the multiplier estimates tend to be skewed to the left, as can be seen by some of the negative

[13] All of the multipliers reported will be at the two year horizon, so as to be consistent with the results reported thus far. Multipliers at the four year horizon have similar implications.

multiplier estimates with large magnitude, but there is a much less obvious skew when the data is generated via a stochastic trend. Lastly, conditional on a data generating process with a stochastic time trend, estimating the multiplier including a deterministic trend leads to a slightly more diffuse distribution.

This paper is primarily interested in nonlinear estimation of the government spending multiplier, so the next exercise considers estimation of a multiplier that depends on the level of the unemployment rate at the time the spending hits the economy. To incorporate this in my Monte Carlo analysis, for each set of simulations, I also generate an unemployment series, which is randomly sampled with replacement from the actual unemployment rate observations from 1948 to 2010. For this exercise, I am obviously assuming that the econometrician is using the unemployment rate as the threshold variable. I follow the same procedure as when estimating the state dependent multipliers in the postwar period for an exogenously imposed unemployment rate threshold. That is, observations where the lagged simulated unemployment rate is below the median in the sample, are considered to have been observed in the "good state," while those where the unemployment rate is above the median are identified as being in the "bad state." Table 4.14 contains the distribution of multiplier estimates for the high- and low-unemployment states, while a graphical respresentation can be found in Figures 4.4 and 4.5.

What can one learn from these results? Consider first the good state of the economy. In this case, if the true process is deterministic and a deterministic time trend is used in estimation, the subsequent results are fairly well behaved. A relatively symmetric and tightly spaced distribution centered around 0.50 emerges. If instead, estimation is executed in first differences, a long right tail arises, raising the possibility of estimating some extremely large good-state multipliers. In fact, however, the same phenomenon occurs when the true process is nonstationary and one runs the regressions in first differences. Even if the process is nonstationary, applying a time trend does not dramatically distort the distribution of multiplier estimates, although there is a slightly greater probability (relative to when the true process is trend-stationary) of arriving at a negative multiplier of a fairly large magnitude. These results imply, therefore, that applying a linear or quadratic time trend to the data does not have an especially strong effect on the estimates of the good state multiplier even when it is inappropriate to do so.

Next, I look at the estimates for the bad state of the economy. Here, the results are generally less sanguine. Applying a deterministic time trend where it is appropriate to do so raises the possibility of extremely negative estimates, as a long tail extends to the left. When the true process is deterministic and stochastic trend estimation is applied, a problem of opposite sign if more muted degree emerges, as the tail extends to the right. The most well-behaved results seem to be when the true process follows a non-stationary process, but deterministic methods are applied to detrend it. This produces the most compact distribution, with the median estimate around 0.68. Still, stochastic estimation applied to a stochastic process performs fairly well also.

The preceding exercise assumes that there are an equal number of an observations in both the good and bad states of the economy. When, however, ones estimates the threshold delineating the two states endogenously, it is quite possible that one regime, usually the one representing bad states of the economy, will have a much smaller number of observations. To address this possible concern, I repeat the above state-dependent simulation exercise setting the threshold for the bad state at the 66^{th} percentile of the unemployment rate distribution. See Table 4.15 and Figures 4.6 and 4.7 for the results of this exercise.

When the threshold unemployment rate that distinguishes the states of the economy is shifted upward, one can see that estimation of the good state multiplier is not greatly affected by choice of the time trend specification no matter the true data generating process. When one considers the bad state of the economy, however, usage of deterministic time trends seems to increase the likelihood of extreme multiplier values, regardless of the true process. This confirms the empirical findings. Estimating the multiplier using first differences is not totally immune to this problem, though. Very large positive and negative values are still fairly likely.

The results of this subsection offer very modest evidence that specifying the local projections regressions using a deterministic time trend can lead to lower precision in estimation of the multipliers. Still, it would be inappropriate to assume that a stochastic trend specification will eliminate this problem. Perhaps due to the relatively small number of observations, especially when one is splitting the sample, and the long horizon regressions being relied upon (which places even more pressure on the data), it could be that, no matter the technique used to estimate the impulse responses, the results can be very erratic. In the concluding section, I will revisit what this simulation exercise might mean for the estimation of government spending multipliers in different states of the world.

4.4 Conclusion: Does the Specification Matter?

This paper has sought to answer why, for all of the empirical research devoted to finding if the government spending multiplier is state dependent, there is yet so little consensus on the issue. Papers that make different specification assumptions across a number of dimensions come to different findings on the issue. For example, two of the most well-known papers on the issue, Auerbach and Gorodnichenko (2012b) and Ramey and Zubairy (2014), have diametrically opposed conclusions, with the former finding that the multiplier is very large during recessions and almost zero during expansions and the latter finding no difference at all in the multiplier for the two different states of the world. I have estimated over two thousand pairs of good- and bad-state multipliers in an attempt to see if any particular specification choice leads to systematically higher or lower estimates.

I find that there some choices that are more likely to result in an extremely high or low multiplier, such as imposing a deterministic time trend or estimating impulse responses via the use of generalized impulse response functions. Multipliers tend to be higher when fiscal shocks are identified using narrative methods or when data from before World War II are included. One's choice of the macroeconomic variable that distinguishes between good and bad states of the world has a clear effect on whether or not there is a statistically significant difference between the multipliers estimated for the good state and the bad state. A Monte Carlo exercise seeks to understand the consequences of how the time trend is specified, and in general it appears that deterministic time trends lead to more extreme values even when the true data generating process is stationary around a deterministic trend, although estimated one's regressions in first differences do not solve these problems completely.

This paper shows that using aggregate time series methods to answer this question leads to serious questions about the robustness of the results to different specifications. Such a finding argues for using more disaggregated data, such as at the state, county, or even individual household or firm level so as to draw sharper conclusions with simpler, more robust techniques. This is the tack taken by such studies as Nakamura and Steinsson (2014), Fishback, Horrace, and Kantor (2005), or Cohen, Coval, and Malloy (2011). Future research may do well to continue along these lines, making use of more data observations and more plausibly exogenous spending shocks so as to arrive at more precise estimates. Even if aggregate data is used, model averaging methods, such as those employed by Durlauf, Navarro, and Rivers (2014), may need to be employed.

Table 4.1: Summary Statistics for State-Dependent Multiplier Estimates

Horizon	State	Mean	Median	Std. Dev.	Max	Min	90th Pct	10th Pct	Skewness	Kurtosis	Normality P-Value
2 Years	Good	-7.84	0.51	349.74	9443.35	-3676.10	5.34	-2.24	17.71	525.37	< 0.001
2 Years	Bad	-8.81	0.56	353.94	9624.62	-3868.89	5.87	-4.03	17.85	540.64	< 0.001
4 Years	Good	-0.54	0.68	345.43	5889.49	-4585.90	7.88	-0.54	3.96	139.90	< 0.001
4 Years	Bad	-1.20	0.76	349.45	5865.37	-4587.39	10.70	-2.84	3.68	135.19	< 0.001

The table gives summary statistics for all multiplier estimates. Each set includes 1920 estimates.

Table 4.2: Do Some Specification Choices Lead to “Extreme” Estimates?

	Extreme (High or Low)	Extremely High	Extremely Low
Choice			
Deterministic Trend	0.323***	0.146***	0.177***
	(0.017)	(0.013)	(0.013)
SVAR Identification	−0.223***	−0.098***	−0.124***
	(0.015)	(0.012)	(0.011)
SVAR Identification (w/ Ramey news variable)	0.117***	0.056***	0.061***
	(0.020)	(0.016)	(0.016)
Ramey news variable	0.105***	0.042***	0.063***
	(0.020)	(0.015)	(0.016)
Monetary Policy	0.004	0.013	−0.008
	(0.018)	(0.014)	(0.014)
Exogenous Threshold	0.023	−0.027**	0.050***
	(0.018)	(0.014)	(0.014)
CBO Output Gap	0.042	0.026	0.015
	(0.045)	(0.034)	(0.033)
Beveridge-Nelson Output Gap	−0.042	−0.059***	0.016
	(0.028)	(0.016)	(0.024)
Unobserved Component Output Gap	0.027	0.028	−0.001
	(0.032)	(0.025)	(0.023)
Hamilton Output Gap	−0.031	−0.047***	0.016
	(0.029)	(0.018)	(0.024)
Bounceback Output Gap	0.009	0.039	−0.030
	(0.031)	(0.026)	(0.020)
Nonlinear Unobserved Component Output Gap	0.009	0.039	−0.030
	(0.031)	(0.026)	(0.020)
Model Averaged Output Gap	−0.035	−0.083***	0.048
	(0.039)	(0.016)	(0.037)

The table gives coefficients from a linear probability model regressing an indicator for the two year good state multiplier being extreme in one of the ways described in the column headings on a dummy indicator taking on a value of one when the specification choice indicated by the row heading was used in computing that multiplier. Heteroskedasticity and autocorrelation robust standard errors are in parentheses. ***, **, and * indicate significance at the one, five, and ten percent levels, respectively.

Table 4.3: Do Some Specification Choices Lead to "Extreme" Estimates? (ctd.)

	Extreme (High or Low)	Extremely High	Extremely Low
Choice			
AAA-10 Year Spread	0.096***	−0.013	0.109***
	(0.034)	(0.022)	(0.030)
BAA-10 Year Spread	0.003	0.034	−0.030
	(0.031)	(0.025)	(0.020)
BAA-AAA Spread	−0.037	−0.007	−0.030
	(0.029)	(0.022)	(0.020)
Unemployment Rate	−0.037	0.016	−0.053***
	(0.029)	(0.024)	(0.018)
Long Sample	0.038**	0.142***	−0.104***
	(0.018)	(0.014)	(0.013)
Jordà IRF Estimation	−0.392***	−0.196***	−0.196***
	(0.016)	(0.013)	(0.013)
Parsimonious Specification	0.008	0.006	0.002
	(0.018)	(0.014)	(0.014)

The table gives coefficients from a linear probability model regressing an indicator for the two year good state multiplier being extreme in one of the ways described in the column headings on a dummy indicator taking on a value of one when the specification choice indicated by the row heading was used in computing that multiplier. Heteroskedasticity and autocorrelation robust standard errors are in parentheses. ***, **, and * indicate significance at the one, five, and ten percent levels, respectively.

Table 4.4: Summary Statistics for State-Dependent Multiplier Estimates (Excluding Extreme Values)

Horizon	State	Mean	Median	Std. Dev.	Max	Min	90th Pct	10th Pct	Skewness	Kurtosis	Normality P-Value
2 Years	Good	1.02	0.51	1.32	5.34	−2.21	3.24	−0.10	1.28	3.84	< 0.001
2 Years	Bad	1.16	0.56	1.55	5.74	−3.97	3.64	−0.28	0.75	3.13	< 0.001
4 Years	Good	1.32	0.68	1.51	7.74	−0.53	3.40	0.12	1.80	6.29	< 0.001
4 Years	Bad	1.44	0.76	2.04	10.68	−2.84	4.05	−0.52	1.43	6.20	< 0.001

The table gives summary statistics for all multiplier estimates. Each set includes 1536 estimates.

Table 4.5: Specification Breakdown After Truncation of Multiplier Distribution

	Two Years; Good	Two Years; Bad	Four Years; Good	Four Years; Bad
Choice				
Deterministic Trend	0.40	0.40	0.40	0.40
Stochastic Trend	0.60	0.60	0.60	0.60
SVAR Identification (w/o Ramey News Variable)	0.40	0.40	0.39	0.40
SVAR Identification (w/ Ramey News Variable)	0.30	0.31	0.30	0.31
Ramey News Variable	0.30	0.29	0.31	0.29
Monetary Policy	0.50	0.50	0.50	0.50
No Monetary Policy	0.50	0.50	0.50	0.50
Exogenous Threshold	0.49	0.49	0.49	0.50
Endogenous Threshold	0.51	0.51	0.51	0.50
CBO Output Gap	0.05	0.05	0.05	0.05
Beveridge-Nelson Output Gap	0.10	0.10	0.10	0.10
Unobserved Component Output Gap	0.10	0.10	0.10	0.10
Hamilton Output Gap	0.10	0.10	0.10	0.10
Bounceback Output Gap	0.10	0.10	0.10	0.10
Nonlinear Unobserved Component Output Gap	0.10	0.10	0.10	0.10
Model Averaged Output Gap	0.05	0.05	0.05	0.05
AAA-10 Year Spread	0.09	0.09	0.09	0.09
BAA-10 Year Spread	0.10	0.10	0.10	0.10
BAA-AAA Spread	0.10	0.10	0.10	0.10
Unemployment Rate	0.10	0.10	0.10	0.10
Long Sample	0.44	0.43	0.46	0.46
Short Sample	0.56	0.57	0.54	0.54
Jordà IRF Estimation	0.62	0.62	0.60	0.59
GIRF Estimation	0.38	0.38	0.40	0.41
Parsimonious Specification	0.50	0.50	0.51	0.51
Expanded Specification	0.50	0.50	0.49	0.49

The table gives the relative frequency of multipliers computed with each given specification choice after dropping extreme value multipliers. Relative frequencies may not add to 1 due to rounding.

Table 4.6: Effect on Multiplier Estimates (Excluding Extreme Values)

Horizon	2	2	4	4	2	2	4	4
State	Good	Bad	Good	Bad	Good	Bad	Good	Bad
Regression	Univ.	Univ.	Univ.	Univ.	Multiv.	Multiv.	Multiv.	Multiv.
Choice								
Deterministic Trend	0.023	−0.089	0.120	−0.035	0.572***	0.519***	0.652***	0.573***
	(0.071)	(0.082)	(0.082)	(0.110)	(0.051)	(0.063)	(0.064)	(0.093)
SVAR Identification	0.044	−0.126	−0.141*	−0.348***				
	(0.071)	(0.081)	(0.079)	(0.103)				
SVAR Identification (w/ Ramey news variable)	−0.272***	0.067	0.018	−0.040	0.189***	0.539***	0.528***	0.627***
	(0.070)	(0.086)	(0.087)	(0.109)	(0.059)	(0.072)	(0.074)	(0.092)
Ramey news variable	0.221***	0.077	0.140*	0.447***	0.497***	0.583***	0.582***	0.946***
	(0.070)	(0.086)	(0.081)	(0.125)	(0.057)	(0.075)	(0.065)	(0.112)
Monetary Policy	−0.030	−0.032	−0.040	−0.119	−0.021	−0.019	−0.058	−0.124
	(0.068)	(0.079)	(0.077)	(0.104)	(0.048)	(0.061)	(0.056)	(0.081)
Exogenous Threshold	−0.034	0.037	0.099	−0.005	0.004	0.083	0.119**	0.088
	(0.068)	(0.079)	(0.077)	(0.104)	(0.048)	(0.061)	(0.056)	(0.083)
CBO Output Gap	−0.477***	−0.635***	−0.384***	−1.056***	−0.206	−0.358**	0.203	−0.348*
	(0.139)	(0.160)	(0.141)	(0.207)	(0.157)	(0.157)	(0.163)	(0.208)
Beveridge-Nelson Output Gap	0.008	−0.052	−0.074	−0.091	−0.075	0.087	0.044	0.193
	(0.109)	(0.128)	(0.107)	(0.186)	(0.097)	(0.125)	(0.107)	(0.164)
Unobserved Component Output Gap	0.229*	−0.031	0.379**	0.052	0.216**	0.048	0.503***	0.197
	(0.130)	(0.131)	(0.156)	(0.154)	(0.105)	(0.112)	(0.122)	(0.156)
Hamilton Output Gap	0.120	−0.094	−0.002	−0.187	0.046	0.066	0.110	0.124
	(0.119)	(0.127)	(0.112)	(0.175)	(0.102)	(0.112)	(0.105)	(0.156)
Bounceback Output Gap	0.038	−0.012	0.026	−0.060	0.036	0.074	0.077	0.054
	(0.099)	(0.135)	(0.124)	(0.165)	(0.087)	(0.120)	(0.117)	(0.174)
Nonlinear Unobserved Component Output Gap	0.076	−0.038	0.255*	−0.046	0.057	0.058	0.341***	0.145
	(0.116)	(0.109)	(0.146)	(0.148)	(0.099)	(0.101)	(0.124)	(0.143)
Model Averaged Output Gap	−0.137	−0.263	−0.112	−0.946***	0.033	−0.068	0.293**	−0.293*
	(0.138)	(0.185)	(0.122)	(0.207)	(0.125)	(0.145)	(0.145)	(0.175)

The table gives coefficients from a regression of the multiplier estimate at the time horizon and for the state of the economy given in the column heading on a dummy indicator taking on a value of one when the specification choice indicated by the row heading was used in computing that multiplier. Heteroskedasticity and autocorrelation robust standard errors are in parentheses. ***, **, and * indicate significance at the one, five, and ten percent levels, respectively.

Table 4.7: Effect on Multiplier Estimates (Excluding Extreme Values, ctd.)

Horizon	2	2	4	4	2	2	4	4
State	Good	Bad	Good	Bad	Good	Bad	Good	Bad
Regression	Univ.	Univ.	Univ.	Univ.	Multiv.	Multiv.	Multiv.	Multiv.
Choice								
AAA-10 Year Spread	−0.143	0.332*	0.181	0.699***	−0.001	0.548***	0.385***	0.899***
	(0.117)	(0.183)	(0.158)	(0.228)	(0.108)	(0.193)	(0.128)	(0.201)
BAA-10 Year Spread	−0.052	0.198	−0.033	0.563***	−0.048	0.378***	0.063	0.749***
	(0.113)	(0.130)	(0.142)	(0.195)	(0.091)	(0.137)	(0.114)	(0.195)
BAA-AAA Spread	−0.050	0.300**	−0.300***	0.431***	−0.113	0.317***	−0.169*	0.535***
	(0.109)	(0.126)	(0.116)	(0.164)	(0.091)	(0.121)	(0.102)	(0.174)
Unemployment Rate	0.076	−0.085	−0.145	−0.232*				
	(0.105)	(0.115)	(0.110)	(0.127)				
Long Sample	0.423***	0.176**	0.571***	0.802***	0.463***	0.209***	0.578***	0.699***
	(0.067)	(0.078)	(0.078)	(0.104)	(0.050)	(0.065)	(0.058)	(0.090)
Jordà IRF Estimation	−1.733***	−1.883***	−1.879***	−2.198***	−1.954***	−2.132***	−2.125***	−2.441***
	(0.066)	(0.075)	(0.074)	(0.097)	(0.060)	(0.071)	(0.069)	(0.095)
Parsimonious Specification	−0.119*	−0.062	0.087	0.161	−0.094*	−0.053	0.034	0.089
	(0.067)	(0.079)	(0.077)	(0.104)	(0.048)	(0.061)	(0.055)	(0.080)

The table gives coefficients from a regression of the multiplier estimate at the time horizon and for the state of the economy given in the column heading on a dummy indicator taking on a value of one when the specification choice indicated by the row heading was used in computing that multiplier. Heteroskedasticity and autocorrelation robust standard errors are in parentheses. ***, **, and * indicate significance at the one, five, and ten percent levels, respectively.

Table 4.8: Effect on Multiplier Estimates (Excluding Extreme Values)

Horizon	2	2	4	4	2	2	4	4
State	Good	Bad	Good	Bad	Good	Bad	Good	Bad
Regression	> 1	> 1	> 1	> 1	< 0	< 0	< 0	< 0
Choice								
Deterministic Trend	0.258***	0.251***	0.294***	0.283***	0.037***	0.018	0.028***	0.065***
	(0.014)	(0.017)	(0.016)	(0.017)	(0.010)	(0.013)	(0.007)	(0.015)
SVAR Identification (w/ Ramey news variable)	−0.095***	−0.083***	−0.060***	−0.076***	0.049***	0.020	0.005	0.022
	(0.014)	(0.019)	(0.017)	(0.019)	(0.013)	(0.014)	(0.010)	(0.017)
Ramey news variable	−0.075***	−0.095***	−0.038**	−0.065***	−0.027***	−0.025*	−0.022***	−0.006
	(0.015)	(0.019)	(0.018)	(0.020)	(0.008)	(0.013)	(0.007)	(0.017)
Monetary Policy	−0.003	0.004	−0.007	0.000	0.015*	0.031***	0.032***	0.029***
	(0.013)	(0.016)	(0.015)	(0.016)	(0.009)	(0.011)	(0.007)	(0.013)
Exogenous Threshold	−0.006	−0.009	0.003	−0.021	0.020**	−0.062***	0.011	−0.054***
	(0.013)	(0.016)	(0.015)	(0.016)	(0.009)	(0.011)	(0.007)	(0.013)
CBO Output Gap	−0.006	−0.078**	0.112**	−0.081**	−0.007	0.106**	−0.035***	0.184***
	(0.029)	(0.033)	(0.046)	(0.037)	(0.025)	(0.045)	(0.011)	(0.052)
Beveridge-Nelson Output Gap	−0.009	−0.014	0.022	−0.029	0.025	0.035	−0.011	0.093***
	(0.025)	(0.032)	(0.027)	(0.034)	(0.017)	(0.023)	(0.009)	(0.028)
Unobserved Component Output Gap	0.095***	0.013	0.133***	0.000	0.000	0.031	−0.012	0.014
	(0.030)	(0.032)	(0.032)	(0.035)	(0.014)	(0.021)	(0.009)	(0.022)
Hamilton Output Gap	0.003	−0.031	0.023	−0.053*	0.019	0.035	−0.011	0.088***
	(0.025)	(0.029)	(0.027)	(0.031)	(0.016)	(0.023)	(0.009)	(0.028)
Bounceback Output Gap	0.017	−0.016	0.056*	−0.053	−0.002	0.037*	0.002	0.056**
	(0.026)	(0.032)	(0.031)	(0.033)	(0.013)	(0.022)	(0.013)	(0.025)
Nonlinear Unobserved Component Output Gap	0.082***	0.006	0.113***	0.006	−0.007	−0.007	−0.004	0.019
	(0.030)	(0.032)	(0.032)	(0.035)	(0.012)	(0.017)	(0.011)	(0.022)
Model Averaged Output Gap	0.049	−0.050	0.118***	−0.064*	0.027	0.100**	−0.020	0.228***
	(0.036)	(0.035)	(0.046)	(0.038)	(0.030)	(0.041)	(0.016)	(0.049)

The table gives coefficients from a regression of a dummy variable for whether or not the multiplier is significantly greater than 1 (left panel) or below 0 (right panel) at the time horizon and for the state of the economy given in the column heading on a dummy indicator taking on a value of one when the specification choice indicated by the row heading was used in computing that multiplier. Heteroskedasticity and autocorrelation robust standard errors are in parentheses. ***, **, and * indicate significance at the one, five, and ten percent levels, respectively.

Table 4.9: Effect on Multiplier Estimates (Excluding Extreme Values, ctd.)

Horizon	2	2	4	4	2	2	4	4
State	Good	Bad	Good	Bad	Good	Bad	Good	Bad
Regression	> 1	> 1	> 1	> 1	< 0	< 0	< 0	< 0
Choice								
AAA-10 Year Spread	0.028	0.195***	0.097***	0.154***	0.005	−0.003	0.012	0.100***
	(0.026)	(0.045)	(0.031)	(0.044)	(0.016)	(0.021)	(0.016)	(0.049)
BAA-10 Year Spread	0.035	0.136***	0.052*	0.132***	0.084***	−0.014	0.052**	−0.008
	(0.028)	(0.041)	(0.028)	(0.043)	(0.024)	(0.018)	(0.020)	(0.021)
BAA-AAA Spread	0.000	0.038	0.002	0.051	0.063***	−0.011	0.052**	−0.024
	(0.025)	(0.036)	(0.025)	(0.040)	(0.021)	(0.017)	(0.020)	(0.018)
Long Sample	0.049***	−0.025	0.048***	−0.013	−0.070***	−0.075***	−0.043***	−0.124***
	(0.014)	(0.018)	(0.015)	(0.018)	(0.009)	(0.010)	(0.008)	(0.013)
Jordà IRF Estimation	−0.352***	−0.288***	−0.375***	−0.317***	0.051***	0.084***	0.028***	0.132***
	(0.017)	(0.019)	(0.017)	(0.018)	(0.007)	(0.010)	(0.005)	(0.013)
Parsimonious Specification	−0.005	−0.006	−0.030**	−0.005	0.018**	−0.026**	−0.002	0.014
	(0.013)	(0.016)	(0.015)	(0.016)	(0.009)	(0.011)	(0.007)	(0.013)

The table gives coefficients from a regression of a dummy variable for whether or not the multiplier is significantly greater than 1 (left panel) or below 0 (right panel) at the time horizon and for the state of the economy given in the column heading on a dummy indicator taking on a value of one when the specification choice indicated by the row heading was used in computing that multiplier. Heteroskedasticity and autocorrelation robust standard errors are in parentheses. ***, **, and * indicate significance at the one, five, and ten percent levels, respectively.

Table 4.10: Summary Statistics for Difference Between State Dependent Multipliers

Horizon	Extreme Values Included?	Mean	Median	Std. Dev.	Max	Min	90th Pct	10th Pct	Skewness	Kurtosis
2 Years	Yes	−0.968	0.002	46.933	181.273	−2018.992	1.535	−0.944	−41.459	1782.697
2 Years	No	0.191	0.030	0.954	5.653	−5.774	1.467	−0.702	0.663	8.784
4 Years	Yes	−0.667	0.000	26.203	582.511	−707.087	1.556	−1.872	−8.116	464.373
4 Years	No	0.273	0.069	0.945	4.312	−4.909	1.502	−0.626	0.875	7.108

The table gives summary statistics the difference between good and bad state multiplier estimates. Positive values indicate that the bad state multiplier is larger. When extreme values are included, there are 1920 pairs of multipliers at each horizon. When extreme values are excluded, there are 1385 multiplier pairs at the two year horizon and 1253 at the four year horizon.

Table 4.11: Effect on Difference between Good and Bad State Multiplier Estimates (Excluding Extreme Values)

	2 Years	4 Years
Deterministic Trend	-0.085^{*}	-0.050
	(0.048)	(0.051)
SVAR Identification (w/ Ramey news variable)	0.295^{***}	0.378^{***}
	(0.059)	(0.063)
Ramey news variable	0.230^{***}	0.287^{***}
	(0.061)	(0.060)
Monetary Policy	0.033	0.026
	(0.048)	(0.049)
Exogenous Threshold	-0.020	-0.096^{*}
	(0.048)	(0.050)
CBO Output Gap	-0.241^{*}	-0.185
	(0.129)	(0.140)
Beveridge-Nelson Output Gap	0.235^{***}	0.221^{**}
	(0.091)	(0.089)
Unobserved Component Output Gap	-0.182^{**}	-0.199^{**}
	(0.078)	(0.093)
Hamilton Output Gap	0.100	0.111
	(0.097)	(0.101)
Bounceback Output Gap	0.011	-0.022
	(0.089)	(0.096)
Nonlinear Unobserved Component Output Gap	-0.171^{**}	-0.187^{***}
	(0.076)	(0.092)
Model Averaged Output Gap	0.023	0.045
	(0.129)	(0.141)
AAA-10 Year Spread	0.809^{***}	0.726^{***}
	(0.143)	(0.147)
BAA-10 Year Spread	0.551^{***}	0.528^{***}
	(0.092)	(0.105)
BAA-AAA Spread	0.424^{***}	0.472^{***}
	(0.072)	(0.087)
Long Sample	-0.306^{***}	-0.316^{***}
	(0.052)	(0.059)
Jordà IRF Estimation	-0.079^{*}	0.024
	(0.048)	(0.051)
Parsimonious Specification	-0.029	-0.067
	(0.047)	(0.049)

The table gives coefficients from a regression of the difference between the good and bad state multipliers for the horizon in the column heading on a dummy indicator taking on a value of one when the specification choice indicated by the row heading was used in computing that multiplier. Heteroskedasticity and autocorrelation robust standard errors are in parentheses. ***, **, and * indicate significance at the one, five, and ten percent levels, respectively.

Table 4.12: Coefficient Estimates for DGPs for Monte Carlo Simulation

	Deterministic		Stochastic	
Independent Variable	GDP	Govt Exp.	GDP	Govt Exp.
Constant	0.5795^{***}	-0.0764	0.0037^{***}	0.0018^{*}
	(0.178)	(0.244)	(0.001)	(0.001)
GDP_{t-1}	1.3076^{***}	0.0599	0.3575^{***}	0.0730
	(0.064)	(0.088)	(0.064)	(0.086)
GDP_{t-2}	-0.2594^{**}	-0.0397	0.0892	-0.0208
	(0.106)	(0.145)	(0.069)	(0.093)
GDP_{t-3}	-0.1394	-0.0394	-0.0183	-0.0697
	(0.109)	(0.150)	(0.070)	(0.094)
GDP_{t-4}	0.0331	0.0478	-0.1245^{*}	0.1981^{**}
	(0.067)	(0.092)	(0.065)	(0.087)
$GovtExp_{t-1}$	-0.1107^{**}	1.0961^{***}	-0.1279^{***}	0.1044^{*}
	(0.046)	(0.063)	(0.046)	(0.062)
$GovtExp_{t-2}$	0.1444^{**}	-0.0744	0.0181	0.0312
	(0.070)	(0.096)	(0.046)	(0.061)
$GovtExp_{t-3}$	-0.0550	0.0691	-0.0521	0.0760
	(0.070)	(0.096)	(0.044)	(0.059)
$GovtExp_{t-4}$	0.0172	-0.1152^{**}	0.0477	0.0644
	(0.040)	(0.055)	(0.040)	(0.053)
$Milnews_{t}$	0.0360^{**}	-0.0420^{**}	0.0371^{***}	-0.0364^{*}
	(0.015)	(0.020)	(0.014)	(0.019)
$Milnews_{t-1}$	-0.0395^{**}	0.1000^{***}	-0.0457^{***}	0.1119^{***}
	(0.017)	(0.023)	(0.017)	(0.023)
$Milnews_{t-2}$	0.0331^{*}	0.0795^{***}	0.0426^{**}	0.0773^{***}
	(0.018)	(0.025)	(0.018)	(0.025)
$Milnews_{t-3}$	0.0092	0.0942^{***}	0.0080	0.0990^{***}
	(0.018)	(0.024)	(0.018)	(0.024)
$Milnews_{t-4}$	0.0173	0.0602	0.0293^{*}	0.0612^{***}
	(0.016)	(0.022)	(0.017)	(0.023)
$TimeTrend$	0.0004^{***}	0.0000		
	(0.000)	(0.000)		
$TimeTrend^2$	-0.0000^{**}	-0.0000		
	(0.000)	(0.000)		

The table gives coefficients from regressions of the equation systems given in Equations 4.9 and 4.10. The columns indicate the dependent variables and the trend specifications. ***, **, and * indicate significance at the one, five, and ten percent levels, respectively.

Table 4.13: Distribution of Linear Multiplier Estimates)

DGP	Estimation	5%	10%	25%	50%	75%	90%	95%
Deterministic	Deterministic	−1.36	−0.67	−0.03	0.46	0.89	1.42	2.08
Deterministic	Stochastic	−1.48	−0.59	0.02	0.50	0.97	1.70	2.32
Stochastic	Deterministic	−1.43	−0.52	0.14	0.74	1.53	2.66	4.28
Stochastic	Stochastic	−1.24	−0.42	0.24	0.79	1.42	2.44	3.84

The table gives the multiplier representing the percentile given by the column heading from 500 Monte Carlo simulations generated by the process indicated in the first column and estimated assuming the process was generated as given by the second column.

Table 4.14: Distribution of Nonlinear Multiplier Estimates (Median Unemployment Rate Distinguishes States))

DGP	Estimation	5%	10%	25%	50%	75%	90%	95%
Good State								
Deterministic	Deterministic	−6.36	−2.12	−0.25	0.45	1.16	3.52	7.89
Deterministic	Stochastic	−4.53	−1.51	−0.18	0.50	1.31	3.17	6.01
Stochastic	Deterministic	−7.97	−3.33	−0.24	0.80	2.11	5.20	9.49
Stochastic	Stochastic	−6.86	−2.35	−0.26	0.81	2.03	5.37	10.40
Bad State								
Deterministic	Deterministic	−9.23	−4.98	−0.80	0.96	3.09	8.49	17.10
Deterministic	Stochastic	−5.36	−2.35	−0.40	0.41	1.33	2.68	5.50
Stochastic	Deterministic	−8.69	−3.66	−0.46	0.68	1.81	3.54	5.56
Stochastic	Stochastic	−5.14	−2.12	−0.14	0.80	2.04	4.09	8.20

The table gives the multiplier representing the percentile given by the column heading from 500 Monte Carlo simulations generated by the process indicated in the first column and estimated assuming the process was generated as given by the second column. The "good state" of the economy refers to observations when the lagged unemployment rate is below the median in the sample, while the "bad state" refers to observations when the lagged unemployment rate is above the median in the sample.

Table 4.15: Distribution of Nonlinear Multiplier Estimates (67^{th} Percentile of Unemployment Rate Distinguishes States))

DGP	Estimation	5%	10%	25%	50%	75%	90%	95%
Good State								
Deterministic	Deterministic	−3.20	−1.20	−0.27	0.38	1.01	2.25	5.01
Deterministic	Stochastic	−4.64	−1.40	−0.19	0.38	0.99	2.06	3.27
Stochastic	Deterministic	−3.64	−1.64	−0.18	0.73	1.63	4.12	7.79
Stochastic	Stochastic	−3.94	−1.60	−0.10	0.73	1.63	3.38	5.85
Bad State								
Deterministic	Deterministic	−16.98	−8.03	−2.06	1.00	4.16	11.14	21.65
Deterministic	Stochastic	−7.16	−3.95	−0.70	0.44	2.07	6.25	15.31
Stochastic	Deterministic	−11.87	−4.38	−0.80	0.64	2.32	7.10	14.26
Stochastic	Stochastic	−8.66	−3.71	−0.58	0.76	2.22	6.66	15.27

The table gives the multiplier representing the percentile given by the column heading from 500 Monte Carlo simulations generated by the process indicated in the first column and estimated assuming the process was generated as given by the second column. The "good state" of the economy refers to observations when the lagged unemployment rate is below the 67^{th} percentile in the sample, while the "bad state" refers to observations when the lagged unemployment rate is above the 67^{th} percentile in the sample.

Figure 4.1: Density Plots of State-Dependent Multipliers

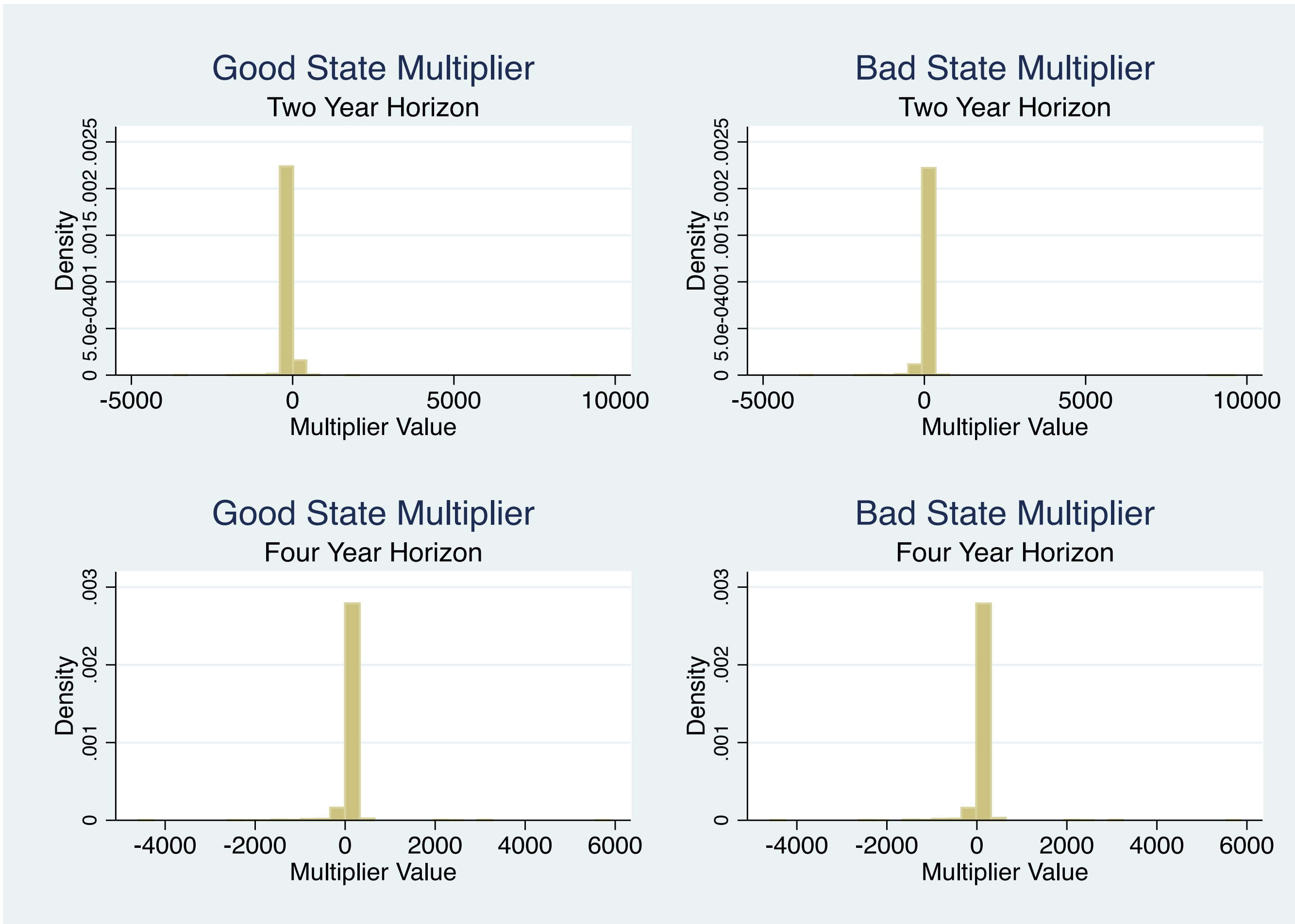


Figure plots density of multiplier estimates for the time horizon and the state of the economy indicated.

Figure 4.2: Density Plots of State-Dependent Multipliers

Figure plots density of multiplier estimates for the time horizon and the state of the economy indicated. Multiplier estimates in the top and bottom ten percentiles of the distribution have been excluded.

Figure 4.3: Monte Carlo Simulation, Linear Specification

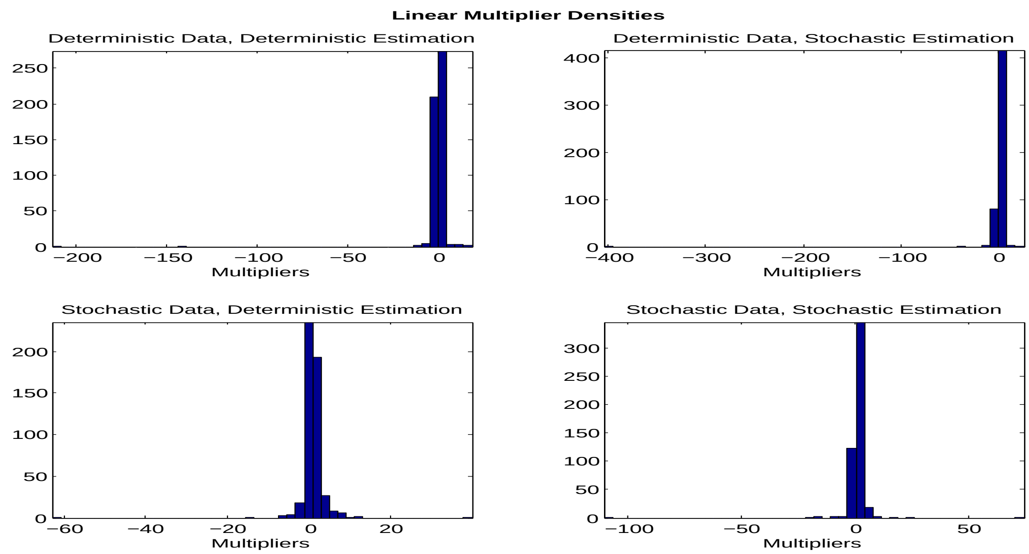


Figure plots density of multiplier estimates (not taking into account state dependence) for the for the data generating process and estimation approach indicated. Multipliers are computed at the two year horizon.

Figure 4.4: Monte Carlo Simulation, Unemployment Rate below Median Level

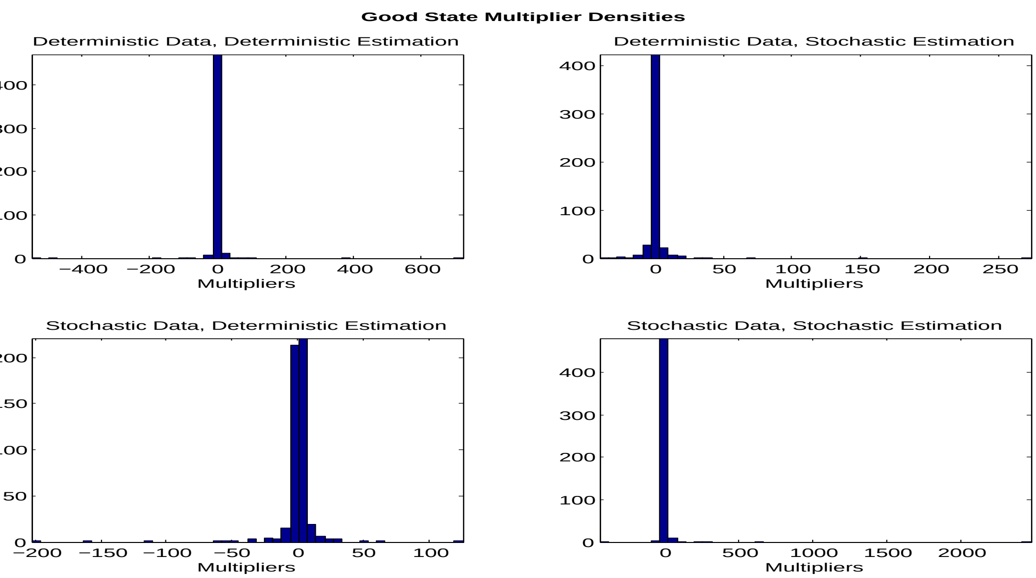


Figure plots density of multiplier estimates (not taking into account state dependence) for the for the data generating process and estimation approach indicated. Multipliers are computed at the two year horizon. The "good state" refers to observations when the unemployment rate is below the median in the sample.

Figure 4.5: Monte Carlo Simulation, Unemployment Rate above Median Level

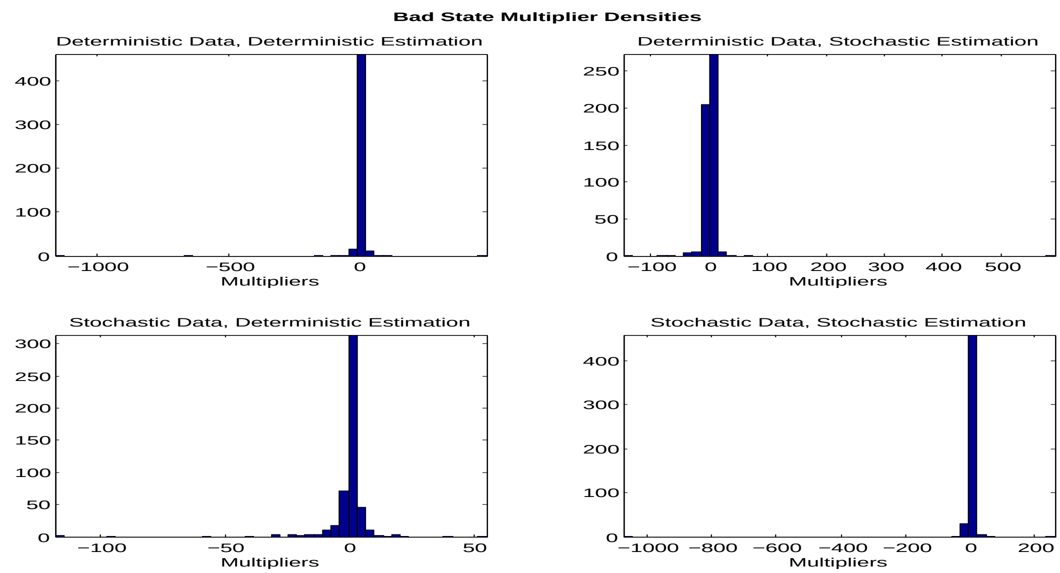


Figure plots density of multiplier estimates (not taking into account state dependence) for the for the data generating process and estimation approach indicated. Multipliers are computed at the two year horizon. The “bad state” refers to observations when the unemployment rate is above the median in the sample.

Figure 4.6: Monte Carlo Simulation, Unemployment Rate below 66th Percentile

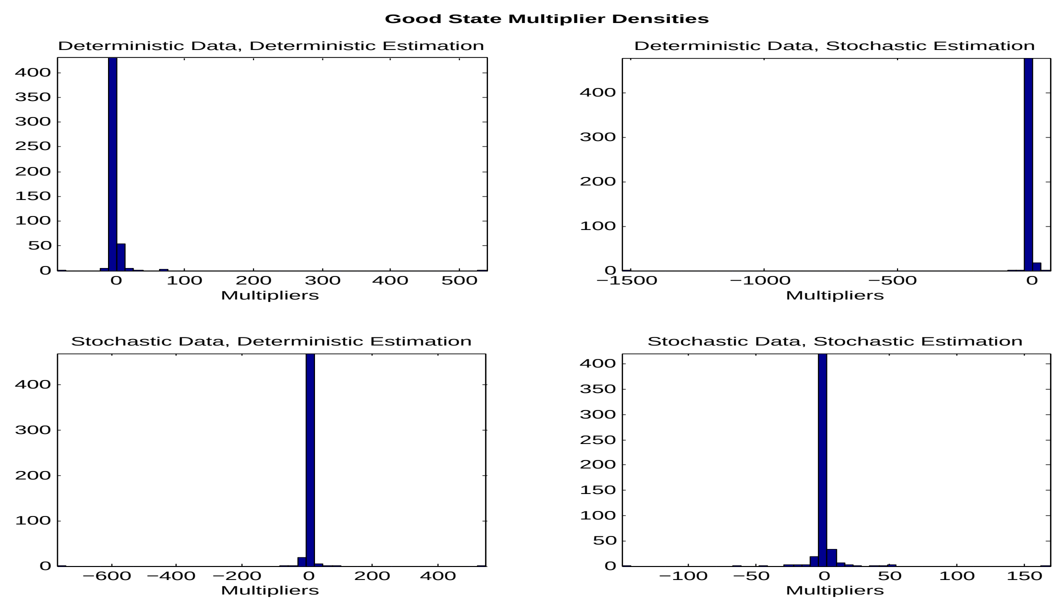


Figure plots density of multiplier estimates (not taking into account state dependence) for the for the data generating process and estimation approach indicated. Multipliers are computed at the two year horizon. The "good state" refers to observations when the unemployment rate is below the 67^{th} percentile in the sample.

Figure 4.7: Monte Carlo Simulation, Unemployment Rate above 66th Percentile

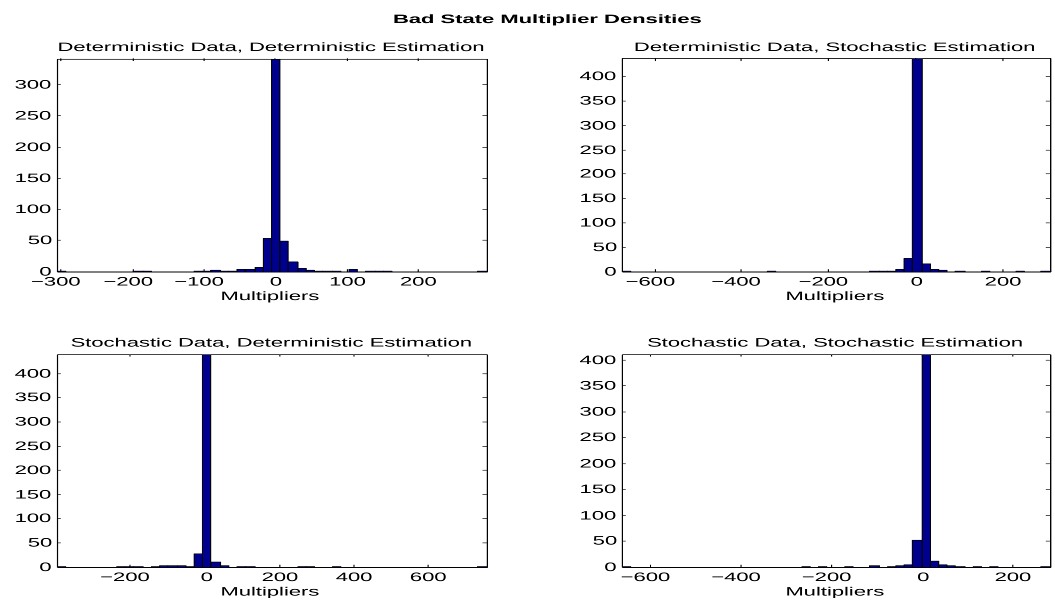


Figure plots density of multiplier estimates (not taking into account state dependence) for the for the data generating process and estimation approach indicated. Multipliers are computed at the two year horizon. The "bad state" refers to observations when the unemployment rate is above the 67^{th} percentile in the sample.

Chapter 5

Local Effects of a Military Spending Shock: Evidence from Shipbuilding in the 1930s and 1940s

5.1 Introduction

What are the effects of government purchases on local economies, especially when the aggregate economy is in a state of weakness? Normally, these effects are summarized in terms of a “multiplier,” defined as the amount of extra output generated by an additional dollar of government purchases. One of the benefits of examining fiscal multipliers at the local level is that one can observe a larger number of regions, who are all subject to the same national monetary policy. This is helpful, because potentially endogenous monetary policy changes can confound the estimation of aggregate government spending multipliers. Knowledge of how purchases affect local economies can also provide insight into the transmission mechanisms of government spending on a broader scale. This is because one might expect purchases to impact the areas in which the funds are directly spent most quickly and powerfully. Spillover effects can also be important, because positive spillovers to neighboring areas are indicative of a large multiplier overall, while negative spillovers to neighboring areas suggest that the government's activity is merely inducing a reallocation of resources from one region to another. Another advantage of tackling this question at the local level is that identification of federal spending shocks might be easier. Especially in a time of war or the threat of war, it may be more plausible that the federal government is not increasing spending solely in response to local area conditions.

This paper fills a gap in the literature on the effects of federal government spending at the local level by exploiting a previously understudied spending episode, the Vinson-Trammell Act of 1934, which aimed to build the United States Navy up to treaty limitations imposed at the end of World War I, and was a response to Japanese naval expansion. Using historical sources, I am able to identify the counties receiving shipbuilding contracts, and I track the evolution of their economies throughout the 1930s. In particular, I examine the responses of manufacturing output, employment, and retail sales, among other outcomes. The timing of this act is fortuitous for me, as the 1930s were a period in which nominal interest rates were pinned to the zero lower bound throughout. This is important, because during this period monetary policy did not react to the fiscal shock with higher interest rates. Christiano, Eichenbaum, and Rebelo (2011) argue that fiscal multipliers are particularly large when there are no changes in interest rates to offset fiscal policy. In addition, Kuhn and George (2014) show that occasionally binding capacity constraints can also produce countercyclicality of the government spending multiplier. As a point of comparison, I also examine the local economies of areas building ships for the government during World War II, when economic capacity was more constrained.

My results show that counties hosting shipyards in 1934 (the year the Vinson-Trammell Act was passed) experienced relatively greater manufacturing output growth and relatively greater retail sales growth in the latter part of the decade, compared to counties not hosting shipyards. In particular, counties that hosted shipyards at the time of the act's passage saw an extra 12-13 percentage points of output growth over the two-year periods from 1935-1937 and 1937-1939, relative to otherwise identical counties that did not host shipyards. In addition, retail sales growth in these counties was 3-4 percentage points higher in the latter half of the 1930s. There is evidence that the naval spending spilled over into neighboring counties, boosting retail sales growth there as well. At the household level, consumers in shipyard counties spent more on consumption goods the more they were exposed to the naval spending. This result holds even when controlling for the household's income, which is consistent with the idea that labor supply and consumption were complements in the utility function. This is important, because Nakamura and Steinsson (2014) demonstrate that such complementarity could be key to generating an aggregate multiplier greater than one. What is more, these effects generally do not hold for the period during World War II, when shipbuilding activity really ramped up and when capacity constraints began to impede the economy. With regard to other economic variables, the spending did not alter education choices on the extensive margin, but seems to have had a negative effect on the resources devoted to schooling. The spending by the government also alters the relative importance of the shipbuilding industry in these counties, compared to other durable goods industries. These effects are robust to controlling for the initial level of industrialization in the counties, the percentage of the area that is classified as "urban," and for New Deal spending at the county level.

Of course, for the purposes of policy, what is of interest is the actual aggregate government spending multiplier. One of the drawbacks of conducting the analysis at the county level is the difficulty of "scaling up" the local results. The results of this paper are consistent with those of Nakamura and Steinsson (2014), who use a multi-region DSGE model to translate their local results to an aggregate multiplier. In the context of their model, the spending shock identified here may have had a multiplier of around 9. Although this figure is huge, it is also worth considering that in the depths of the Great Depression, with a vast amount of unused capacity and an economy relatively closed to trade, it may not be out of the question. I also consider other means of scaling the multiplier, and these give very different results, however.

The rest of the paper proceeds as follows. Section 5.2 contains a brief literature review. In Section 5.3, I discuss my empirical methodology, including background information on the Vinson-Trammell Act of 1934, my method for identifying shipyard locations, the data, and my regression specifications. Results follow in Section

5.4, and this is followed by an attempt to interpret the baseline local results as an aggregate government spending multiplier in Section 5.5. Section 5.6 consludes.

5.2 Literature Review

The empirical literature on the output effects of government purchases has grown rapidly in recent years, as many national and local governments have pursued fiscal stimulus in an attempt to boost flagging economies. Still, there is little consensus on whether this spending has been a net positive or if it has rather crowded out private activity. Most papers have concentrated on aggregate government spending multipliers, with an offshoot of the literature focusing on whether these output effects depend on the condition of the economy in the period when the spending hits. See, for example, Ramey (2011b), Barro and Redlick (2011), Auerbach and Gorodnichenko (2012b), Auerbach and Gorodnichenko (2012a), and Auerbach and Gorodnichenko (2013) among many others. Another substrand of this field has attempted to estimate government spending multipliers at the local level. In the United States, this has included multipliers on spending in the fifty states or at county level. In some studies, the source of the spending is the federal government, while in others, the source of the spending is the state government.

The literature estimating local government spending multipliers has exploited a variety of indentification strategies in a handful of settings, and many have found strong positive effects on local economies, although that is not a uniform conclusion. Serrato and Wingender (2014) use population revisions after decennial Censuses to instrument for federal spending that is a function of county population, and they find a local income multiplier of 1.57. Nakamura and Steinsson (2014) utilize military procurement contracts, which indicate the particular U.S. state at the receiving end of funding, to find an output multiplier of around 1.5. Hooker (1996) undertakes a similar analysis to find that military spending cuts are particularly harmful to state economies. Chodorow-Reich et al. (2012) use expansions in Medicaid funding enacted in 2009 to estimate that for every extra $100,000 in transfers from the federal government to the states, 3.8 job-years are created, including 3.2 in the private sector. Shoag (2010) and Shoag (2013) identifies state government spending shocks generated by windfalls in pension fund returns, and he finds an income multiplier of 2.12 for the years 1987 to 2008 and a multiplier of 1.43 for the period of the Great Recession. Fishback and Kachanovskaya (2010) and Fishback, Horrace, and Kantor (2005) estimate the effect of New Deal grants (instrumented by a number of political variables) on real per capita income at the state level and retail sales per capita at the state and county level. They find an income multiplier just above 1 and a positive impact on retail sales for many types of grants. Acconcia, Corsetti, and Simonelli (2014) use an Italian law which mandates the removal of local councils upon evidence of Mafia infiltration. Such council dismissals were often associated with dramatic declines in public investment, so they use dismissals as an instrument to find a local multiplier of 1.2 and a longer-term multiplier of 1.8 for provinces in Italy in the 1990s. Hausman (2013) takes a slightly different tack, examining the impacts of the unexpected early payment of the 1936 Veterans' Bonus to find that many bonus recipients quickly went out and spent their windfall.

Many papers have, however, produced dissenting views on the local effects of government spending. Fishback and Cullen (2013) demonstrate that World War II spending at the county level did not influence many economic indicators. Of course, there are limitations to data availability in this period, and performing any analysis on the U.S. economy during World War II is necessarily dealing with an economic environment unlike any other in the nation's history. Bruckner and Tuladhar (2014) do not find a local government spending multiplier greater than one when looking at the effects of central government spending in Japanese prefectures during the 1990s, although they do find evidence that different kinds of spending produce different results, as well as stronger effects when local economies are relatively weak. Clemens and Miran (2012), taking advantage of heterogeneity in the stringency of balanced budget requirements of U.S. states, find that the multiplier on investment spending is likely less than one. Finally, Cohen, Coval, and Malloy (2011) use the ascendance of local representatives to powerful positions on influential congressional committees as an instrument for federal earmarks, and they find a significantly negative effect on corporate investment by firms headquartered in those districts. They attribute this negative impact to the government crowding out private activity.

This paper will take cues from a number of different studies. First, like Nakamura and Steinsson (2014), I will estimate the effects of a military spending shock at the local level. To do so, I will make use of a spending program that has not been greatly exploited to this point, the Vinson-Trammell Act of 1934, which generated a significant amount of naval shipbuilding during the Great Depression. The Depression context differentiates my study from that of Nakamura and Steinsson

(2014), who only study the period following World War II. I can also study a lower geographical entity than they can (counties as opposed to states). The Depression context ties my paper to that of Fishback, Horrace, and Kantor (2005), who also look at county-level economic outcomes during the Depression. By focusing on the era of the Great Depression, I am able to offer some insight on the question of whether government spending is more effective at a time of severe economic weakness. I will contrast my results from the Vinson-Trammell shipbuilding activity with those deriving from similar regressions on spending authorized by the United States Maritime Commission (USMC) during World War II, which tended to take place at a time when capacity utilization was higher. Also, this paper obtains some inspiration from Hausman (2013) by seeking information on the effects of this shipbuilding from several different sources, including a county level dataset assembled by Fishback et al. (2011b),[1] IPUMS samples from the 1930 and 1940 Censuses,[2] the Study of Consumer Purchases in the United States, 1935-1936, the County and City Data Book [United States] Consolidated File: County Data, 1947-1977, and a listing of USMC spending at various shipyards scattered across the country provided by Fischer (1946), as well as further information on shipyard locations found in Lane (1951) and contemporary newspaper accounts.

Unlike the work of Chodorow-Reich et al. (2012) or Hausman (2013), I will be looking at the effects of government purchases, not transfers. Unlike Nakamura and Steinsson (2014), the purchases that I study take place solely during a period of

[1] I wish to note that this citation refers to both the paper, "Information and the Impact of Climate and Weather on Mortality Rates During the Great Depression," as well as the associated data set, "Weather, Demography, Economy, and the New Deal at the County Level, 1930-1940."

[2] Ruggles et al. (2010).

severe economic weakness. Unlike Fishback and Kachanovskaya (2010), I will be able to delve to the county level, as opposed to the state level. Unlike Clemens and Miran (2012) and Acconcia, Corsetti, and Simonelli (2014), I deal with increases in government purchases, as opposed to cuts to government investment. Although my paper is not the first to exploit military shipbuilding in this era (see, for example, Thornton and Thompson (2001)), I am not aware of any others that explore its wider effects.

5.3 Empirical Methodology

This section will outline the process for identifying a military spending shock and estimating its local effects. First, I will describe the Vinson-Trammell Act of 1934. Then, I will demonstrate how I identify shipyard (and neighboring) counties. The third subsection will provide information on the various sets of data that I will employ, and the fourth section will detail the regression specifications.

5.3.1 The Vinson-Trammell Act of 1934

The Washington Naval Treaty was signed in 1922 by representatives of the United States, the United Kingdom, France, Italy, and Japan.[3] The aim of the treaty was to prevent the sort of arms race that was believed to have contributed to the outbreak of the First World War nearly a decade earlier. The treaty placed limits on the amount

[3]The source for much of the information contained in this subsection is Cook (2004), Chapters 3 and 5.

of tonnage that the signatories' navies could employ, as well as limits on the types of weapons that could be carried on naval vessels. In addition, many shipbuilding programs that were underway in these countries were to be halted and scrapped. The stipulations of the agreement were extended and reinforced in the London Naval Treaty signed by the same five powers in 1930.

Throughout the 1920s and the early part of the 1930s, the United States Navy did not build up to its treaty allowance. Cook (2004) reports that in the ten years after the initial Washington Naval Treaty, the United States had built more than a hundred fewer ships than any of the other signatories and a total of zero destroyers. This inactivity was due partly to greater isolationist and pacifist sentiment and partly to a lack of political will. President Hoover, for example, staunchly opposed naval expansion. This was not the case in Japan, which had built its fleet up quickly with more modern, capable ships. Some in the policy-making establishment, such as Carl Vinson, a U.S. Senator from Georgia, had begun to get nervous about Japanese intentions and started to agitate for increased naval spending. In late 1931, Japan invaded Manchuria, in clear violation of several treaties it had signed, and in 1933, it announced plans to increase spending on its navy by 25%. These concerns convinced President Roosevelt "that a longterm building program was essential if the navy were to keep pace with Japan." (Cook (2004), p. 87).

What would become the Vinson-Trammell Act of 1934[4] was introduced by Senator Vinson in January of 1934 and passed Congress on 20 March 1934, to be signed by President Roosevelt a week later. The bill authorized the government to build the

[4]Senator Park Trammell of Florida had authored a competing bill that he eventually dropped to support Senator Vinson's.

Navy up to the country's treaty allowances. The passage of the Vinson-Trammell Act also raised expectations of future government spending, as it is listed as an exogenous spending news shock equal to about 1.5% of GDP in the series constructed by Owyang, Ramey, and Zubairy (2013) and Ramey and Zubairy (2014).[5] Although there had been some appropriations to naval shipbuilding made as part of the 1932 National Industrial Recovery Act (which appropriations were also pushed by Senator Vinson with eyes focused on the emerging Japanese threat), the spending that was anticipated as a result of this bill was much larger. Also, unlike the 1932 bill, the motivation behind the Vinson-Trammell Act was not economic revitalization.

Opposition to the passage of the bill came mainly from pacifists, who argued that the supposed Japanese threat was an illusion manufactured by shipbuilders so as to obtain government contracts. Cook (2004, Chapter 5) offers some specific examples. Indeed, if that was the case, it would threaten the exogeneity of this spending. Senator Vinson, the main proponent of the bill seems not to have believed in this notion. He had been involved in a special audit into aircraft manufacturers that examined whether they had made "excessive profits" from 1927 to 1933. No evidence was ultimately found, but the senator was concerned enough to push for a more formal investigation. On the possibility of the government being exploited by private firms, Senator Vinson said, "We are not going to stand by and let the Government be at the mercy of any private company; we are not going to be held up. If they're making too much, we'll put a stop to it," (Cook (2004), p. 96). In fact, the bill included a provision limiting profits on shipbuilding contracts to ten

[5]The passage of the bill is not explicitly mentioned in either of these papers, but its inclusion is indicated in the narrative of the data series available on Valerie Ramey's website.

percent (Lane (1951), p.798). It seems, then, unlikely that this spending program was implemented so as to benefit shipbuilding firms.

Figure 5.1 shows annual real defense spending for every year from 1929 to 1940. One can see a distinct jump that occurs in Fiscal Year 1935 (the first year for which the Vinson-Trammell spending would be occurring). Fiscal 1935 saw a real increase in defense spending of 26.7% compared to a year earlier. The following year saw a further 14.2% increase. Average spending for the years from 1935 to 1939, before the big spending shocks that would be associated with World War II, was 17.81 billion 2009 dollars, a 38.2% increase on average spending for the years from 1929 to 1934, the years before the implementation of the Vinson-Trammell Act. A visual examination of the time series data for real defense consumption, real nondefense consumption and current tax receipts suggests that much of this extra spending for ships was initially financed by allocating funds away from nondefense items and eventually by higher tax revenues. This interpretation is supported by calculations of the average marginal tax rate by Barro and Redlick (2011), who show that taxes were raised in 1934, 1935, and 1936. In any event, because these purchases were financed at the federal level, one should expect that the shock should be interpreted as a windfall for the counties hosting shipyards. While the shipyard counties receive the entirety of the spending, they bear the cost relatively equally compared to other counties in the country, whether the purchases are financed by a reallocation of resources or higher taxes.

The first shipbuilding contracts awarded in conjunction with the bill were placed in August of 1934. According to *The New York Times*, "plans have nearly all been completed so that work can start, not only in private but in government yards, within a reasonable time."[6] The kinds of ships that Senator Vinson envisioned being constructed required about three years for completion (Cook (2004), p. 96), so one might expect that the spending beginning in 1935 and extending into 1936 would have effects until the end of the decade.

5.3.2 Shipyard Locations

Identifying the locations of shipyards active at the time of the bill's passage in 1934 will be key to understanding the effects of the spending. Although there is evidence of further yards opening in the latter half of the 1930s, I exclude these from my baseline analysis because of concerns that their opening may have been endogenous to the spending. The central assumption that I will make in my empirical analysis is that the counties that received the Vinson-Trammell spending did so because of pre-existing shipbuilding facilities and not because of any other local economic conditions.

My primary source for identifying shipbuilding locations around the country is the fifth part of the first chapter of Lane (1951). Further information on yard locations comes from contemporaneous newspaper sources, such as the article from *The New York Times* referenced in Footnote 9. Figure 5.2 displays the geographical county locations of the identified shipyards, and Table 5.1 gives a listing of the counties. The identified shipbuilding centers also include counties hosting major steel producing

[6] *The New York Times*, "Awards Contracts for 24 Warships," 23 August 1934.

facilities owned by the Bethlehem Steel Company, which also owned several shipyards. The locations of these facilities are also obtained from Lane (1951). I include them on the presumption that any economic benefit as a result of this spending shock accruing to counties hosting Bethlehem shipyards would also be experienced by counties hosting the steel facilities supplying them. This is the reason for several inland counties in Pennsylvania being included in the list of shipyard counties.

Since I will also estimate whether the supposed economic benefits spilled over into neighboring counties, for each identified shipyard county, I gather a list of counties bordering it or that have strong economic links to it, as defined by the 1991 Contiguous County File.[7]

From examining the list of shipyard counties and eyeing the associated map, it is the case that the shipyard counties cluster around urban areas, particularly in the northeastern part of the country. Cities such as New York, Boston, Philadelphia, Baltimore, Los Angeles, San Francisco, and Seattle are included,although other major cities such as Chicago, Detroit, and St. Louis are not. Admittedly, this poses a concern with regard to whether the effects that I will pick up are not rather due to, for example, relatively faster growth in urban areas. I will attempt to demonstrate that this is not the case by controlling for the percentage of each county that is urban, as well as state fixed effects. On the other hand, none of the shipyards that were open at the time of the bill's passage were located in Georgia or Florida, the home states of the senators for which it is named, which relieves any concern about spending being allocated for politically motivated reasons.

[7]ICPSR Data Set 9835, United States Department of Commerce, Bureau of the Census (1992).

As a comparison, I will also be examining the effects on counties that hosted USMC shipbuilding activity in the lead-up to and during World War II. I have detailed data on ships purchased by the USMC from various yards around the country between 1936 and 1946 from Fischer (1946). I will describe this data source in greater detail in the next subsection, but at this point, I note that the Fischer (1946) document lists 62 shipyards from which the USMC purchased ships. Figure 5.3 displays the locations of these yards, and Table 5.2 lists the affected counties. One will note that there is much greater geographical heterogeneity in this group of shipyard locations compared to the list of yards active earlier. As with the shipyards active by 1934, counties that border those receiving USMC contracts are identified with the help of the 1991 Contiguous County File.

5.3.3 Data

The data used to conduct the analysis in this paper comes from a number of different sources. The primary dataset is that of Fishback et al. (2011b). This dataset is an annual county level panel that covers the years from 1930 to 1940. It includes a large number of variables, of which I will make use of a smaller subset. The data set includes information on county population in 1930 and 1940 (as well as linearly interpolated figures for the intervening years). It also has information on the number of manufacturing establishments, along with the average number of employees at each establishment, manufacturing output and value added, and wage payments to manufacturing workers and average earnings. This manufacturing data is available for the years 1931, 1933, 1935, 1937, and 1939. It also includes retail sales data for the years 1933, 1935, and 1939. The manufacturing data I will use as a county level proxy for output, while the retail sales data will stand in for consumption. The data set has variables for retail and wholesale employment, wholesale net sales, and average retail and wholesale earnings for the years 1935 and 1939. Also included is the number of automobile registrations for the years 1930, 1931, and 1936. Finally, it has information on the number of tax returns filed in each county for every year in the sample.

In addition to these series, which will provide the bulk of the outcomes I consider in the analysis, this data identifies the percentage of each county which is "urban," and has an indicator for whether each individual county is located on the Great Lakes or the Atlantic, Pacific, or Gulf coasts. These will help me to control for the urban character of each county as well as whether or not it depends greatly on maritime industries. It also has an interpolated series of New Deal spending for each county. That is, there is information on total New Deal spending over the course of 1933 to 1939 in each county, and this sum is interpolated into an annual time series using information on New Deal grants at the state level.

Similar to this data set from Fishback et al. (2011b) is the County and City Data Book [United States] Consolidated File: County Data, 1947-1977.[8] This second set includes many of the same variables that are found in the above-mentioned dataset including the number of manufacturing establishments, manufacturing output, manufacturing value added, retail and wholsale sales, retail and wholesale employment,

[8]ICPSR Data Set 7736, United States Department of Commerce, Bureau of the Census (2012).

and automobile registrations. The years covered by this set are those after the Second World War (all of the variables listed have an observation between 1947 and 1949), so the value of this information is to see if the USMC shipbuilding activity benefited recipient counties relative to counties without shipbuilding facilities during the war. Although it is perilous to estimate effects spanning World War II due to the unique circumstances surrounding the wartime economy, such as rationing and greater government intervention, it may be useful to compare the effects of government ship purchases in a time of relatively constrained capacity to those when the economy is in the midst of a depression.

The third source of data to be employed is the Study of Consumer Purchases in the United States, 1935-1936,[9] which was also featured in Hausman (2013). That paper contains extensive details on this survey, but it is worth noting here that it has information on where households are located (which I use to map them to shipyard counties, counties bordering shipyard counties, or counties unrelated to shipyards), their income, their age, their race, and their expenditures on a large number of items. The survey was conducted over the course of 1935 and 1936 and is meant to capture expenditures in the preceding calendar year. There are problems with using this data, since it is certainly not nationally representative and limited to urban areas, as noted in Hausman (2013). The time span covered by the survey is at the very start of the period seeing spending associated with the Vinson-Trammell Act of 1934. Critically, I am able to identify the extent to which the household's survey year overlaps with

[9]ICPSR Data Set 8908, U.S. Dept. of Labor. BLS. Cost of Living Division, U.S. Dept. of Agriculture. BHE. Economics Division, U.S. Natural Resources Committee. Consumption Research Staff. Industrial Section, U.S. Central Statistical Board, and U.S. WPA (2009).

spending on naval vessels by relying on newspaper articles reporting on the awarding of contracts. This provides crucial identification. Also, among the counties hosting active shipyards at the time of passage, only New York City and Mobile, Alabama are represented in this survey, although there are respondents living in a number of counties bordering shipyard counties. I follow Hausman (2013) in constructing my categories of consumption expenditure.

Next, I will extract information from the Census of Population for 1930 and 1940 on schooling so that I can estimate whether the spending shocks that I identify changed individuals' education choices.[10] These files have information on the number of individuals in each county in various age ranges, including people who are 14 or 15 years old, people who are 16 and 17 years old, and people who are from 18 to 20 years old, as well as the proportion of these people who are in school, for both 1930 and 1940.

From IPUMS 1% samples of the same two censuses, I also get an idea of the industrial structure of each county.[11] Specifically, I tabulate the number of people in each county who are employed and calculate the number of employed people who describe themselves as working in a certain industry. Each worker is classified into one of 16 major categories, which break down further into 149 subcategories. Again, I gather this information in both census years to see if the spending associated with the Vinson-Trammell Act impacted the industrial landscapes in shipyard

[10]Specifically, I obtain county level census data for 1930 and 1940 from "Historical, Demographic, Economic, and Social Data: The United States, 1790-1970," ICPSR Data Set 00003, Interuniversity Consortium for Political and Social Research (2005).

[11]I employ 1% IPUMS samples, because 5% samples are not available for 1940, and I wish to maintain consistency.

counties, counties bordering shipyard counties, and counties unrelated to shipyards differentially, a possibility noted in the model of Ramey and Shapiro (1998), where government purchases tend to incentivize the usage of capital for military production as opposed to civilian production.

The last bit of data is the list of ships built on United States Maritime Commission contracts between 1936 and 1946, which comes from Fischer (1946). This document lists the name and total nominal cost of each ship built in this period, as well as the name of the firm that built it and its location. Using this file, I can construct the total amount spent in nominal terms in each locality over the course of this decade. Unfortunately, this file does not include the specific months or years over which each ship was built, impeding the possibility of creating a real total cost for each county or conducting a finer year-by-year analysis using this information. By consulting a number of contemporary newspaper sources, I can assign a handful of ships to some particular date using the name of the ship and the yard it was launched from. This method seems to imply that a majority of the vessels contained in this document were launched after 1940.[12] Since they were built at a time when economic capacity was beginning to be constrained by the build-up to U.S. entry into World War II (see

[12]For example, a series of articles from the *Daily Boston Globe* from July 1939 to November 1940 catalog the launch of most of the vessels built for the USMC (according to the Fischer (1946) file) in the Boston area, but other articles name other ships in the file as being launched after 1940. In any event, my identification strategy for the main part of my analysis rests on the shipyards that were in operation in 1934, when the bill passed. This information comes from Lane (1951) and some other newspaper sources. Whether the ships built at these yards were done so under Vinson-Trammell or by the USMC should not be relevant to the question as long as these counties receive spending solely due to the fact that they hosted shipyards in 1934. It is clear from Lane (1951) that the expanded program of the USMC caused several new yards to open (in Houston and Tampa, for example). This will raise problems when looking at the outcomes that span World War II.

Gordon and Krenn (2014)), I will seek to use the purchase of these ships to compare the effects of government ship purchases in times of slack to the analogous effects of spending when the economy is operating nearer its potential.

5.3.4 Regression Specification

With a varied set of outcomes with differing time observations, it is necessary for me to estimate a number of different regressions. I will start by considering outcomes available in the Fishback et al. (2011b) data set. For a number of variables related to manufacturing, which have observations on the years 1931, 1933, 1935, 1937, and 1939, I estimate

$$\begin{aligned}\Delta Y_{it} = {} & \alpha + \beta_1 Shipyard_{1934,i} + \beta_2 BordersShipyard_{1934,i} + \sum_{t=1935}^{1939} \delta_t I(Year = t) \\ & + \sum_{t=1935}^{1939} \gamma_{1,t} Shipyard_{1934,i} * I(Year = t) + \sum_{t=1935}^{1939} \gamma_{2,t} BordersShipyard_{1934,i} * I(Year = t) \\ & + X_i' \Omega + \eta_{it} \ ,\end{aligned} \tag{5.1}$$

where ΔY_{it} is the two-year growth rate in some manufacturing variable, such as real manufacturing output or the number of manufacturing establishments, $Shipyard_{1934,i}$ is a dummy indicator for whether or not county i hosted a shipyard at the time of passage of the Vinson-Trammell Act of 1934, $BordersShipyard_{1934,i}$ is a dummy variable indicating whether county i bordered a county with a shipyard in 1934,[13] I(Year=t) is a series of dummies that stand in for time fixed effects and proceed

[13] In cases where a county hosted a shipyard and bordered another county also hosting a shipyard, I coded the $BordersShipyard_{1934}$ variable to be 0.

in two-year intervals, and X_i is a vector of control variables including state fixed effects, a dummy for whether the county is located on a coast, and whether the county is relatively industrialized or urban.[14] "Relatively industrialized" means that its percentage of the population employed in manufacturing before passage of the bill is greater than the national average in that year.[15] For a county's relative urban nature, I control for the percentage of the county considered urban in 1930. With several of these variables being invariant over time, including the $Shipyard_{1934,i}$ and $BordersShipyard_{1934,i}$ variables, county fixed effects would lead to identification problems. At the same time, however, state fixed effects help control for the possibility that relative strength or weakness of balanced budget rules at the state level confound the results. When conducting the regression analysis, I first exclude the top and bottom percentiles of the dependent variable so as to remove outliers. Then, because I am interested in running this regression on a balanced panel, I drop any county's observations if it is missing data for any year in the sample.[16]

It is worthwhile to take a moment to consider how to interpret the coefficients from this regression. The coefficient on the term $Shipyard_{1934,i}$ is identified only by variation in the first two-year interval. Thus, one can read this coefficient as the difference between the growth rates for shipyard counties and non-shipyard counties for the years between 1931 and 1933, i.e. before the spending shock took place. The

[14] Given the heavy northeastern concentration of the shipyards, one might think that region fixed effects would be more appropriate than state fixed effects, but inclusion of Census Bureau Region or Division fixed effects did not impact the results.

[15] Effectively, this means the share of the population employed in manufacturing in 1933 must be above the national average.

[16] The results are entirely robust to including large observations of the dependent variable and allowing the panel to be unbalanced.

coefficients on the three interactions between $Shipyard_{1934,i}$ and the fixed effects for the intervals from 1933 to 1935, 1935 to 1937, and 1937 to 1939 are read as the difference between the growth rates for shipyard and non-shipyard counties for these respective time periods, holding everything else equal. An analogous interpretation holds for all terms with the $BordersShipyard_{1934,i}$ variable. In a sense, one can read this regression as a sort of disaggregated difference-in-difference specification.[17] Because I do not have reliable data on where exactly among the shipyard counties the spending was allocated, I use dummy variables in the regression. Thus, one can interpret the effects that I uncover as an "Intention to Treat" (ITT) effect.

One possible threat to identification is that the shipbuilding industry was well placed for a return to health after a particularly nasty few years at the beginning of the Great Depression. It is hard to rule this idea out entirely, due to the relative paucity of data before the act's passage. I can show that when I only look at the observations on manufacturing up to the year 1933, there is little evidence that counties with shipyards were performing statistically differently from other counties. As an attempt to refute this "mean-reversion" story, I also run regressions on pre-1934 data only for the outcomes for which it is available. I also pursue an alternative route using a "propensity score"-type of methodology.

A further robustness check includes a variable that captures the change in or the level of New Deal spending for each year for which I have manufacturing data. I define New Deal spending per county as the sum of grants and loans from a number of

[17] When a more conventional difference-in-differences specification is employed, the results are broadly similar, but I cannot observe the detailed changes over time.

programs, for which information is available in the Fishback et al. (2011b) data set.[18] There is no annual data at the county level for New Deal spending. Fishback et al. (2011a) interpolate a county-level series for this type of aid by using the total amount of New Deal spending over the 1930s at the county level and state-level year-by-year fluctuations. New Deal spending is likely to be endogenous as the explicit purpose of the program was to help the economy emerge from the Depression. With this in mind, I follow Fishback, Horrace, and Kantor (2005) by employing a set of instruments in a Two-Stage Least Squares framework. As in that paper, my instruments for New Deal spending are the standard deviation of the share of the vote going to the Democratic Party in presidential elections from 1896 to 1928, voter turnout in the 1928 election, the log of the area (in square miles) of the county, the latitude and longitude of its county seat, and the share of the population that belonged to a church in 1926. It is not clear *ex ante* whether New Deal spending should enter the regression in log levels or in growth rates, so I try both specifications.

The next set of variables that I am interested in are those pertaining to consumption, such as retail sales data. These variables are only available in the years 1933, 1935, and 1939, necessitating a somewhat simpler specification. The associated

[18] This includes Reconstruction Finance Corporation loans, Disaster Loan Corporation loans, Public Works Administration Nonfederal loans, United States Housing Authority loan contracts, Farm Credit Administration loans, Farm Security Administration Rural Rehabilitation loans, Farm Security Administration Tenant Purchase loans, Rural Electrification loans, Home Owners Loan Corporation loans, Federal Housing Administration Title 1 insured loans, Federal Housing Administration Title 2 insured loans, Agricultural Adjustment Administration grants, Farm Security Administration Rural Rehabilitation grants, Public Roads Administration completed grants, Public Works Administration Nonfederal grants, Public Works Administration federal grants, Public Building Administration grants, Works Progress Administration grants, other works program grants, Social Security Administration grants, United States Housing Authority Public House grants, Federal Emergency Relief Administration grants, and Civil Works Administration grants.

regression equation that I estimate is

$$\Delta Y_{it} = \alpha + \beta_1 Shipyard_{1934,i} + \beta_2 BordersShipyard_{1934,i} + \delta I(Year = 1939)$$
$$+ \gamma_1 Shipyard_{1934,i} * I(Year = 1939) + \gamma_2 BordersShipyard_{1934,i} * I(Year = 1939)$$
$$+ X_i'\Omega + \eta_{it} \ , \tag{5.2}$$

where here, ΔY_{it} is the average annual growth rate since the last observation.[19] $I(Year = 1939)$ is an indicator variable for observations in 1939, and the coefficient on its interaction term with either the shipyard dummy variable or the shipyard border dummy variable is the coefficient of interest. These four years that follow 1935 are the only information I have on the possibly differential behavior of retail sales after the passage of the Vinson-Trammell Act. Like with the manufacturing outcomes, I drop the top and bottom percentiles of the distribution of the dependent variable and then also any counties missing data for one of the three years that in which I have observations.

There are also several variables that only have observations in the years 1935 and 1939, including retail earnings, retail employment, wholesale earnings, wholesale employment, and wholesale net sales. For these, the regression specification is yet simpler, written as

$$\Delta Y_i = \alpha + \beta_1 Shipyard_{1934,i} + \beta_2 BordersShipyard_{1934,i} + X_i'\Omega + \eta_i \ . \tag{5.3}$$

Now, the coefficients that I am interested in are those on the $Shipyard_{1934,i}$ and

[19] That is, for observations in 1935, $\Delta Y_{it} = \frac{ln(Y_{1935}) - ln(Y_{1933})}{2}$, and for observations in 1939, $\Delta Y_{it} = \frac{ln(Y_{1939}) - ln(Y_{1935})}{4}$.

$BordersShipyard_{1934,i}$ variables themselves, as I simply have a cross-sectional regression in differences. The dependent variables in this set of regressions are specified as $ln(Y_{i,1939}) - ln(Y_{i,1935})$. As with the previous two sets of regressions, outlier observations are excluded from the regression, as well as (necessarily) any counties that are missing data for either of the two years considered.

To place my results from this series of regressions in context and to get some kind of an idea as to whether the results are different when the economy moves from the very slack Great Depression period to the highly constrained World War II period, I follow in the spirit of Fishback and Cullen (2013) by regressing another set of equations that are essentially identical in form to those in Equation 5.3, with the essential difference being that the dependent variable measures the growth rate of some indicator, such as population, manufacturing establishments, manufacturing value added, sales or employment in the retail or wholesale sectors, or auto registrations, over a time period spanning World War II. For example, for most of the listed variables, I have an observation in 1933 or 1935 (before the passage of the spending bill or just as the spending was being implemented) and an observation in some post-war year, like 1947, 1948, or 1954. I will test the hypothesis that the growth rate in these variables is greater in countries that hosted a shipyard by 1934. In addition, since I have a clearer idea on the actual allocation of World War II spending on shipbuilding from the Fischer (1946) data, I can run a very similar regression, specified as

$$\Delta Y_i = \alpha + \beta_1 Shipyard_{USMC,i} + \beta_2 BordersShipyard_{USMC,i} + X_i'\Omega + \eta_i \ , \qquad (5.4)$$

where now $Shipyard_{USMC,i}$ and $BordersShipyard_{USMC,i}$ are dummy variables indicating that the given county hosted or bordered, respectively, a shipyard receiving

USMC contracts. Because I have data on how much was spent at each shipyard (at least in nominal terms), a sister regression to Equation 5.4 that estimates the effects of the government purchases on the intensive margin would be

$$\Delta Y_i = \alpha + \beta_1 ShipSpending_{Host,i} + \beta_2 ShipSpending_{Border,i} + X_i'\Omega + \eta_i \ . \tag{5.5}$$

In this equation, $ShipSpending_{Host,i}$ gives the log of nominal spending on ships by the USMC in that county. $ShipSpending_{Border,i}$ gives the log of total spending in all counties bordering county i.

To investigate the effects of the Vinson-Trammell Act on other economic outcomes, such as the proportion of children from various age groups in school or the change in the percentage of the employed workforce working in a particular industry, the specification used is that in Equation 5.3 and the dependent variables are the change in the variable of interest (all of which come from Census data) between 1930 and 1940.

Finally, I run regressions based on the 1935-1936 Consumer Survey. I exploit variation in the residence of the respondents (i.e., whether they live in a county hosting a shipyard or not), as well as in the extent to which the schedule year the household reports on overlaps with the initial burst of spending. Specifically, I estimate an equation of the form,

$$\begin{aligned} Y_i = \alpha &+ \beta_1 Shipyard_{1934,i} + \beta_2 BordersShipyard_{1934,i} + \beta_3 Overlap_i \\ &+ \beta_4 Shipyard_{1934,i} * Overlap_i + \beta_5 BordersShipyard_{1934,i} * Overlap_i \\ &+ X_i'\Omega + \eta_i \ . \end{aligned} \tag{5.6}$$

Here, as before $Shipyard_{1934,i}$ and $BordersShipyard_{1934,i}$ refer to whether the respondent household lives in a county hosting a shipyard or one of its bordering counties, respectively. Y_i denotes dollars spent in the past twelve months on some consumption category. To construct the variable $Overlap_i$, I take the difference between the end of the survey year for household i and the date when the first set of contracts were awarded as part of the Vinson-Trammell Act of 1934, which, according to *The New York Times*[20], was 22 August 1934. This variable is measured in days. The assumption underlying this variable's construction is the following. The article referred to makes plain that this set of contracts awarded was the first associated with the new navy spending and that building would start "promptly." Therefore, if the household is reporting on consumption before this date, then that consumption occurred without knowing when or where the new spending would be taking place. Also, the article alludes to the fact that more contracts would be awarded later, so to the extent that the household's consumption year moves further from this date, the more one might expect it to be influenced by the government spending. By interacting this variable with whether or not the household lives in a county hosting a shipyard or near a shipyard, I can examine the differential effect experienced by households in shipyard counties exposed to greater amounts of spending relative to those who do not live in shipyard counties and those who live in shipyard counties but are exposed to smaller amounts of government spending. Following Ozer-Balli and Sørensen (2013), I demean the $Overlap_i$ variable in the interaction term. X_i is a vector of controls that include the age and age squared of the husband and wife in the household, a dummy for whether the head of household is not white, and the

[20] *The New York Times*, "Awards Contracts for 24 Warships," 23 August 1934.

household's income.[21]

5.4 Results

Before discussing the results, I report summary statistics on a number of key outcomes that I will be examining in Table 5.3. The table shows that there was considerable variation in these dependent variables, with many of the standard deviations in growth rates far above the mean values. This is to be expected, given that the sample period that I examine is one of the more volatile economic episodes in the history of the modern United States. The table also illustrates the attractiveness of winsorizing the data, given the substantial outliers on both the high and low ends of the distributions. Next, I report the results of the regression analysis.

5.4.1 Results on Fishback et al. (2011b) Data

The data set constructed by Fishback et al. (2011b) contains a large number of variables that are of interest for this study. I will start by examining outcomes related to the manufacturing industry, for which the data has some of the best detail. Following that, I will consider retail sales outcomes, for which the analysis resembles a more conventional difference-in-differences framework, and conclude this subsection with a treatment of a number of miscellaneous outcomes.

[21] The regression results are robust to the exclusion of the income term.

5.4.1.1 Manufacturing Outcomes

I start by examining the results on growth in manufacturing output and manufacturing value added. Plots of the coefficients are found in Figures 5.4 and 5.6. These regressions come from a specification of the regression with a full set of control variables, excluding New Deal grants and loans. Results from a regression without controls are very similar. From the plot, one can see that manufacturing output in counties with shipyards grew over thirteen faster in the two years to 1937 than they otherwise would have been expected to. This difference is significant at the one percent confidence level. This is followed by growth of nearly identical magnitude and significance in the following two year period to 1939. The figure for manufacturing value added tells a very similar story. For neither outcome do I see significant effects in bordering counties. The sum of the extra growth in shipyard counties between 1933 and 1939 is 32%, with an associated p-value of 0.053. If I only consider the extra growth from 1935 to 1939, the sum is 28%, significant at the one percent level. No significant effects are seen for border counties in Figures 5.5 or 5.7.

The results of these regressions imply that the effects of the spending on local economies' manufacturing output and value added were extremely large. This is interesting in light of the fact that the entire economy was in a very dire state at the time the spending bill was passed. By including time fixed effects, I am able to disentangle the effects of spending on the treated counties from a more general tendency on the part of the entire United States to recover from the trough of the Depression. It is also interesting that there seems to have been no significant effect on the manufacturing output of nearby counties, although the signs of the coefficients are positive (in the latter part of the decade). At least with regard to areas in very close proximity, it is not apparent that the large output increases in shipyard counties drew resources away from their neighbors. Below, I will consider how these effects on output and value added may have also had an impact on employment, consumption, and other variables.

Table 5.4 contains estimates from a battery of robustness checks applied to the baseline regressions for manufacturing output. The first concern is that concurrent with this increased spending on warships was the New Deal spending program instituted by the Roosevelt Administration. Many of the programs associated with the New Deal were transfer payments, loans, and subsidies (not, as in the case of the Vinson-Trammell Act studied here, purchases of goods and services. In any event, it is beyond the scope of this paper to evaluate the effectiveness of the New Deal in stimulating economic activity. My only concern is that, for some reason, New Deal spending may have been systematically allocated to areas also likely to have shipyards. It is not clear, *ex ante*, whether when controlling for New Deal spending, the spending should be specified in log levels or in log differences, especially considering the biannual nature of my observations on manufacturing variables.[22] Therefore, I try both specifications, as well as one that controls for the sum of New Deal spending over the two year interval. What is clear is that New Deal spending is endogenous, as it was allocated to areas suffering from weaker economic activity. I follow Fishback, Horrace, and Kantor (2005) then in using an instrumental variables approach. The

[22]Specifying the New Deal spending in log levels seems to be the more natural approach, given the temporary nature of the programs, but this risks throwing out information on spending that took place in the intervening year.

instrument set for total New Deal grants and loans is that described in Section 5.3.4, as well as state fixed effects.

The first column of Table 5.4 contains the baseline specification already reported. The next three columns demonstrate that the inclusion of New Deal grants and loans do not qualitatively affect the results, and in the case where New Deal spending is specified in levels (over intervals of one or two years), the results do not change very much at all. Thus, I can conclude that the positive effects on manufacturing output that I am finding are due to the shipbuilding program and not to simultaneous New Deal payments.

In the baseline estimation, standard errors are clustered at the state level and I use state fixed effects. The next two columns of Table 5.4 consider whether or not the baseline findings are senstive to these specification choices. The fifth column of the table demonstrates that clustering the standard errors at county, rather than state, level leaves the point estimates unaltered and the significance levels nearly so. The same outcome is the case when regional fixed effects are substituted for state fixed effects. One may be concerned that the heavy Northeastern concentration of the shipyards still operating in 1934 is partly driving the estimated effects, but it is clear that this is not an issue.

A reasonable question to ask is whether it is not the case that counties that are home to shipyards are not in some respect fundamentally different from other counties. That is, it may not be appropriate to pool these relatively urban, highly industrialized areas with more rural, sparsely populated local economies. In an effort to address this concern, I undertake the following exercise, which is similar to a propensity score-type analysis. I first run a cross-sectional regression, in which the dependent variable is the presence of a shipyard in 1934 and the independent variables are state fixed effects, location on a coast, the percentage of the county that is urban, and whether or not it is highly industrialized. I then sort the counties by the fitted values from this regression and limit the sample to only the top 25% by this "propensity score." I then rerun the baseline regression on this smaller, theoretically more homogeneous sample. The results are in the seventh column of Table 5.4 (the one labeled "Propensity Score 1"), where it is apparent that even among like counties, those hosting shipyards see significantly faster growth in the latter part of the 1930s.

I conduct a further robustness check by examining whether counties with heavy concentrations in other industries see a similar time path of output and retail sales over the 1930s. The results of this experiment can be found in Tables 5.5. To make the regressions in this experiment comparable to those evaluating the outcomes of shipyard counties, I exploit the fact that the Fishback et al. (2011b) data set gives the number of employees in a variety of industries as of 1930. I divide the workforce in each industry by the population in the county and then rank each county by the industry's share of the population. Because I have 26 shipyard counties, I code the top 26 counties in each industry with a dummy variable indicating them as having a heavy concentration of that industry. Then, I repeat the regression of Equations 5.1 replacing the $Shipyard_{1934}$ dummy variables with the dummy variables for each of the industries that I consider. One can interpret these as a sort of placebo test. It is not obvious that any of these industries were explicitly subject to a government spending shock, so, on balance, there should not be any significant difference in output. At the least, they should have sequences different from those of shipyards.

Table 5.5 gives the results for the output growth regressions. As can be seen in the table, for many industries, the effect is insignificant in all three years considered. No industry sees a pattern that matches the trajectory of shipyards exactly (with large significant increases in the last two biannual periods in the decades). Even those that do see significant increases tend to be those that would be related to shipbuilding, such as iron, lumber, and rubber.[23] Thus, it is clear that shipyard counties see a unique combination of effects on output and retail sales that I conclude is due to the sizable shock to demand emanating from the government starting in 1934 and 1935.

The number of treated counties is relatively small, and this might produce worries that the results are driven by particularly large responses in one or two shipyard counties. In Table 5.6, I attempt to address this concern by dropping individual shipyard counties, one by one, from the regression equation. Each column in the table reports the coefficient on $Shipyard_{1934,i}$ interacted with the fixed effect for the indicated year. From the table, it is clear that the results are robust to dropping any one individual shipyard county from the sample.

Before moving on to other outcomes, it may be important to demonstrate that the results found so far are not due to mean reversion. That is, I would like to argue against the notion that the positive effects on manufacturing and consumption reported above are due solely to the natural recovery of the shipbuilding industry and

[23]One surprising result of this exercise is the really poor performance seen by counties for whom cotton was an important industry. This is likely due to policies associated with the Agricultural Adjustment Act of the New Deal, which incentivized farmers not to plant and may have had very negative effects on other industries in those counties as well.

its environs. To be sure, this is an extremely difficult story to rule out, especially considering the relative paucity of data available to me before the passage of the Vinson-Trammell Act. I do attempt to argue against this explanation in a couple of different ways. Table 5.7 gives the results from regressions including only data before 1934.[24] For most of the outcomes under consideration, this reduces to the two-year period from 1931 to 1933, although for manufacturing employment and retail sales per capita, I can also include the two-year period from 1929 to 1931. The table shows that there is only weak evidence (seen in coefficients significant at the ten percent level for manufacturing value added and wage payments) that shipyard counties were doing especially badly before the passage of the bill. This is inconsistent with the idea that they were subsequently "due" for a stronger-than-average recovery.

In the last column of Table 5.4, I conduct another propensity score-type analysis, in which I include, along with the variables mentioned above, the manufacturing output growth rate from 1931 to 1933. In this way, I hope to limit the sample not only to counties similar to shipyard counties in terms of their demographic and structural characteristics, but also to those that had a similar experience economically in the last observed two-year period wholly previous to the passage of the bill. In this regression, it is clear that the signs and magnitudes of all of the coefficients are roughly the same as in the baseline estimation, but the significance is weakened somewhat, especially for the two years from 1935 to 1937. Still, manufacturing output in shipyard counties grew nearly 12% faster than in other counties (significant at a confidence level of five

[24] As can be very clearly seen from the table, I include all of the outcomes that I am considering in this section of the paper, although the main results on these outcomes will be discussed formally below.

percent) from 1937 to 1939, even when limiting the sample to areas that had similar economic dynamics leading up to the authorization of the spending program.

I turn now to results on some other manufacturing outcomes. Looking at total manufacturing employment in Figure 5.8 and Figure 5.9, the point estimates on the three post-1934 interaction terms are all positive, but they are imprecisely estimated. The p-values on the interaction terms with the 1935 and 1937 fixed effects range from 0.11 to 0.15 for shipyard counties. There is no significant effect on bordering counties. It is interesting that output should be so positively effected, while the effect on employment is more muted. I will use the next series of graphs to try to untangle why this is so.

Figures 5.10 and 5.11 illustrate that total wage payments by manufacturers were positively affected by the spending in shipyard counties. Again, no significant effect is discernible in bordering counties. The magnitude of the effect on wage payments is similar to that on manufacturing output, and all three post-1934 interaction terms are significant at the fiver percent level. When I look at the results for average earnings per manufacturing employee (Figures 5.12 and 5.13), I can see that the post-1934 interaction terms all have positive point estimates. In bordering counties, this positive estimate is statistically significant for 1935, and for shipyard counties, it is significant for 1939. Therefore, the significant effect on wage payments that I observe must be due to some combination of firms hiring more workers and paying their existing workers more.

Figures 5.14 provides an additional layer of detail. Here, one can see that the effect of the spending bill on manufacturing establishments was negative. This negative impact is significant at the ten percent level in the two year period to 1935 and has a p-value of about 0.11 in the two year period to 1937. This decline in the number of manufacturing firms is accompanied by a rise in the average number of employees per firm that is strongly significant (see Figure 5.16). Additionally, one can see from Figure 5.18 that manufacturing output per worker also grew significantly faster in shipyard counties than elsewhere from 1937 to 1939 (again, with little significant impact on border counties).

Therefore, the data reveals a story in which the spending on ships has a negative impact on the number of firms, possibly through higher wages, while surviving firms are larger and more productive (at least with regards to labor productivity). The increased hiring of the existing firms is offset by a decline in the number of firms, muddying the effect on total employment. The result seems to be modestly higher employment with modestly higher earnings per worker, causing a rise in total wage payments and having a negative impact on the number of establishments. Although a detailed examination of the effect of this aggregate demand shock on the industrial organization of the affected counties is beyond the scope of this paper, these firm distribution dynamics are interesting and merit further research.[25]

[25]Kehrig (2015), for example, builds a model intended to explain the observation that in recessions, dispersion in productivity among firms becomes greater as all firms, even productive ones, use resources less efficiently. As a result of the (positive) shock I study, the number of firms declines and the survivors use more labor more efficiently, so it appears, at first glance, that these results are consistent with the model of Kehrig (2015).

5.4.1.2 Retail Sales Outcomes

My results suggest an increase in employment in shipyard counties. If I see an increase in consumption as well, this finding would be consistent with the assumption of nonseparable preferences, such as complementarity between consumption and labor. Although I do not have data on consumption at the county level for this period, I do have evidence on retail sales. Retail sales are by no means a perfect proxy for consumption, but they have been used for this purpose in previous studies, such as Ostergaard, Sørensen, and Yosha (2002), Fishback, Horrace, and Kantor (2005), Shoag (2010), and Romer and Romer (2014). The results from estimating Equation 5.2 can be found in Table 5.8.

The table shows a positive impact of a shipyard's presence on retail sales growth in a county.[26] The coefficient on the $Shipyard_{1934,i}$ dummy variable is 0.038, and is significant at the one percent confidence level. What is more, this positive effect on retail sales also spilled over into bordering counties, where the coefficient is 0.045 and is also significant at the one percent level. This is an interesting result in that an aggregate fiscal multiplier greater than unity should involve positive effects even outside the area that directly receives the spending. This is also found in the international context of Auerbach and Gorodnichenko (2013b). These results also hold up in an instrumental variables regression that includes New Deal spending (see the additional columns in Table 5.8) and when shipyard counties are dropped from the regression on an individual basis (see the results reported in Table 5.10.

[26] The results for per-capita retail sales growth, data for which are present in Fishback et al. (2011b), are very similar.

In addition, when I consider counties with high concentrations in other industries, for very few of them does the same pattern emerge (see Table 5.9). Therefore, the data reveals the complementarity between labor and consumption implied by nonseparable preferences and necessary for the large aggregate multiplier suggested by the Nakamura and Steinsson (2014) model. This is important not least because they are not able to explicitly test for this complementarity since they do not have reliable consumption data. In this sense, my results support those of Nakamura and Steinsson (2014) by estimating results for output that are similar to theirs and providing direct evidence for nonseparable preferences, an assumption critical to their model.

5.4.1.3 Miscellaneous Outcomes

Further evidence on the effects on local economic activity can be seen in the regression results on Equation 5.3 reported in Table 5.11. These include the effects on outcomes including wholesale employment, earnings, and net sales, as well as employment and earnings in the retail sector. The period covered by these regressions is from 1935 to 1939, so that these are simple cross-sectional regressions. Here, it is clear that growth in wholesale and retail employment was much slower in shipyard counties than in like counties without shipyards. In particular, retail employment grew six percent slower, significant at the 5% level. This suggests some crowding out within these counties, as workers gravitated toward the manufacturing work stimulated by the shipbuilding program, and it is consistent with a story in which sectoral shifts took place in the local economies, like in the model of Ramey and Shapiro (1998). In addition, when

I instrument for the change in New Deal spending at the county level, it turns out that retail earnings growth was itself much slower in these counties. These results are a bit puzzling, especially in light of the fact that both manufacturing output and retail sales were concurrently growing so fast in shipyard counties.

5.4.2 Results from Census Data

I now turn to results based on Census data. All of the regressions reported below take the form of Equation 5.3, and the results are contained in Table 5.12. I note first that none of the dependent variables for which results are reported above are in per capita terms, but rather show the growth rate in aggregate quantities. This is because, for the intercensal years, I only have population data that is arrived at via straight-line interpolation between the 1930 and 1940 values. The first line in Table 5.12 illustrates that the significant effects found above are not due merely to an influx of people to shipyard counties, as these counties see no significant difference in their population growth relative to other counties. Border counties, however, do see significantly faster growth over the decade.[27] Despite this, for example, retail sales growth per capita in counties neighboring shipyards is still significantly greater than in counties not located near shipyards, assuaging some concerns about population inflows confounding the results. Still, it is interesting that, as a result of the spending, there is a reallocation of resources, although not to the directly affected areas. To my knowledge, this is the first paper that uncovers such a result.

[27] I also run a specification of this regression in which I control for population growth from 1920 to 1930 in order to take account of long run trends in migration, but these do not affect the results at all.

Also from census data, I can estimate the impact of the shipbuilding spending on educational choices in the county. Recent research by Charles, Hurst, and Notowidigdo (2014) suggests that speculative housing price booms prior to the Great Recession altered the opportunity cost calculations of marginal college students, leading more of them to foresake the pursuit of an advanced degree. Gupta (2015) also finds that macroeconomic policies can have unintended consequences on education choices. I can use data on the proportion of 14-15 year-olds, 16-17 year-olds, and 18-20 year-olds who are enrolled in school to see if a similar dynamic is at play in this context. The intuition is that greater economic activity resulting from the spending on ships (captured by the higher employment and wages seen in Figures 5.12 and 5.8, may have raised the opportunity cost of staying in school, leading to a drop in enrollment. On the other hand, if the boom in shipbuilding stabilized the earnings of heads of households, it is possible that it may have relieved the pressure on adolescents to seek work. The second three lines of Table 5.12 show that neither of these effects makes itself evident in this context, as there are no significants effects on the proportion of the variously aged groups in school. This would support the notion that agents operating in the 1930s viewed education as an investment (rather than as consumption), which should not be affected by short run fluctuations in income.

Finally, I examine whether this new spending on ships had any impact on the industrial structure of the counties hosting shipyards, such as whether it made the shipbuilding industry relatively more important to counties that hosted shipyards. This is motivated in part by the model of Ramey and Shapiro (1998), who develop a neoclassical model of government spending, in which the spending is sector-specific (as is the case in the spending I study), and it is costly to reallocate capital from one sector to the other. To the extent that there are frictions impeding the mobility of capital, their model offers differing predictions on the paths of such indicators as consumption and real interest rates. To answer this question, I calculate the proportion of employed workers indicated to be working in various sectors of the economy from the 1930 and 1940 Censuses, and I look to see if the relative change in counties with shipyards or their neighbors is greater than elsewhere. In the model of Ramey and Shapiro (1998), labor is perfectly flexible, so relative sectoral shifts in government spending should draw labor from non-military sectors to the shipbuilding sector. The second panel of Table 5.12 reveals that there does not seem to be much of an effect in Agriculture, Fisheries, and Forestry, Mining, Construction, Durable Goods Manufacturing, Nondurable Goods Manufacturing, Transportation, Communication, Utilities, Wholesale, Retail, Financials, Business Services, Professional Services, or Government in shipyard counties. While there are some significant coefficients on the indicator variable for counties that border shipyards, the overall picture is of little industrial flux engendered by the new government spending on ships, at least when looking at such broad categories. As naval vessels are durable goods, it is particularly surprising that I see no significant effect on the share of the workforce employed in durable goods manufacturing. By digging a little deeper, however, I do see that there are shifts within the durable goods sector, if not across manufacturing sectors. The share of the workforce employed in shipbuilding rises by 2.3% in shipyard counties relative to other counties. There is also a slight uptick in the share employed in aircraft and parts manufacturing (not reported). The durable goods sectors that are most negatively affected are Other Primary Iron and Steel Industries and Miscellaneous Manufacturing Industries. The fact that the shares of durables and nondurables manufacturing do not change supports the notion (applied in the translation of my results into the Nakamura and Steinsson (2014) model) that manufacturing output increases did not crowd out or crowd in activity in other industries.

5.4.3 Results from Consumer Survey

For the last set of regressions with 1930s data, I consider the consumption habits of households living in counties home to shipyards in 1934. I follow Hausman (2013) in making use of the Study of Consumer Purchases in the United States, 1935-1936, an early attempt by the government to gain an understanding of individual consumption behavior. It is an imperfect measure of consumption in many ways,[28] but this survey ought to provide at least some insight into whether households living near shipyards were able to consume more as a result.

Table 5.13 gives the first set of regressions of Equation 5.6. The first column of the table demonstrates that, on average, consumption in shipyard counties is significantly greater than in non-shipyard counties. Though this number is stark, it does not, in itself, carry much information, because it does not say anything about whether consumption increased as a result of the naval spending. Similarly, the coefficient on the variable measuring the number of days overlapping the household's survey

[28]See the detailed description in Hausman (2013) or Section 5.3.4 above. Also, the spending categories discussed below follow directly from the definitions in Hausman (2013).

year and the time since the announcement of the first contracts awarded suggests that overall consumption throughout the country began to rise later in the survey period, but it is not possible to attribute this to the Vinson-Trammell spending. On the other hand, the significant coefficient on the interaction between the shipyard indicator variable and the number of days overlapping is quite informative. It implies that for every day more than the national average that a particular household's survey year overlapped with the Vinson-Trammell spending when they lived in a shipyad county, they consumed an extra \$2.33 relative to households living in a non-shipyard county. This coefficient is significant at the five percent level. This is on top of the extra \$0.93 per day that they consumed relative to their neighbors whose survey year overlapped less with the spending. Although the signs of the coefficients when income is regressed on the same equation are the same, they are not significant.

It is worthwhile to put this result into context. A household living in a shipyard county spends \$2.33 per day (relative to the average) that they are exposed to the shipbuilding program. The median number of extra days of exposure (again, relative to the average) is 11, implying that the median household with a greater than average exposure to the program spends an extra \$25.63 (\$2.33 per day $\times$ 11 days) in their survey year. This translates to about an extra \$325 in 2009 dollars. Thus, the extra spending is large enough to be significant, but it is not an implausible jump in consumption.

The first column in the first panel in Table 5.14 shows that this is not merely due to a relaxing of the household's budget constraint. The regressions in this table include income as a right hand side variable. For total consumption, the coefficient on the interaction term between living in a shipyard county and the overlap between the survey year and the spending barely changes. Consumption rises by an extra \$2.34 per day even holding income constant. There are two possible explanations for this. It could be that households know that further spending is on ships is on the way as made clear in the newspaper article already mentioned. Thus, their expectations for higher income in the future are driving higher consumption now. It is unclear how much weight to give this explanation given the depressed economic environment and the parlous state of the banking sector at this time. The other explanation could be, as argued above, that labor supply and consumption are complements in the utility function, and the increased employment in shipyard counties is causing an increase in consumption as well.[29]

The rest of Table 5.14 gives a more detailed breakdown of the type of spending that consumers were increasing. The most significant effects are on housing operation, medical care, recreation, and food. Interestingly, spending on education declines significantly by \$0.35 per day of overlap. Although the census regressions in Section 5.4.2 did not show any significant change in whether children were attending school, it does seem that, on the intensive margin, they were investing less in schooling. This would be consistent with a story in which the increased public spending raised the opportunity cost of education and made working a more viable alternative for younger agents.

[29]The coefficients on variables relating to households living in counties bordering shipyards were almost all insignificant, so I do not report them to conserve space.

5.4.4 Results on Outcomes Spanning World War II

To this point, I have only considered the effects of spending on ships on various economic indicators between the passage of the bill (1934) and the end of the 1930s. This is partly because my most detailed data covers this particular period, but it is also because I am interested in whether the effects of government spending are different when there is a considerable degree of slack in the economy, which aptly describes the years of the Great Depression. As a point of comparison, I would like to separately consider whether the effects of this spending persisted after the economy exited the Depression and entered the World War II period. I can do this first by considering the counties I have already identified which are likely to have received spending associated with the Vinson-Trammell Act of 1934. I can also consider the effects on counties that hosted spending sponsored by the United States Maritime Commission over the period from 1936 to 1946. One of the advantages of this experiment is that I have at least a rough idea about the allocation of spending across yards, thanks to Fischer (1946), which I do not have for the Vinson-Trammell spending. Of course, studying the effects of government spending during World War II is a veritable minefield, due to the very different nature of the U.S. economy, including a likely greater degree of "command-and-control" than at any other time in the nation's history. Still, some previous papers have dared to tread on this ground, such as Fishback and Cullen (2013), who find little evidence that local war spending affected local economic outcomes, and McGrattan and Ohanian (2010), who argue that a standard neoclassical model can account well for aggregate fluctuations during the war. It may yet be informative to consider whether the same effects uncovered for the 1930s exist for the 1940s as well.

I can measure the effect of shipyards' presence by 1934 on the growth of a number of variables between a point in time before the spending (1933 for most variables), and the first observation on that variable after the war. In some cases, data availability restricts me to using 1935 or 1936 as a starting year. The ending year is either 1947, 1948, or 1949, depending on what is available for each variable. For reference, rationing ended by 1945, so the restrictions associated with rationing had only been lifted for two years by the time I observe my first outcomes. That said, I am hesitant to allow the focus of this experiment to drift too far past the end of the war. Table 5.15 displays the results from regressions of Equation 5.3 on data that spans from the era of the Depression to after the end of the war.

Overall, the results are not as strong for the period spanning the war, as the first panel of Table 5.15 indicates. For counties hosting shipyards, the growth in manufacturing value added is nearly 17% slower than in the country as a whole, and this is significant at the five percent level. Wholesale employment growth is also marginally significantly slower by about 14.5%. I cannot say that this is the result of the spending associated with shipbuilding crowding out other activity or if activity in the rest of the country was starting to catch up as military spending was broadly spread during World War II. Either way, the counties directly exposed to the Vinson-Trammell spending are seeing no additional benefit from it after the 1930s end. This is not the case for their neighbors. Interestingly, counties that border shipyard counties but that have no shipyard of their own see significantly higher growth in auto registrations (implying higher consumption on durable goods) as well as significantly higher retail employment growth, relative to the rest of the country.

I can also control for total military spending in each county over the course of World War II. The results are generally robust to accounting for this spending. Growth in manufacturing value added and retail sales is significantly slower in counties that were home to shipyards in 1934, while their neighbors still see significantly faster growth in automobile registrations, although the positive effects on retail sales and employment growth disappear with the inclusion of World War II spending.

It is not easy to pin down the mechanism at work here. It is possible, though hardly certain, that the benefits of the spending during the 1930s in shipyard counties are finally spilling over into their neighboring counties. This would support the notion of a large aggregate multiplier, and though the observed extra growth is some time later, restrictions associated with the war effort may have delayed the manifestation of the spillovers. Unfortunately, however, I cannot conclusively answer this question. Complicating matters is that some new yards began to open towards the end of the 1930s and into 1940 (see Lane (1951), p. 34), quite possibly so as to obtain spending contracts, and these may very well be confounding the results.

The counties affected by the Vinson-Trammell Act were not, however, the only areas to experience shipbuilding spending once the war started. I now turn to the effects of the USMC ship purchase program, which touched a considerably larger group of localities. I consider two specifications, as discussed in Section 5.3.4, one in which the right hand side variable is a dummy indicating whether the USMC bought ships from a given county or not (with a separate dummy for their neighbors) and a second that specifies the independent variable as the log of the total nominal amount spent on ships by the USMC in each county. In this second specification, the border counties are attributed the total nominal amount spent in all neighboring counties that hosted shipyards. Thus, the former specification can be considered as examining the extensive margin, while the latter focuses on the intensive margin. I consider the effects on the same set of outcomes considered in Table 5.15, and the results can be found in Table 5.16.

A number of results stand out from these regressions. Firstly, with the exception of a significantly faster rate of auto registration growth (on the extensive margin) and a slightly higher rate of growth in the number of manufacturing establishments, counties that hosted shipyards building USMC ships do not seem to experience better economic outcomes than areas with no shipyard connection. In this sense, it is possible that there is relatively more crowding out of private activity in these shipbuilding counties, especially since the period that I study in this exercise is one of severe capacity constraints. The second point to note is that border counties see genuine spillover effects from the spending next door. On the extensive margin, bordering counties have significantly faster growth in the number of auto registrations, retail sales, and retail employment. The same holds on the intensive margin. For a given county that may or may not host a shipyard, an increase of one percent in the nominal total of spending in all neighboring counties over the decade spanning World War II raises the growth rate of auto registrations by 0.47%, the growth rate of retail sales by 0.37%, and the growth rate of retail employment by 0.39%. All of these are significant at the one percent level. When I include an additional control variable for overall military spending in the county, these results mostly hold up.

This has curious implications. Presumably, the military spending variable includes the spending by the USMC on ships, so the regression results suggest that spending on shipyards had effects on border counties over and above that of military spending more generally. It is not clear why this should be so, but it is also the case that the data under consideration here is sparse enough that a more detailed analysis may yet be informative.

5.5 Scaling the Local Multiplier to the Aggregate Level

In the literature, it is the aggregate government spending multiplier that is often of greatest interest. Local government spending multipliers may not adequately convey information about general equilibrium effects that could cause the aggregate multiplier to fall below unity even as a dollar of spending in a given county generates more than a dollar of output in that county. If output in counties that do not receive spending (or that have spending taken away) falls by more than the lost government purchases, these negative effects could, in the aggregate, outweigh the booms experienced by areas that receive government spending. For example, I have already shown in Section 5.4.2 that the spending program compelled a movement of individuals into bordering counties. In this section, I will attempt to take the results that I have presented thus far and interpret what they imply for the government spending multiplier that is often estimated in the literature on fiscal policy. The first exercise will be to see what the model of Nakamura and Steinsson (2014) implies for my results.

To do so, I must alter my baseline regression so that it looks a little more like that estimated in the empirical section of Nakamura and Steinsson (2014). I first observe that Ramey and Zubairy (2014) estimate the Vinson-Trammell Act spending at about 1.5% of 1933 nominal GDP, which was about 57.2 billion dollars. This implies a spending program of about 858 million dollars. It is implausible to assume that the spending was distributed evenly among all the shipyard counties, but, for the purposes of this exercise, I will do so, since I cannot well defend any other allocation assumption without more detailed data. In Section 5.3.2, I identify 27 shipyard counties, but I do not have manufacturing data for Newport News, Virginia, so I will assume that the other 26 counties split the spending equally among themselves. This obviously raises potential problems, as the regression will be understating the effects of spending in counties that received less than average, while overstating the effects of spending in the counties that received more than average. Add to this the fact that, if any funds were allocated to Newport News, then the regression is now distributing those funds elsewhere, thus potentially further understating the effects of spending overall. Again, however, I do not mean this to be a formal multiplier estimate, but rather to see what the model of Nakamura and Steinsson (2014) implies for this data.

I also would need to scale the amount of spending by overall output in order to match the regression of Nakamura and Steinsson (2014). Since I do not have overall output at the county level for this time period, I create a rough measure by scaling manufacturing output in 1933 by the percentage of the population employed in manufacturing in that year. While this likely introduces further possible measurement error into the hypothetical regression, it is the best option available to me. I rerun Equation 5.1, substituting the per-county amount of spending scaled by overall output in the county for the $Shipyard_{1934,i}$ dummy variable and the $BordersShipyard_{1934,i}$ dummy variable in the interaction terms. The results of this regression are found in Table 5.17. This table shows that, if the ship purchases were distributed evenly across the shipyard counties, the additional manufacturing output over the course of 1933 to 1939 that could be attributed to them summed to 2.18 dollars for every dollar spent by the federal government on ships. This scales up to a multiplier of 2.64.[30] This is the "Open Economy Relative Multiplier" of Nakamura and Steinsson (2014).[31][32]

Nakamura and Steinsson (2014) consider develop a model in which regions within a monetary union are subject to differential government spending shocks.[33] That is, they examine how the aggregate economy will respond when only one region in their

[30] That is, each coefficient $\beta_t \; for \; t = 1935, 1937, 1939$ is multiplied by the inferred growth rate of output over the period from 1933. For example, to interpret β_{1935} as a "multiplier," (β^M_{1935}) I calculate

$$\beta^M_{1935} = \beta_{1935} \times \frac{Y_{1935}}{Y_{1933}} = \frac{\Delta(Y_{1937} - Y_{1935})}{Y_{1935}} \times \frac{Y_{1933}}{\Delta Shock} \times \frac{Y_{1935}}{Y_{1933}}. \tag{5.7}$$

I follow a similar process for β_{1937} and β_{1939}, with Y_{1937} and Y_{1939}, respectively, substituting for the numerator in the final term of the expression.

[31] In their paper, the open economy relative multipliers on total output range from 1.4 to 1.9.

[32] Of course, this extra 2.64 dollars in manufacturing output may have crowded out some other kind of output, but the data is not capable of revealing this explicitly. For this exercise, I will assume that no crowding-out or crowding-in results from this extra manufacturing output. Results below on relative changes in the industrial composition in shipyard counties suggest no crowding in or out.

[33] A brief summary of the model can be found in Appendix A.

model economy is subject to an increase in government spending. They consider several different specifications of their model. For my purposes, the one that is likely to be most relevant is that where there is nominal price rigidity, nominal interest rates are held constant by the monetary authority (because rates were at zero during the Great Depression), and there is complementarity between consumption and labor in the representative agent's utility function. This last point is supported by my empirical results that show that manufacturing output and retail sales rose simultaneously in shipyard counties in response to the Vinson-Trammell Act and that individual households exposed to the spending spent an extra \$2.33 per day that they were exposed in spite of the fact that their incomes had not yet risen. The results of this specification of their model can be found in the third and fourth rows of Table 7 in Nakamura and Steinsson (2014). When the government spending shock is relatively short-lived, the model implies a local government spending multiplier of \$2.04, which is not very different from my empirical finding of \$2.64. In this case, Nakamura and Steinsson (2014) find that the aggregate government spending multiplier implied by a local multiplier of this magnitude is 8.73. That would suggest that my empirical results would suggest a multiplier at least this large.

Of course, this figure is huge, and I am not aware of any aggregate multipliers estimated in postwar data that come very close to this. That said, as implausible as such a large multiplier might be in the context of the modern postwar United States economy, it may not be so incredible for the 1930s, when the economy was experiencing an extremely large degree of slack[34] and it was much less open to international

[34]According to the data set accompanying the work of Ramey and Zubairy (2014), the unemployment rate was never below 12% between 1934 and 1940 and in some periods, it was higher than

trade (and likely even intra-national trade).

Some may even consider the \$2.64 figure as a decent approximation to the aggregate multiplier. This number is also large according to modern theory and empirics, but it is much closer to the standard range than something between eight and nine. For this multiplier to approximate the actual amount, however, one would have to take very seriously the idea that there were no spillovers, positive or negative, in counties not hosting shipyards. Further, one would have to assume that, although the tax burden was increased in counties not playing host to shipyards, this did not alter the economic behavior of these counties, which does not seem like a palatable assumption to make.

Another approach might be to consider the argument of Gabaix (2011), who posits that when the distribution of firms is sufficiently fat-tailed, idiosyncratic fluctuations for particularly large firms can have effects in the aggregate. That is, they do not die out according to the central limit theorem. Seeing that the distribution of manufacturing output across all counties may be fat-tailed as well (which can be observed from Figure 5.20), it may be useful to adapt the notion of the "granular residual," as developed in Gabaix (2011), to this context.

The procedure is fairly straightforward. First, I consider a series of cross-sectional regressions of the following form for the years 1933, 1935, 1937, and 1939 (the years for which I have observations on manufacturing output growth),

$$\Delta Y_i = \alpha + x_i'\beta + \varepsilon_i \ . \tag{5.8}$$

20%.

where ΔY_i is the log difference of manufacturing output from two years earlier and x_i is the vector of control variables considered in the baseline regression above (location on a coast, percentage urban, indicator for being industrialized in 1933, and state fixed effects), but I do not include the $Shipyard_{1934}$ dummy variable in this regression. I take the residuals from this regression and weight them by the share of overall manufacturing output of County i two years earlier. The sum of these weighted residuals is the granular residual of Gabaix (2011) applied to the current context. By sorting the counties hosting shipyards from those not hosting shipyards, I can divide the overall granular residual into one from shipyards and one from all other counties, thus quantifying the contribution from each type of county to overall fluctuations in aggregate manufacturing output.

By considering Table 5.18, one can see that the overall granular residual (the last column of the table) follows an expected pattern. It is highly negative in the two years to 1933, positive in the two-year periods to 1935 and 1937 (as the economy recovered from the Depression) and negative again from 1937 to 1939 (when the economy re-entered recession).[35] What is surprising is the contribution of the shipyard counties, especially in the periods from 1933 to 1935 and 1935 to 1937. In the two years to 1935, has a magnitude equal to about −36% of the overall granular residual, implying that aggregate manufacturing output would have grown much faster had it not been for the shipyard counties. In the two years to 1937, it is 122% of the overall granular residual. This is puzzling, considering the results reported in all of

[35]The granular residuals of the shipyard counties, counties bordering shipyards, and all others generally sum to the overall granular residual, although the figures in the table may not do so exactly due to rounding.

the previous sections. Note also that shipyard counties contribute positively to the granular residual in the period to 1939, implying that growth would have been worse had it not been for these counties.

A closer inspection of the data reveals that the negative contribution to the granular residual on the part of the shipyard counties is, in large part, a result of the fact that New York City and a number of its large suburbs in New Jersey are home to shipyards. This reveals a curious result. The shipbuilding program caused manufacturing output growth to be much faster than average in the counties that owned shipyards, but this arithmetic average is driven by the positive growth seen in many smaller shipyard counties. In some larger areas that also happened to host shipyards, growth was still less than would have been expected, even after controlling for a number of covariates. Of course, a tricky aspect of this observation is that the shipbuilding industry was not nearly the kind of driving industry for New York City as it might have been for Bath, Maine. It is not at all clear if New York City would have actually seen even slower manufacturing output growth if it had not had its shipyards. In fact, the results from looking at the consumer survey suggest as much. Thus, although this exercise motivated by the discussion of a granular residual that drives aggregate fluctuations in Gabaix (2011) seems to suggest an overall negative multiplier on the Vinson-Trammell spending, even this does not settle the question.

5.6 Conclusion

In this paper, I have contributed to the study of government spending multipliers at the local level by considering in detail the effects of the Vinson-Trammell Act of 1934, a bill that facilitated the purchase of a substantial number of naval vessels in response to military expansion by Japan and in order to build the United States Navy up to treaty provisions. Using a combination of historical sources and contemporary news media, I am able to identify counties that hosted shipyards before the passage of the act. I combine this with county level data on various economic indicators in the 1930s to investigate the effect of this spending bill on local economic outcomes.

I find that counties that hosted shipyards in 1934 experience significantly faster growth in manufacturing output and value added. Total manufacturing wage payments are also significantly positively impacted, with this likely composed of higher employment and higher average wages per worker. This combination seems to have favored larger firms and negatively affected the number of manufacturing firms in each county. Retail sales growth grew significantly faster in these counties as well, lending support to the use of a model with complementarity between labor and consumption in the utility function. These results are not due to faster population growth and they are robust to the inclusion or exclusion of a number of control variables, including spending associated with the New Deal. By considering a consumer survey that was coincident with much of the spending, I find that households living in shipyard counties spent upwards of two dollars a day more for each day that they were exposed to the government spending.

At the same time, the spending associated with the government's ship buying program did not affect individuals' choices on education or the broader industrial structure of the county, except for some shifting of the relative importance of the shipbuilding industry within the durable goods manufacturing sector.

These results appear not to have lasted through the Second World War, and a look at the effect of more detailed data on government ship purchases during the war reveal no effect on local outcomes, although counties bordering shipyard counties seem to have experienced greater benefits after the war relative to both shipyard counties and counties unrelated to shipyards. This supports the idea that government spending multipliers may be higher when nominal interest rates are pinned to the zero lower bound, as in Christiano, Eichenbaum, and Rebelo (2011), or when there is relatively more slack in the economy. These results hold constant the effects of overall military spending during World War II and the spending received in the 1930s associated with the New Deal.

When attempting to scale these results into an aggregate government spending multiplier, each such exercise gives wildly varying results. The aggregate multiplier on these purchases may have been as high as eight or nine, or it may have been negative. It seems that more research is needed for translating such local multiplier estimates into the aggregate government spending multiplier that most policymakers are interested in.

Although this study has caveats, not least the fact that I do not have hard data on an annual basis that describes the amount of spending in each county, and that I rely on imperfect proxies to identify where the money was likely to be spent, I believe that it is the first to examine local government spending multipliers on purchases (as opposed to transfers) before World War II (when aggregate data collection was harmonized to a lesser degree), comparing the effects in times when capacity was highly utilized and when there was a great deal of slack in the economy. Also, I am able to roughly translate my estimates to an aggregate government spending multiplier (which may be as high as between 7 and 9). Thus, this paper provides evidence consistent with the segment of the literature finding that federal spending can have stimulative effects in local economies.

Table 5.1: Counties Hosting Shipyards Active in 1934

Baltimore (city), MD
Baltimore (county), MD
Cambria, PA
Camden, NJ
Charleston, SC
Dauphin, PA
Delaware, PA
Hudson, NJ
King, WA
Kings, NY
Lackawanna, PA
Lehigh, PA
Los Angeles, CA
Mobile, AL
New Castle, DE
New London, CT
Newport News, VA
Norfolk, MA
Norfolk, VA
Northampton, PA
Philadelphia, PA
Richmond, NY
Rockingham, NH
Sagadahoc, ME
San Francisco, CA
Solano, CA
Suffolk, MA

This list gives counties hosting shipyards active by 1934, the year of passage of the Vinson-Trammell Act of 1934. The list also includes identified major steel suppliers to the shipbuilding industry.

Table 5.2: Counties Hosting Shipyards Receiving USMC Contracts

Alameda, CA	Lorain, OH
Ashtabula, OH	Los Angeles, CA
Baltimore (city), MD	Marin, CA
Baltimore (county), MD	Marinette, WI
Bay, FL	Milwaukee, WI
Brown, WI	Mobile, AL
Camden, NJ	Morgan, AL
Cecil, MD	Multnomah, OR
Chatham, GA	New Castle, DE
Clallam, WA	New Hanover, NC
Clark, WA	New York, NY
Clatsop, OR	Newport News, VA
Cook, IL	Norfolk, MA
Contra Costa, CA	Norfolk, VA
Cumberland, ME	Nueces, TX
Cuyahoga, OH	Orleans, LA
Delaware, PA	Pierce, WA
Door, WI	Providence, RI
Douglas, WI	Richmond, NY
Duval, FL	Sagadahoc, ME
Galveston, TX	Saint Louis, MN
Glynn, GA	San Diego, CA
Harris, TX	San Francisco, CA
Hillsborough, FL	Skagit, WA
Hudson, NJ	Snohomish, WA
Humboldt, CA	St. Tammany, LA
Jackson, MS	Suffolk, MA
Jefferson, LA	Thurston, WA
Jefferson, TX	Tillamook, OR
King, WA	Waldo, ME
Lincoln, OR	Westchester, NY

This list gives counties hosting shipyards that received USMC contracts from 1936 to 1946, as indicated by Fischer (1946).

Table 5.3: Summary Statistics of Main Outcome Variables

Variable	Mean	Std. Dev.	Min	Max	Observations
Δ Number of Manufacturing Establishments	0.04	0.31	−2.22	2.56	9707
Δ Manufacturing Employment	0.06	0.43	−3.86	4.73	8087
Δ Average Employees per Manufacturing Firm	0.06	0.45	−3.86	4.73	9082
Δ Manufacturing Output	0.12	0.40	−3.65	3.98	8087
Δ Manufacturing Wage Payments	0.06	0.45	−3.89	4.16	8087
Δ Average Earnings per Manufacturing Employee	0.01	0.20	−1.64	1.63	8087
Δ Manufacturing Value Added	0.11	0.47	−3.47	16.13	8090
Δ Retail Sales	0.09	0.08	−0.53	0.87	6097
Δ Wholesale Employment	0.20	0.57	−3.00	3.50	2670
Δ Retail Employment	0.15	0.26	−2.35	2.62	3050
Δ Wholesale Net Sales	−0.16	0.81	−13.75	0.50	2757
Δ Average Earnings per Wholesale Employee	−0.11	0.46	−6.21	4.03	2596
Δ Average Earnings per Retail Employee	0.34	0.43	−7.80	8.48	3044
Δ Number of Tax Returns	0.18	0.38	−3.74	3.91	29616

The table gives summary statistics for main outcome variables obtained from the data set constructed by Fishback et al. (2011b).

Table 5.4: Effect of Shipbuilding Program on Manufacturing Output: Robustness Tests

	Baseline	New Deal (Levels) (2SLS)	New Deal (Changes) (2SLS)	New Deal (Two-Year Sum) (2SLS)	County Cluster	Region FE	Propensity Score 1	Propensity Score 2
$1935 \times Shipyard_{1934}$	0.068	0.066	0.259	0.060	0.068	0.068	0.062	0.014
	(0.049)	(0.050)	(0.190)	(0.056)	(0.049)	(0.049)	(0.058)	(0.059)
$1937 \times Shipyard_{1934}$	0.130***	0.127***	0.341**	0.113**	0.130***	0.130***	0.117**	0.080
	(0.040)	(0.041)	(0.150)	(0.045)	(0.044)	(0.040)	(0.046)	(0.050)
$1939 \times Shipyard_{1934}$	0.128**	0.125**	0.368**	0.121**	0.128***	0.128**	0.157***	0.119**
	(0.053)	(0.055)	(0.169)	(0.061)	(0.045)	(0.053)	(0.056)	(0.056)
Num. obs.	6420	6420	6420	6399	6420	6420	2184	1608

Each column reports the coefficients on the $Shipyard_{1934}$ variable interacted with the time fixed effect indicated by the row heading for a given specification of the regression indicated by the column heading. More detailed descriptions of each robustness specification are given in the text. ***, **, and * indicate significance at the 1,5, and 10 percent levels, respectively.

Table 5.5: Effect on Output of Being a County with a High Concentration in Other Industries

Industry	1935	1937	1939
Shipyards	0.068	0.130***	0.128**
Coal	0.057	0.040	0.139
Oil and Gas	−0.020	0.089	−0.009
Other Mining	0.181	0.104	0.049
Mineral Extraction	0.146	0.137***	−0.170
Chemicals	−0.044	−0.008	−0.058
Cigars	−0.049	−0.062	−0.048
Glass	0.046	0.163*	0.046
Bread	0.075	−0.008	0.028
Meat	0.224***	0.039	0.058
Automobiles	0.198***	0.159***	−0.102*
Iron	0.225***	0.272***	0.021
Metals	0.103	0.199***	0.086
Planing Mills	−0.003	0.025	−0.194**
Lumber	0.117**	0.197**	0.068
Boots and Shoes	−0.032	0.031	0.049*
Printing, Publishing, and Engraving	0.093	0.026	0.064
Pulp and Paper	−0.054*	0.048	−0.073
Cotton Textiles	−0.519***	−0.270***	−0.373***
Rubber	0.088***	0.066***	0.091***

Each row in the table reports coefficients on a dummy variable indicating that the county has is in the top 26 for the whole country in employment per population in the given industry interacted with the year indicated by the column heading. All regressions include state fixed effects, a dummy for whether the county is situated on a coast, the proportion of the county considered "urban" in the 1930 census, and whether the county is "industrialized." Standard errors are clustered at the state level. ***, **, and * indicate significance at the 1,5, and 10 percent levels, respectively.

Table 5.6: Sensitivity of Output Results to Exclusion of Individual Shipyards

Sample	1935	1937	1939
All Counties	0.068	0.130***	0.128**
New London, CT	0.077	0.14***	0.138**
Sagadahoc, ME	0.058	0.113***	0.104**
Norfolk, MA	0.066	0.134***	0.123**
Suffolk, MA	0.072	0.133***	0.128**
Rockingham, NH	0.055	0.117***	0.112**
New Castle, DE	0.077	0.128***	0.136**
Camden, NJ	0.065	0.133***	0.118**
Hudson, NJ	0.067	0.126***	0.124**
Cambria, PA	0.075	0.113**	0.146***
Dauphin, PA	0.059	0.124***	0.129**
Delaware, PA	0.079	0.143***	0.142***
Lackawanna, PA	0.081	0.143***	0.128**
Lehigh/Northampton, PA	0.065	0.120***	0.126**
Philadelphia, PA	0.068	0.134***	0.130**
Norfolk, VA	0.034	0.127***	0.118**
Mobile, AL	0.080	0.140***	0.130**
Charleston, SC	0.080	0.133***	0.135**
Baltimore (county), MD	0.072	0.114***	0.132**
Baltimore (city), MD	0.065	0.130***	0.134**
Los Angeles, CA	0.065	0.129***	0.130**
San Francisco, CA	0.070	0.134***	0.130**
Solano, CA	0.070	0.143***	0.138**
King, WA	0.069	0.131***	0.128**
New York, NY	0.068	0.136***	0.123**

Each row in the table reports coefficients on $Shipyard_{1934}$ interacted with the year indicated by the column heading when the row county is excluded from the regression. All regressions include state fixed effects, a dummy for whether the county is situated on a coast, the proportion of the county considered "urban" in the 1930 census, and whether the county is "industrialized." Standard errors are clustered at the state level. ***, **, and * indicate significance at the 1,5, and 10 percent levels, respectively.

Table 5.7: Pre-Vinson-Trammell Act Outcomes

Outcome	$Shipyard_{1934}$	$BordersShipyard_{1934}$
Δ Number of Manufacturing Establishments	−0.024	−0.002
	(0.025)	(0.017)
Δ Manufacturing Employment	0.020	0.032
	(0.024)	(0.020)
Δ Average Employees per Manufacturing Firm	−0.040	−0.006
	(0.054)	(0.032)
Δ Manufacturing Output	−0.071	−0.000
	(0.051)	(0.045)
Δ Manufacturing Wage Payments	-0.073^{*}	−0.010
	(0.038)	(0.034)
Δ Average Earnings per Manufacturing Employee	−0.010	−0.008
	(0.026)	(0.013)
Δ Manufacturing Value Added	-0.095^{*}	0.004
	(0.048)	(0.045)
Δ Retail Sales per capita	0.001	0.002
	(0.005)	(0.005)

This table gives coefficients on $Shipyard_{1934}$ and $BordersShipyard_{1934}$ from regressions on each outcome including only data before 1934. Standard errors clustered at state level are in parentheses. ***, **, and * indicate significance at the 1,5, and 10 percent levels, respectively.

Table 5.8: Effect on Retail Sales Growth of Hosting or Bordering a Shipyard

Independent Variable	Baseline	New Deal Changes (2SLS)	New Deal Levels (2SLS)
$Shipyard_{1934}$	-0.026^{***}	-0.025^{***}	-0.034^{***}
	(0.007)	(0.007)	(0.012)
$I(Year = 1939) * Shipyard_{1934}$	0.038^{***}	0.037^{***}	0.037^{***}
	(0.010)	(0.010)	(0.010)
$BordersShipyard_{1934}$	-0.026^{***}	-0.026^{***}	-0.034^{***}
	(0.007)	(0.007)	(0.009)
$I(Year = 1939) * BordersShipyard_{1934}$	0.045^{***}	0.045^{***}	0.042^{***}
	(0.012)	(0.012)	(0.013)
$I(Year = 1939)$	-0.074^{***}	-0.077^{**}	-0.073^{***}
	(0.007)	(0.037)	(0.006)
Observations	5282	5286	5286
R-Squared	0.342	0.341	0.318

The table gives coefficient estimates from regressions of the average annual change in real retail sales on dummy variables for $Shipyard_{1934}$ and $BordersShipyard_{1934}$ and the interaction of these dummy variables with year fixed effects. Standard errors clustered at state level are in parentheses. $^{***},^{**}, and^{*}$ indicate significance at the 1,5, and 10 percent levels, respectively.

Table 5.9: Effect on Retail Sales of Being a County with a High Concentration in Other Industries

Industry	1939
Shipyards	0.038***
Coal	0.009
Oil and Gas	−0.008
Other Mining	−0.010
Mineral Extraction	0.004
Chemicals	0.038**
Cigars	0.007
Glass	0.048**
Bread	−0.002
Meat	0.027*
Automobiles	−0.023**
Iron	0.001
Metals	0.016
Planing Mills	−0.004
Lumber	−0.016
Boots and Shoes	0.011
Printing, Publishing, and Engraving	0.028
Pulp and Paper	−0.007
Cotton Textiles	0.056***
Rubber	0.020

Each row in the table reports coefficients on a dummy variable indicating that the county has is in the top 26 for the whole country in employment per population in the given industry interacted with the year indicated by the column heading. All regressions include state fixed effects, a dummy for whether the county is situated on a coast, the proportion of the county considered "urban" in the 1930 census, and whether the county is "industrialized." Standard errors are clustered at the state level. ***, **, and * indicate significance at the 1,5, and 10 percent levels, respectively.

Table 5.10: Sensitivity of Retail Sales Results to Exclusion of Individual Shipyards

Sample	$Shipyard_{1934}$
All Counties	0.038***
New London, CT	0.037***
Sagadahoc, ME	0.038***
Norfolk, MA	0.038***
Suffolk, MA	0.037***
Rockingham, NH	0.040***
New Castle, DE	0.038***
Camden, NJ	0.037***
Hudson, NJ	0.038***
Cambria, PA	0.033***
Dauphin, PA	0.038***
Delaware, PA	0.039***
Lackawanna, PA	0.041***
Lehigh/Northampton, PA	0.039***
Philadelphia, PA	0.039***
Norfolk, VA	0.038***
Newport News, VA	0.040***
Mobile, AL	0.036***
Charleston, SC	0.035***
Baltimore (county), MD	0.034***
Baltimore (city), MD	0.036***
Los Angeles, CA	0.039***
San Francisco, CA	0.036***
Solano, CA	0.041***
King, WA	0.038***
New York, NY	0.039***

Each row in the table reports coefficients on $Shipyard_{1934}$ interacted with the year indicated by the column heading when the row county is excluded from the regression. All regressions include state fixed effects, a dummy for whether the county is situated on a coast, the proportion of the county considered "urban" in the 1930 census, and whether the county is "industrialized." Standard errors are clustered at the state level. ***, **, and * indicate significance at the 1,5, and 10 percent levels, respectively.

Table 5.11: Effect on Growth in Various Economic Indicators of Hosting or Bordering a Shipyard

Dep. Variable	Whlse. Emp.	Retail Emp.	Whlsl. Net Sales	Whlsl. Earnings	Retail Earnings
Baseline Controls					
$Shipyard_{1934}$	-0.136^{*}	-0.057^{**}	0.043	-0.020	-0.037
	(0.051)	(0.024)	(0.035)	(0.034)	(0.034)
$BordersShipyard_{1934}$	-0.021	-0.002	0.015	0.003	0.001
	(0.055)	(0.022)	(0.027)	(0.026)	(0.020)
Observations	2351	2678	2413	2287	2685
R-Squared	0.096	0.177	0.247	0.132	0.264
New Deal Changes (2SLS)					
$Shipyard_{1934}$	-0.054	-0.057^{***}	-0.023	0.045	-0.120^{***}
	(0.057)	(0.019)	(0.033)	(0.033)	(0.046)
$BordersShipyard_{1934}$	-0.033	-0.009	0.052	0.078^{***}	-0.063^{**}
	(0.052)	(0.019)	(0.032)	(0.030)	(0.030)
Observations	2350	2677	2413	2287	2685

The table gives coefficient estimates from regressions of changes in the dependent variable (in real terms, where applicable) from 1935 to 1939 on dummy variables for $Shipyard_{1934}$ and $BordersShipyard_{1934}$. Standard errors clustered at state level are in parentheses. ***, **, and * indicate significance at the 1,5, and 10 percent levels, respectively.

Table 5.12: Effect on Growth in Various Indicators of Hosting or Bordering a Shipyard

Dep. Variable	$Shipyard_{1934}$	$BordersShipyard_{1934}$	Observations	R-Squared
Population	-1.124	3.447^{**}	2960	0.246
	(2.302)	(1.390)		
Pct of 14,15 year-olds in school	-1.110	0.567	2959	0.291
	(0.682)	(0.42)		
Pct of 16,17 year-olds in school	-0.642	0.483	2962	0.369
	(1.297)	(0.656)		
Pct of 18,19,20 year-olds in school	0.017	0.628	2960	0.212
	(0.614)	(0.485)		
Share of workforce in:				
Agriculture, Fisheries, and Forestry	-0.004	-0.006	2987	0.083
	(0.014)	(0.011)		
Mining	-0.005	-0.005	2987	0.037
	(0.007)	(0.004)		
Construction	-0.013	0.003	2987	0.068
	(0.010)	(0.009)		
Durables Mfg.	0.012	-0.008	2987	0.025
	(0.015)	(0.007)		
Nondurables Mfg.	-0.005	-0.003	2987	0.061
	(0.007)	(0.006)		
Transportation	0.000	0.000	2987	0.045
	(0.008)	(0.004)		
Communication	0.000	-0.002	2987	0.022
	(0.002)	(0.001)		
Utilities	0.000	-0.002	2987	0.013
	(0.003)	(0.002)		
Wholesale	0.001	-0.001	2987	0.021
	(0.003)	(0.002)		
Retail	-0.001	0.008	2987	0.027
	(0.008)	(0.005)		
Financials	0.003	0.004^{*}	2987	0.025
	(0.003)	(0.002)		
Business Services	-0.001	-0.002	2987	0.019
	(0.003)	(0.002)		
Personal Services	-0.012	0.000	2987	0.026
	(0.008)	(0.006)		
Entertainment	0.004	-0.001	2987	0.023
	(0.002)	(0.002)		
Professional Services	0.011	0.004	2987	0.021
	(0.007)	(0.005)		
Government	0.011	0.010^{**}	2987	0.027
	(0.009)	(0.005)		
Shipbuilding and Repair	0.023^{*}	0.003	2987	0.095
	(0.012)	(0.002)		

The table gives coefficient estimates from regressions of changes in the dependent variable from 1930 to 1940 on dummy variables for $Shipyard_{1934}$ and $BordersShipyard_{1934}$. Ordinary least squares regressions include a full set of control variables. Standard errors clustered at state level in parentheses. ***, **, and * indicate significance at the 1,5, and 10 percent levels, respectively.

Table 5.13: Results from Regressions based on 1935-1936 Consumer Survey

Independent Variable	Total Consumption	Household Income
Shipyard	1244.50***	1353.94
	(299.98)	(1336.67)
Borders Shipyard	−360.99	−867.56
	(312.36)	(1278.86)
Overlap	0.93***	0.36
	(0.29)	(1.27)
Shipyard*Overlap	2.33**	2.63
	(0.94)	(3.50)
Borders Shipyard*Overlap	0.42	−1.90
	(0.86)	(3.77)

The table gives coefficient estimates from regressions of Equation 5.6. Ordinary least squares regressions include controls for the age and age squared of the husband and wife of the household as well as a dummy for whether the household is not white and state dummies. *Overlap* and its interaction terms are described in the text. Standard errors are in parentheses. ***, **, and * indicate significance at the 1,5, and 10 percent levels, respectively.

Table 5.14: Results from Regressions based on 1935-1936 Consumer Survey

Indep. Var.	Total Consump.	Housing	Housing Op.	Medical Care	Recreation	Tobacco	Reading	Education
Shipyard	1212.02***	369.22***	43.90	40.62	39.07	22.79**	15.19***	42.27
	(291.93)	(65.16)	(45.34)	(37.26)	(35.73)	(9.57)	(4.31)	(37.51)
Overlap	0.93***	0.17***	0.16***	0.00	0.06*	0.04***	0.02***	0.02
	(0.28)	(0.06)	(0.04)	(0.04)	(0.03)	(0.01)	(0.00)	(0.04)
Shipyard*Overlap	2.34**	0.26	0.47***	0.18*	0.31***	0.04*	0.01	−0.35***
	(0.91)	(0.17)	(0.12)	(0.10)	(0.09)	(0.03)	(0.01)	(0.10)

Indep. Var.	Occupational Exp.	Gifts	Food	Autos	Clothing	Travel	Personal Care	Equipment
Shipyard	12.53	22.30*	363.94***	−4.24	75.60	76.32***	21.30***	4.04
	(30.72)	(13.57)	(64.85)	(66.03)	(49.83)	(15.87)	(7.85)	(6.30)
Overlap	0.03	0.06***	0.26***	0.32***	0.27***	0.00	0.04***	−0.00
	(0.03)	(0.01)	(0.06)	(0.06)	(0.05)	(0.02)	(0.01)	(0.01)
Shipyard*Overlap	−0.02	0.02	0.66***	−0.09	0.04	0.01	0.05**	−0.01
	(0.08)	(0.04)	(0.17)	(0.17)	(0.13)	(0.04)	(0.02)	(0.02)

The table gives coefficient estimates from regressions of Equation 5.6. Ordinary least squares regressions include controls for the age and age squared of the husband and wife of the household and household income as well as a dummy for whether the household is not white and state dummies. *Overlap* and its interaction terms are described in the text. Standard errors are in parentheses. ***, **, and * indicate significance at the 1,5, and 10 percent levels, respectively.

Table 5.15: Long Run (Spanning World War II) Effects of Shipbuilding Spending

Dep. Variable	$Shipyard_{1934}$	$BordersShipyard_{1934}$
OLS-Baseline Controls		
Δ Manufacturing Establishments (1933-1947)	-0.084	0.020
	(0.064)	(0.048)
Δ Manufacturing Value Added (1933-1947)	-0.169^{**}	-0.076
	(0.079)	(0.060)
Δ Auto Registrations (1936-1947)	0.027	0.088^{***}
	(0.034)	(0.017)
Δ Retail Sales (1933-1948)	-0.085	0.034
	(0.053)	(0.024)
Δ Retail Employment (1935-1948)	-0.081	0.049^{**}
	(0.055)	(0.025)
Δ Wholesale Net Sales (1935-1948)	0.043	0.037
	(0.069)	(0.046)
Δ Wholesale Employment (1935-1948)	-0.145^{*}	-0.002
	(0.082)	(0.071)
OLS-Controls Include WWII Spending		
Δ Manufacturing Establishments (1933-1947)	-0.051	0.044
	(0.062)	(0.043)
Δ Manufacturing Value Added (1933-1947)	-0.151^{**}	-0.063
	(0.075)	(0.058)
Δ Auto Registrations (1936-1947)	0.003	0.060^{***}
	(0.030)	(0.015)
Δ Retail Sales (1933-1948)	-0.096^{*}	0.022
	(0.050)	(0.025)
Δ Retail Employment (1935-1948)	-0.102^{**}	0.033
	(0.050)	(0.025)
Δ Wholesale Net Sales (1935-1948)	0.049	0.040
	(0.072)	(0.046)
Δ Wholesale Employment (1935-1948)	-0.102	0.042
	(0.082)	(0.062)

The table gives coefficient estimates from regressions of changes in the dependent variables over the years given on dummy variables for $Shipyard_{1934}$ and $BordersShipyard_{1934}$. A full set of control variables is included in the regression. Robust standard errors in parentheses. ***, **, and * indicate significance at the 1,5, and 10 percent levels, respectively.

Table 5.16: Effects of USMC Spending (1936-1945)

Dep. Variable	$Shipyard_{USMC}$	$BordersShipyard_{USMC}$	$ShipSpending_{HOST}$	$ShipSpending_{BORDER}$
OLS-Baseline Controls				
Δ Manufacturing Establishments (1933-1947)	0.095*	0.080**	0.161	0.329
	(0.056)	(0.040)	(0.295)	(0.222)
Δ Manufacturing Value Added (1933-1947)	0.069	0.051	0.147	0.082
	(0.083)	(0.055)	(0.408)	(0.299)
Δ Auto Registrations (1936-1947)	0.082***	0.087***	0.126	0.470***
	(0.026)	(0.013)	(0.153)	(0.070)
Δ Retail Sales (1933-1948)	0.032	0.068***	−0.126	0.365***
	(0.040)	(0.019)	(0.224)	(0.107)
Δ Retail Employment (1935-1948)	0.030	0.076***	−0.136	0.389***
	(0.041)	(0.022)	(0.232)	(0.123)
Δ Wholesale Net Sales (1935-1948)	0.025	0.014	0.148	−0.013
	(0.061)	(0.042)	(0.336)	(0.224)
Δ Wholesale Employment (1935-1948)	−0.069	0.075	−0.674*	0.305
	(0.070)	(0.055)	(0.373)	(0.302)
OLS-Controls Include WWII Spending				
Δ Manufacturing Establishments (1933-1947)	0.109**	0.102***	0.141	0.444
	(0.055)	(0.036)	(0.287)	(0.200)
Δ Manufacturing Value Added (1933-1947)	0.067	0.068	0.065	0.152
	(0.084)	(0.056)	(0.406)	(0.306)
Δ Auto Registrations (1936-1947)	0.040*	0.069***	−0.038	0.378***
	(0.024)	(0.014)	(0.139)	(0.073)
Δ Retail Sales (1933-1948)	−0.009	0.056***	−0.315	0.309***
	(0.039)	(0.020)	(0.212)	(0.112)
Δ Retail Employment (1935-1948)	−0.024	0.060***	−0.375*	0.302**
	(0.039)	(0.023)	(0.216)	(0.129)
Δ Wholesale Net Sales (1935-1948)	0.022	0.023	0.096	0.009
	(0.064)	(0.047)	(0.342)	(0.252)
Δ Wholesale Employment (1935-1948)	−0.059	0.098*	−0.715*	0.434
	(0.070)	(0.055)	(0.371)	(0.304)

The table gives coefficient estimates from regressions of changes in the dependent variables over the years given on dummy variables for $Shipyard_{USMC}$ and $BordersShipyard_{USMC}$ (first two columns) or the total nominal amount spent over 1936 to 1945 on ship contracts (second two columns). Control variables are specified as indicated in the table. Robust standard errors are in parentheses. ***, **, and * indicate significance at the 1,5, and 10 percent levels, respectively.

Table 5.17: Effect of "Per-County" Shipyard Spending on Growth in Manufacturing Output

$Shipyard_{1934}$	-0.035^{**}
	(0.014)
$I(Year = 1935)*$"Per-County" Spending	0.293^{**}
	(0.116)
$I(Year = 1937)*$"Per-County" Spending	0.844^{***}
	(0.100)
$I(Year = 1939)*$"Per-County" Spending	1.051^{***}
	(0.097)

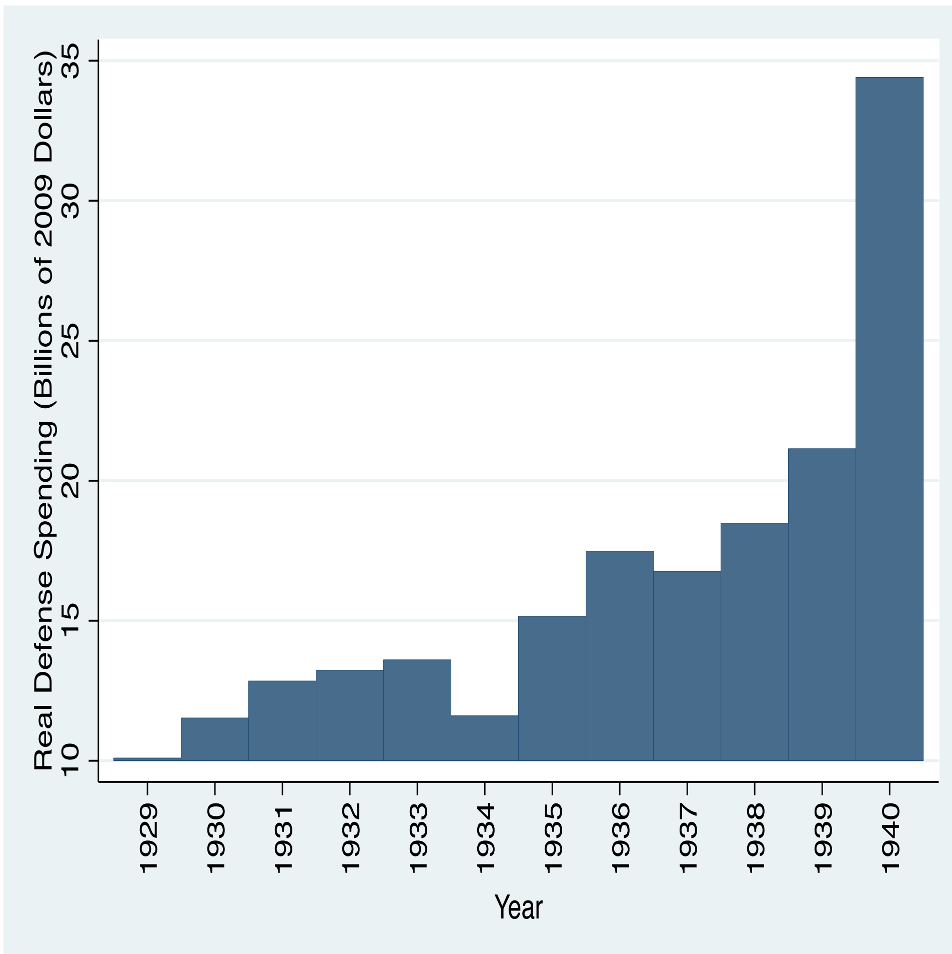
This table gives the coefficient estimates from a regression of the two year change in manufacturing output on dummy variables for *Shipyard* and *BordersShipyard* and interaction of "per-county" spending associated with the Vinson-Trammell Act (defined in the text) interacted with year fixed effects. Standard errors clustered at state level are in parentheses. ***, **, and * indicate significance at the 1,5, and 10 percent levels, respectively.

Table 5.18: Granular Contributions from Shipyard and Non-Shipyard Counties

Year	Shipyard	Borders Shipyard	No Shipyard	Granular Residual
1933	−0.011	−0.000	−0.011	−0.021
1935	−0.008	0.004	0.027	0.022
1937	−0.022	0.002	0.003	−0.018
1939	0.008	−0.004	−0.038	−0.034

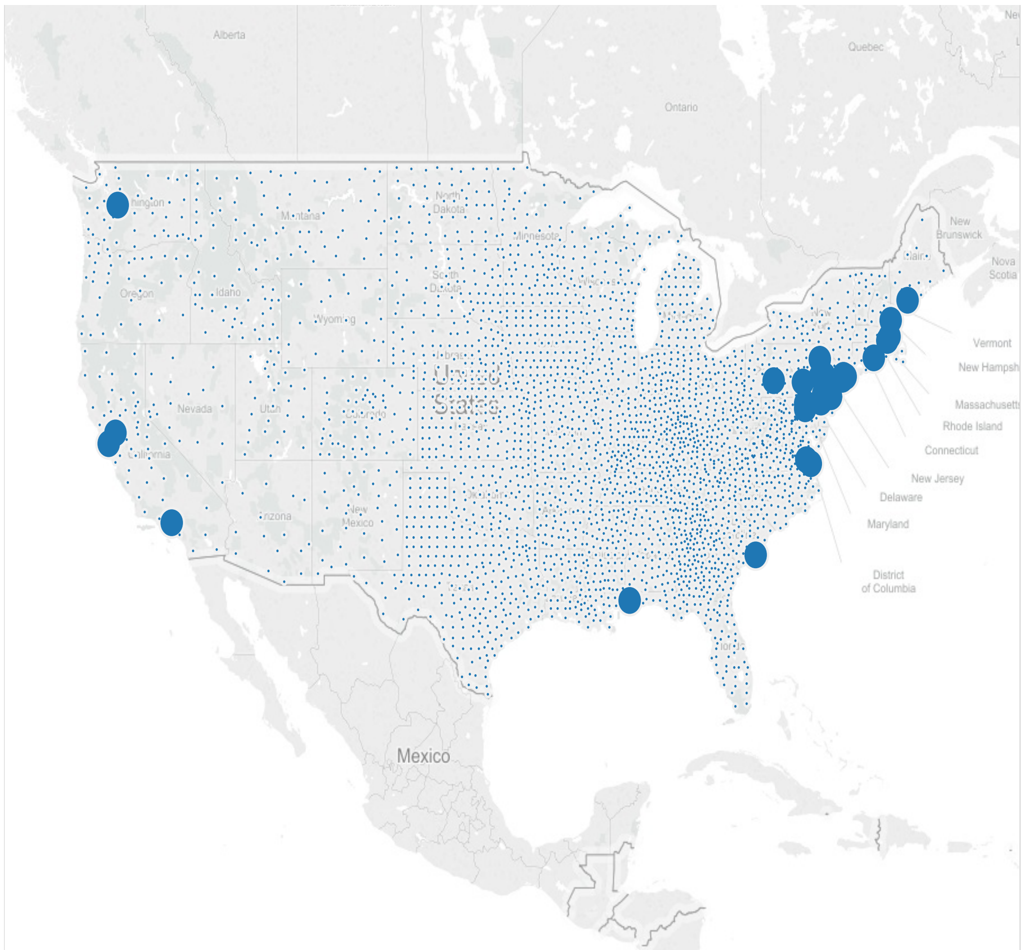
Each cell in the first three columns of the table represents the contribution towards the overall "granular residual" (found in the fourth column), as described by Gabaix (2011) for the year given by the row header. See the text for a description of how the granular residual is computed.

Figure 5.1: Evolution of Defense Spending by the Federal Government in the 1930s



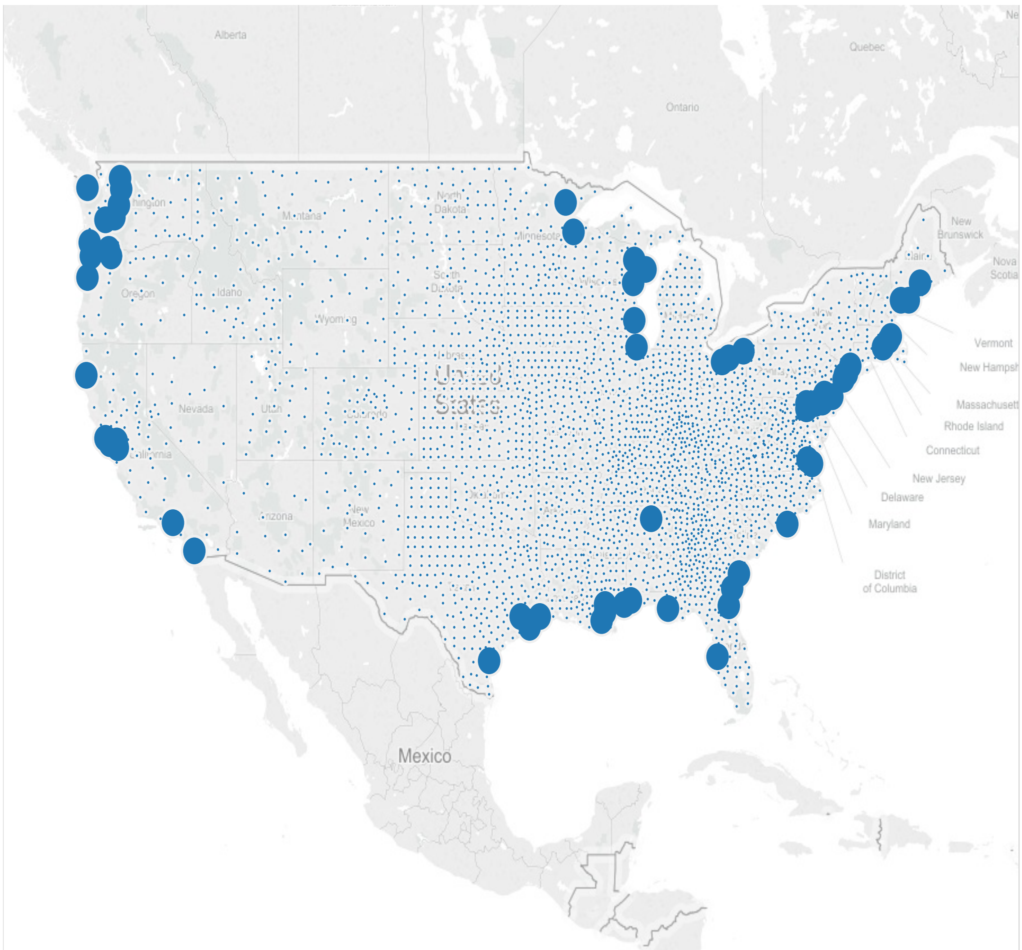
The figure shows annual real defense spending (in 2009 dollars) for the years from 1929-1940. The source is BEA Series A824RC1A027NBEA deflated by the GDP deflator.

Figure 5.2: Locations of Shipyards Active in 1934



Large dots indicate the locations of shipyards and major shipyard suppliers, as indicated in Lane (1951) and contemporary news sources.

Figure 5.3: Locations of Shipyards Receiving USMC Contracts in 1936-1946



Large dots indicate the locations of shipyards receiving USMC contracts from 1936 to 1946, as indicated in Fischer (1946).

Figure 5.4: Additional Growth in Total Manufacturing Output Associated with Shipyard Counties

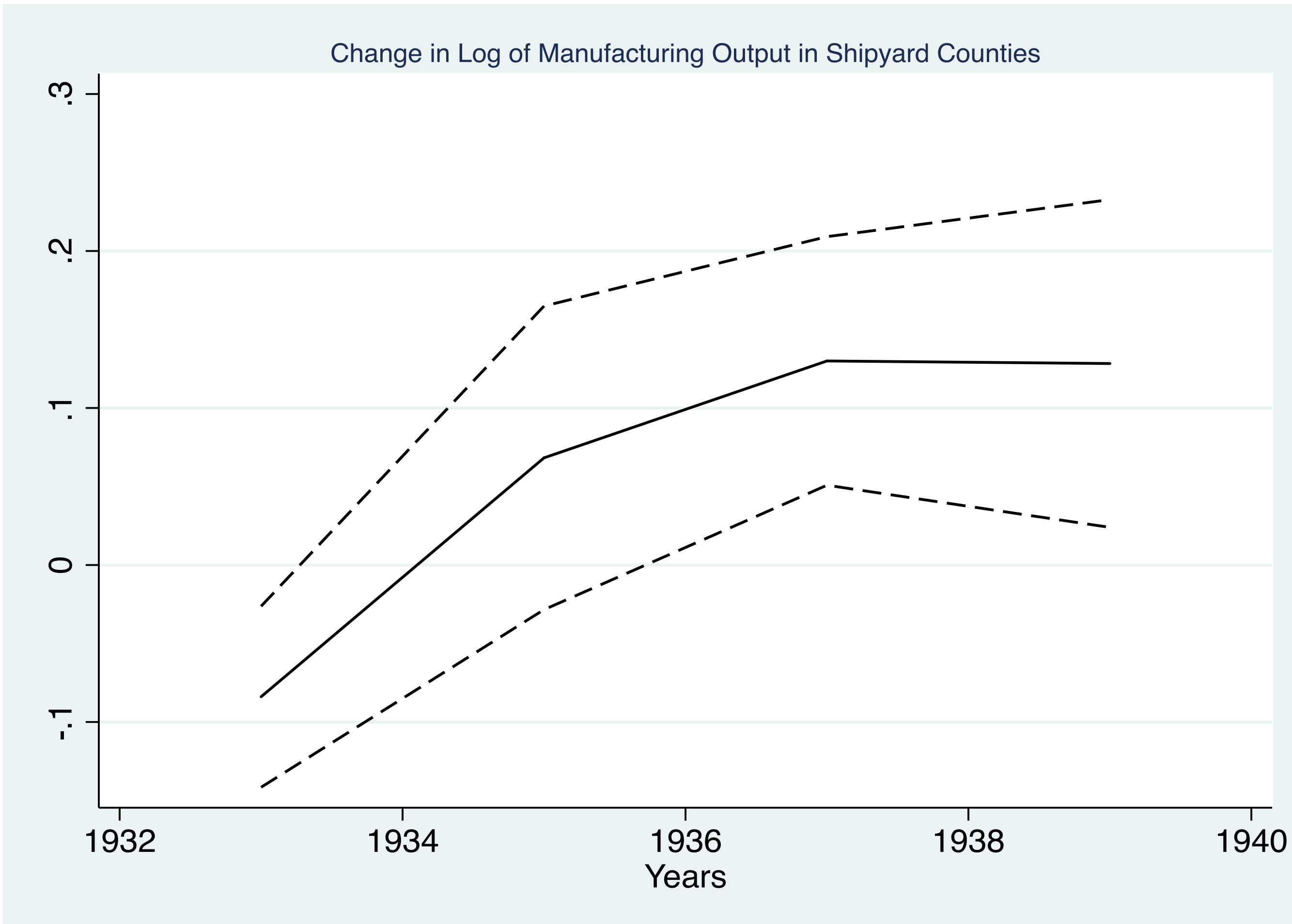


Figure plots coefficients on $Shipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $Shipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is "industrialized," on a coast, and its urban percentage in 1930.

Figure 5.5: Additional Growth in Total Manufacturing Output Associated with Shipyard Border Counties

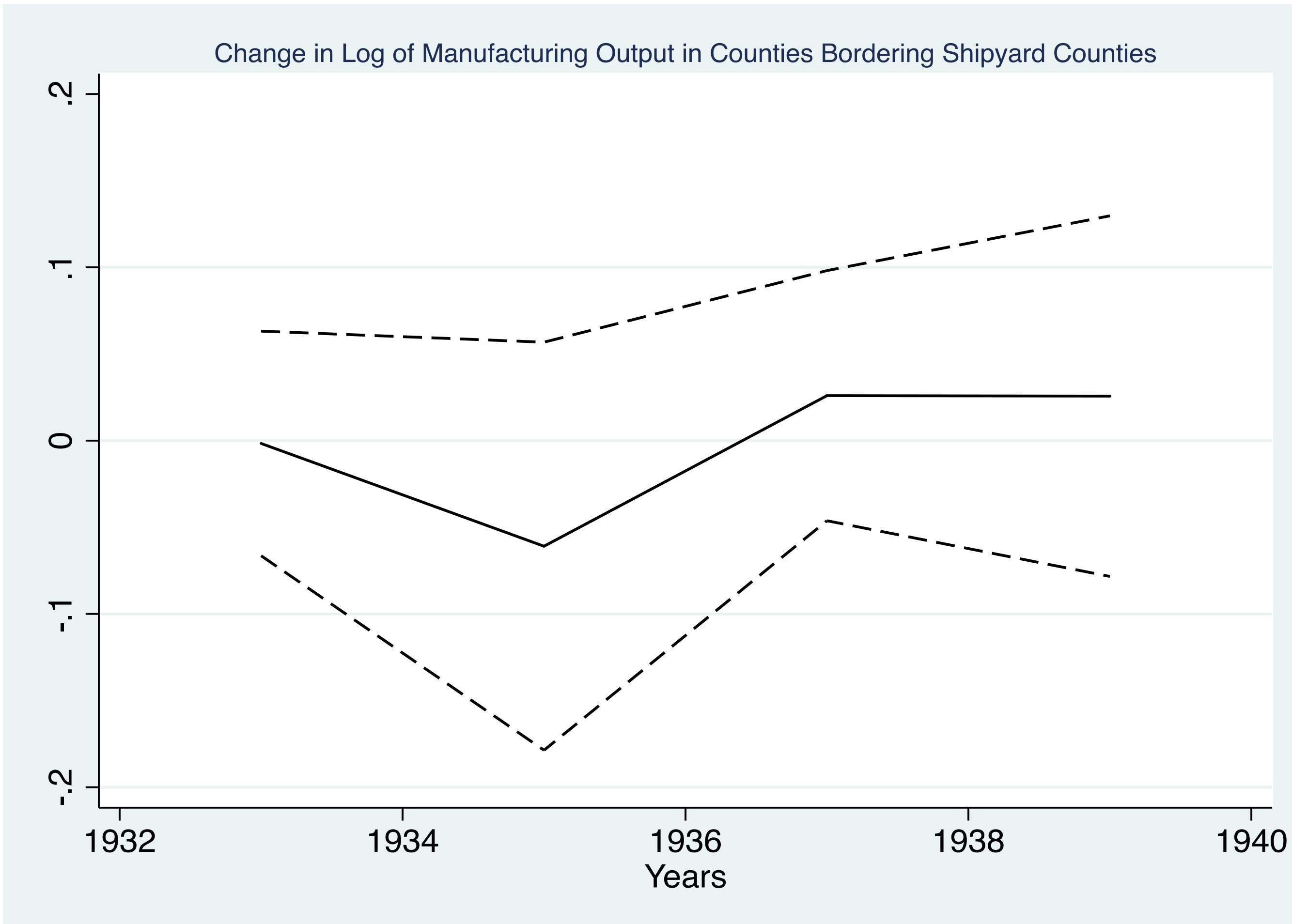


Figure plots coefficients on $BordersShipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $BordersShipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is "industrialized," on a coast, and its urban percentage in 1930.

Figure 5.6: Additional Growth in Total Manufacturing Value Added Associated with Shipyard Counties

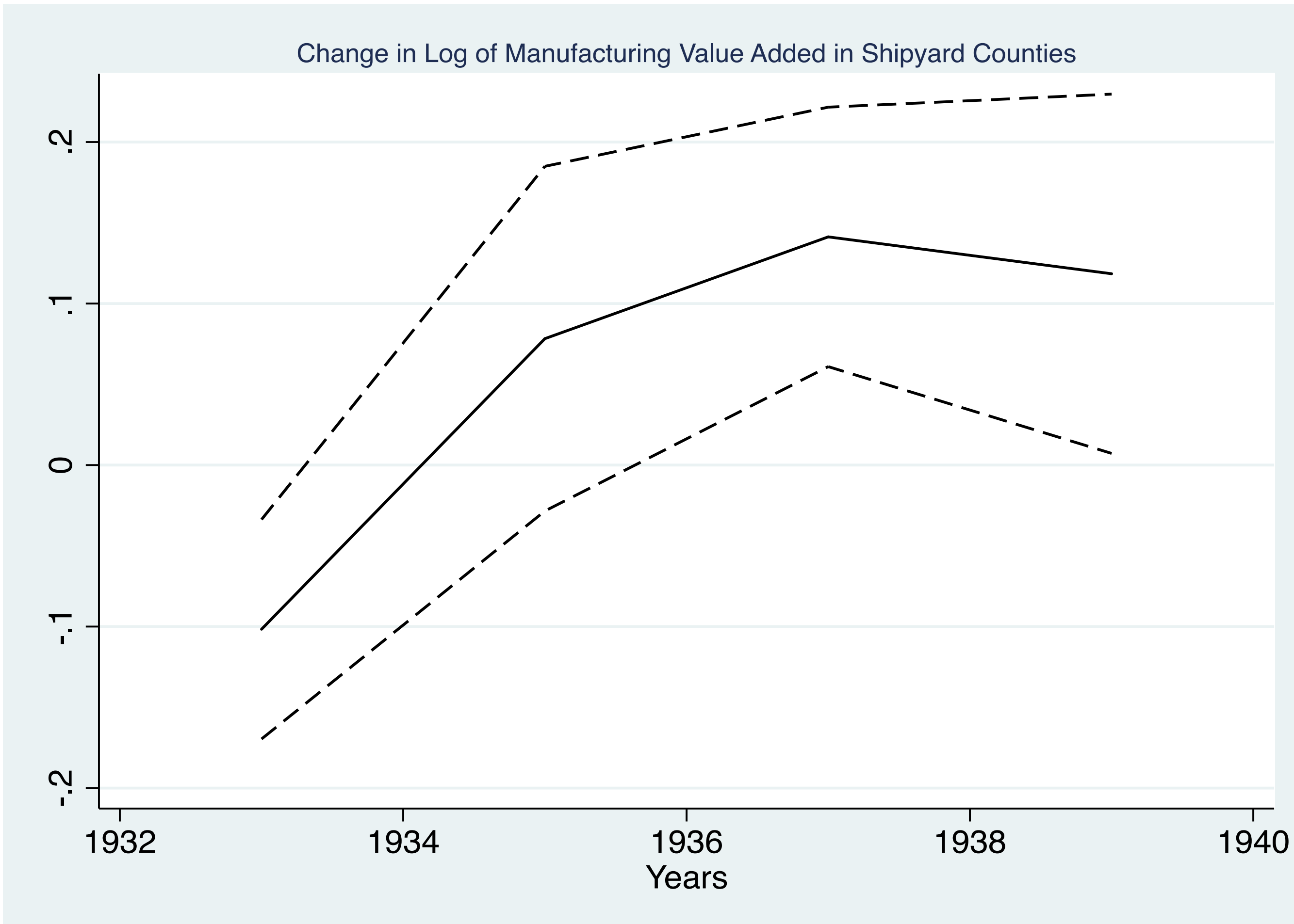


Figure plots coefficients on $Shipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $Shipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is “industrialized,” on a coast, and its urban percentage in 1930.

Figure 5.7: Additional Growth in Total Manufacturing Value Added Associated with Shipyard Border Counties

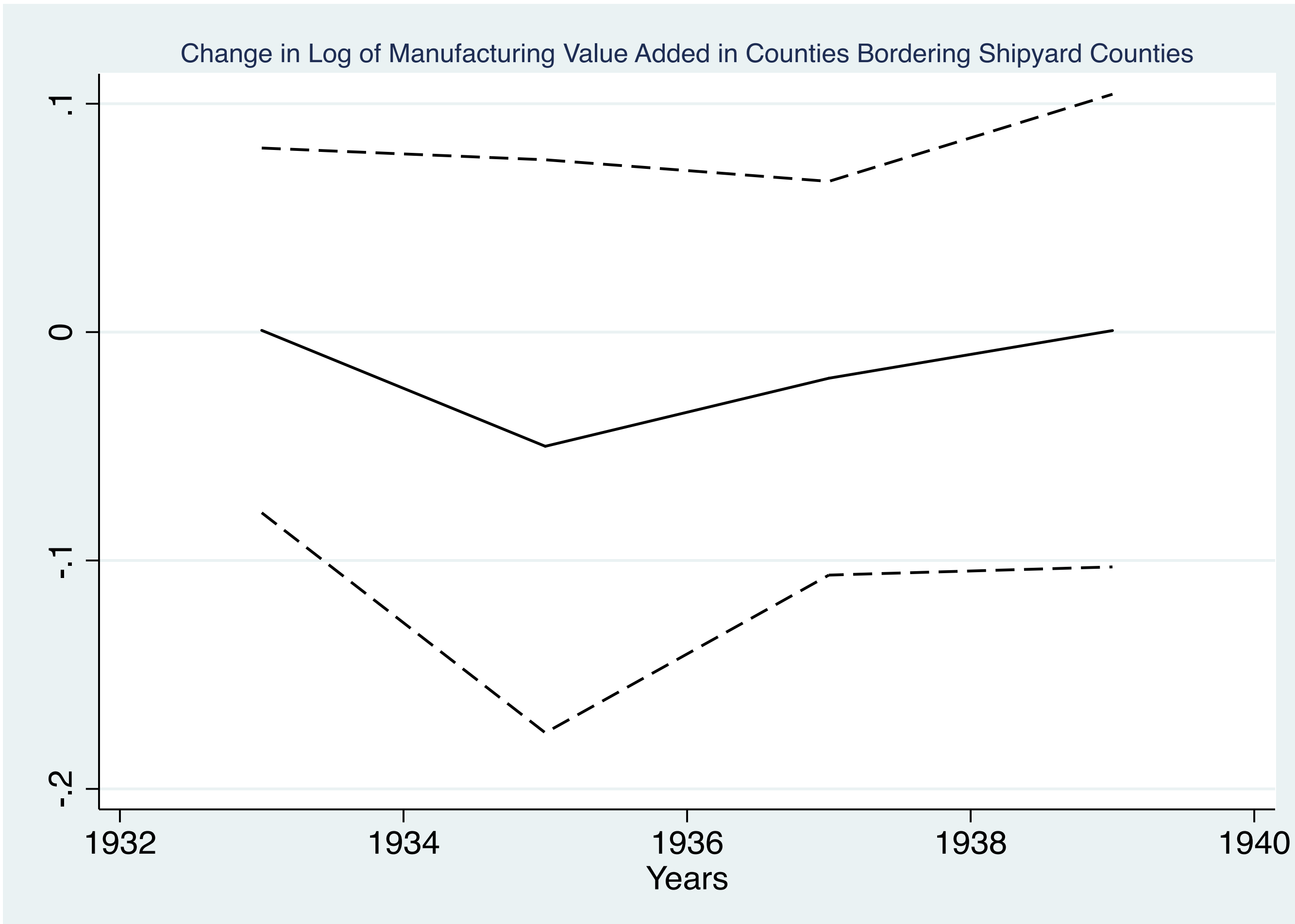


Figure plots coefficients on $BordersShipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $BordersShipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is "industrialized," on a coast, and its urban percentage in 1930.

Figure 5.8: Additional Growth in Manufacturing Employees Associated with Shipyard Counties

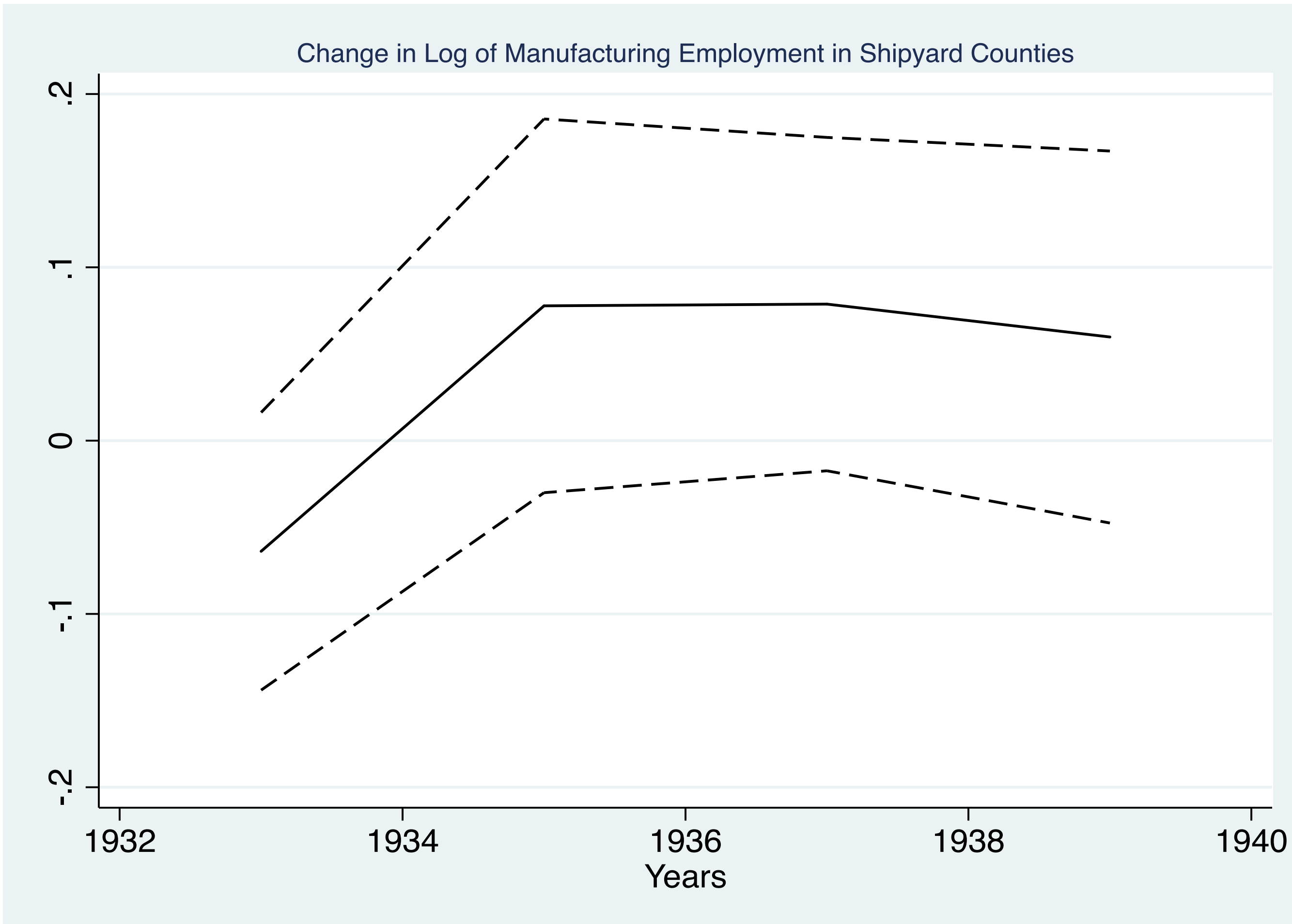


Figure plots coefficients on $Shipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $Shipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is "industrialized," on a coast, and its urban percentage in 1930.

Figure 5.9: Additional Growth in Manufacturing Employees Associated with Shipyard Border Counties

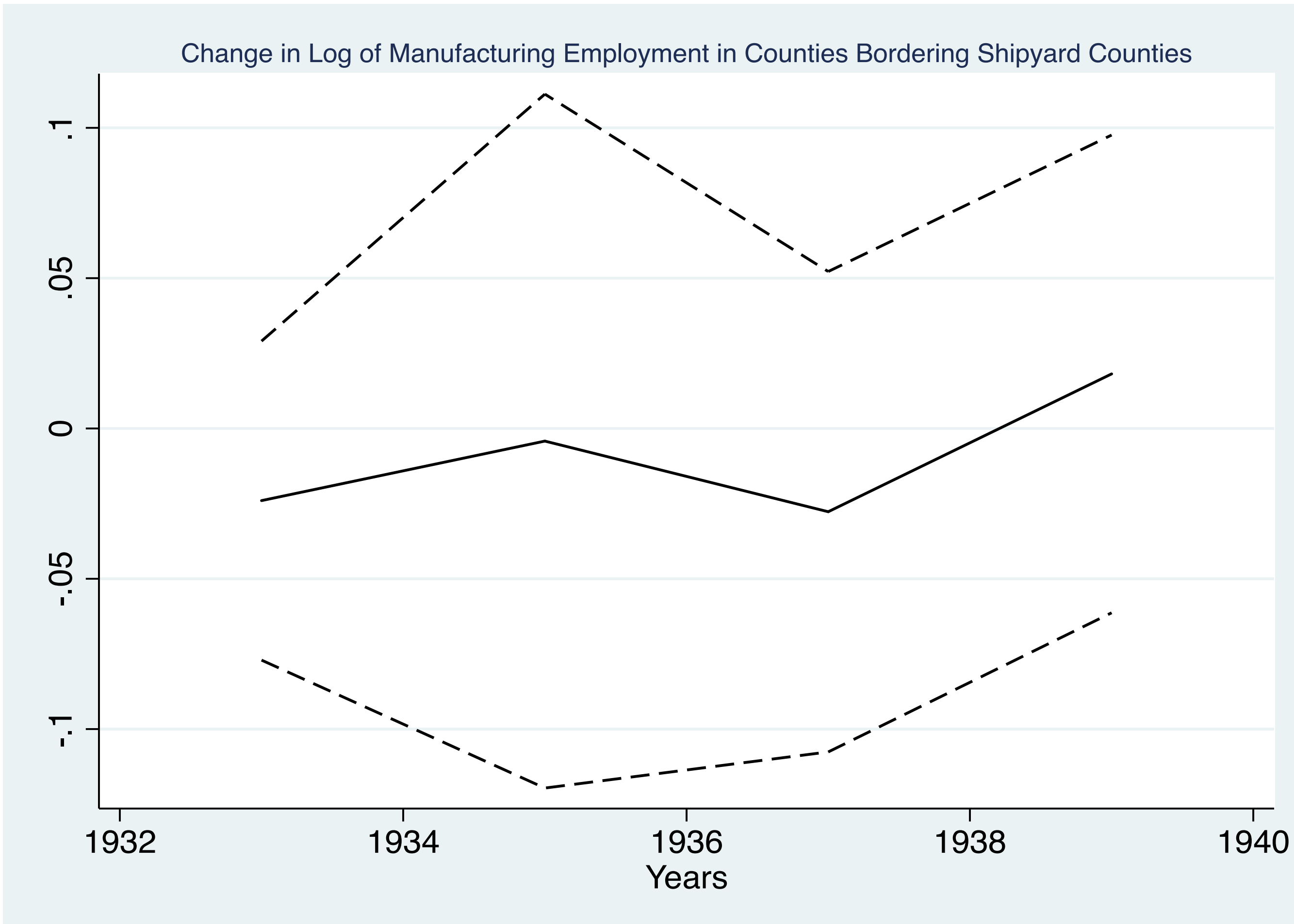


Figure plots coefficients on $BordersShipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $BordersShipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is "industrialized," on a coast, and its urban percentage in 1930.

Figure 5.10: Additional Growth in Total Manufacturing Wage Payments Associated with Shipyard Counties

Figure plots coefficients on $Shipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $Shipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is "industrialized," on a coast, and its urban percentage in 1930.

Figure 5.11: Additional Growth in Total Manufacturing Wage Payments Associated with Shipyard Border Counties

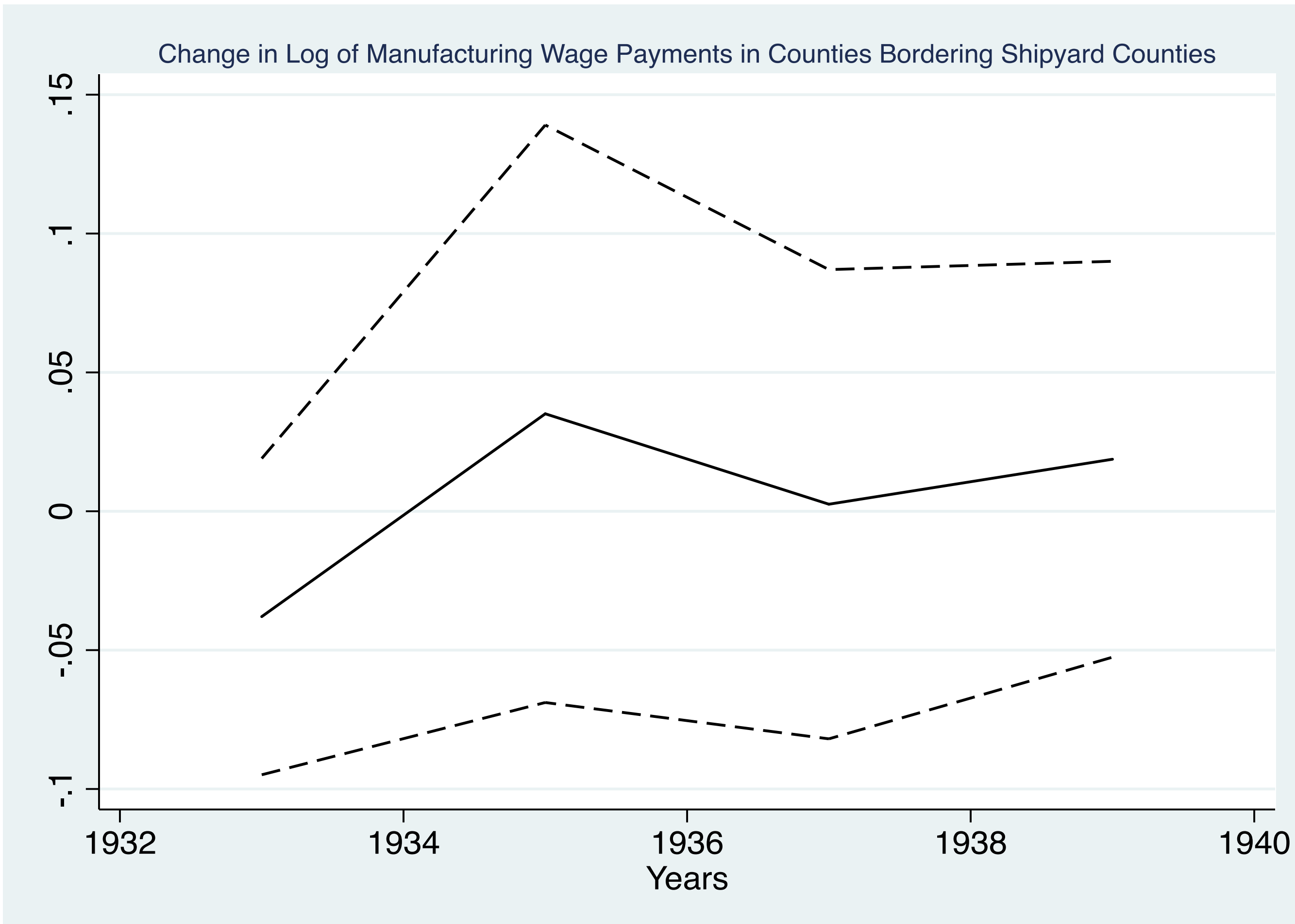


Figure plots coefficients on $BordersShipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $BordersShipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is "industrialized," on a coast, and its urban percentage in 1930.

Figure 5.12: Additional Growth in Average Earnings per Manufacturing Employee Associated with Shipyard Counties

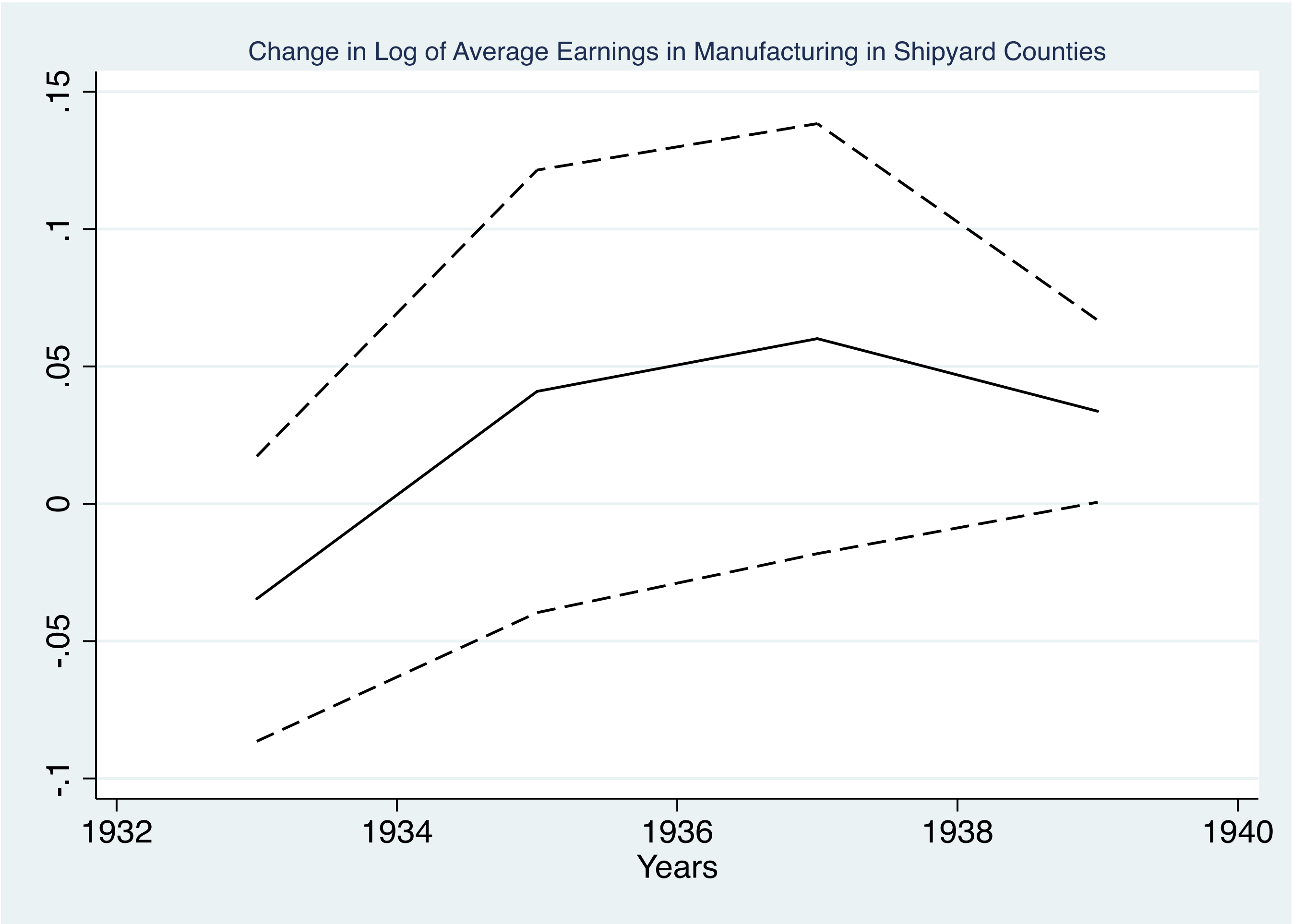


Figure plots coefficients on $Shipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $Shipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is “industrialized,” on a coast, and its urban percentage in 1930.

Figure 5.13: Additional Growth in Average Earnings per Manufacturing Employee Associated with Shipyard Border Counties

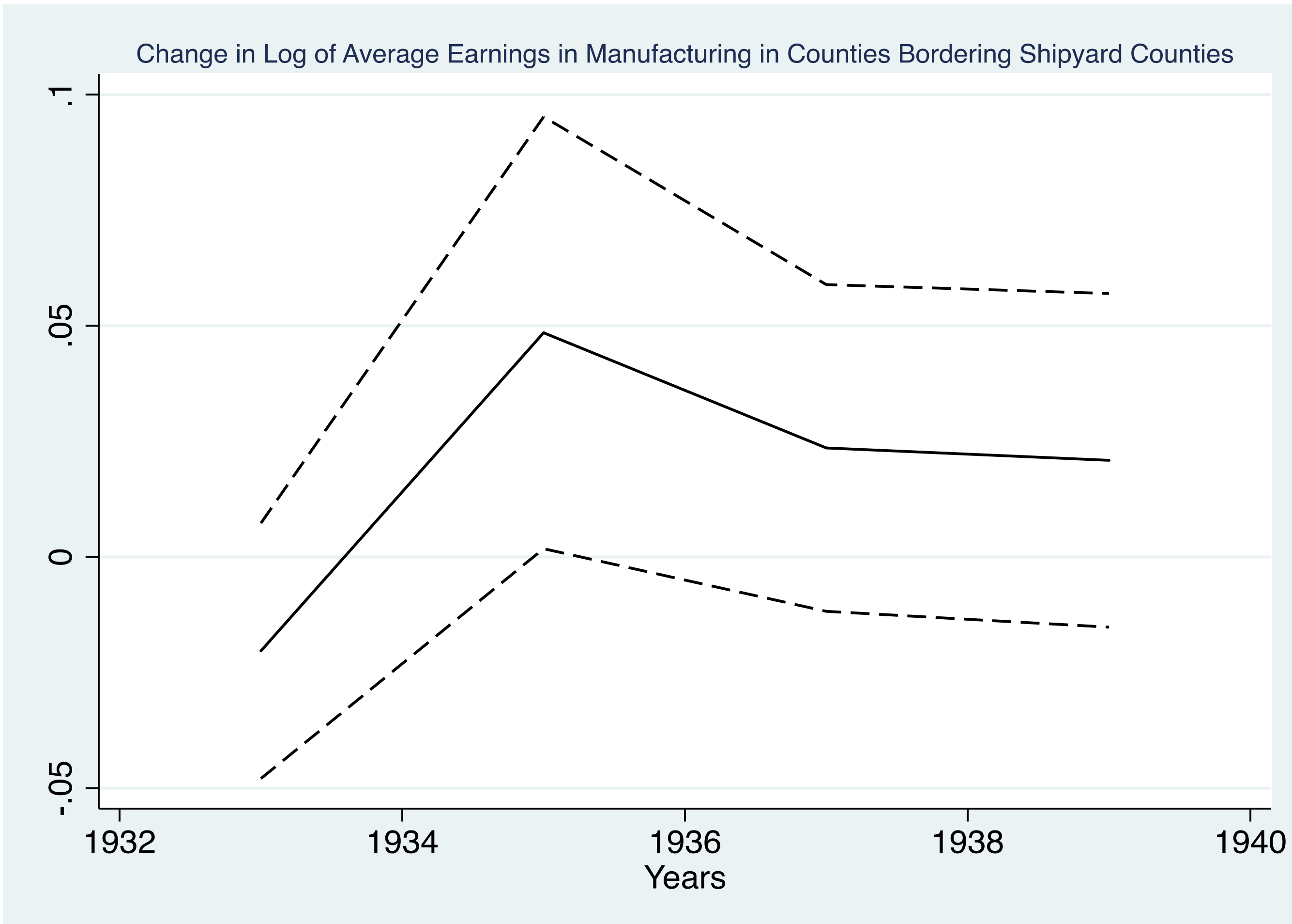


Figure plots coefficients on $BordersShipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $BordersShipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is "industrialized," on a coast, and its urban percentage in 1930.

Figure 5.14: Additional Growth in Number of Manufacturing Establishments Associated with Shipyard Counties

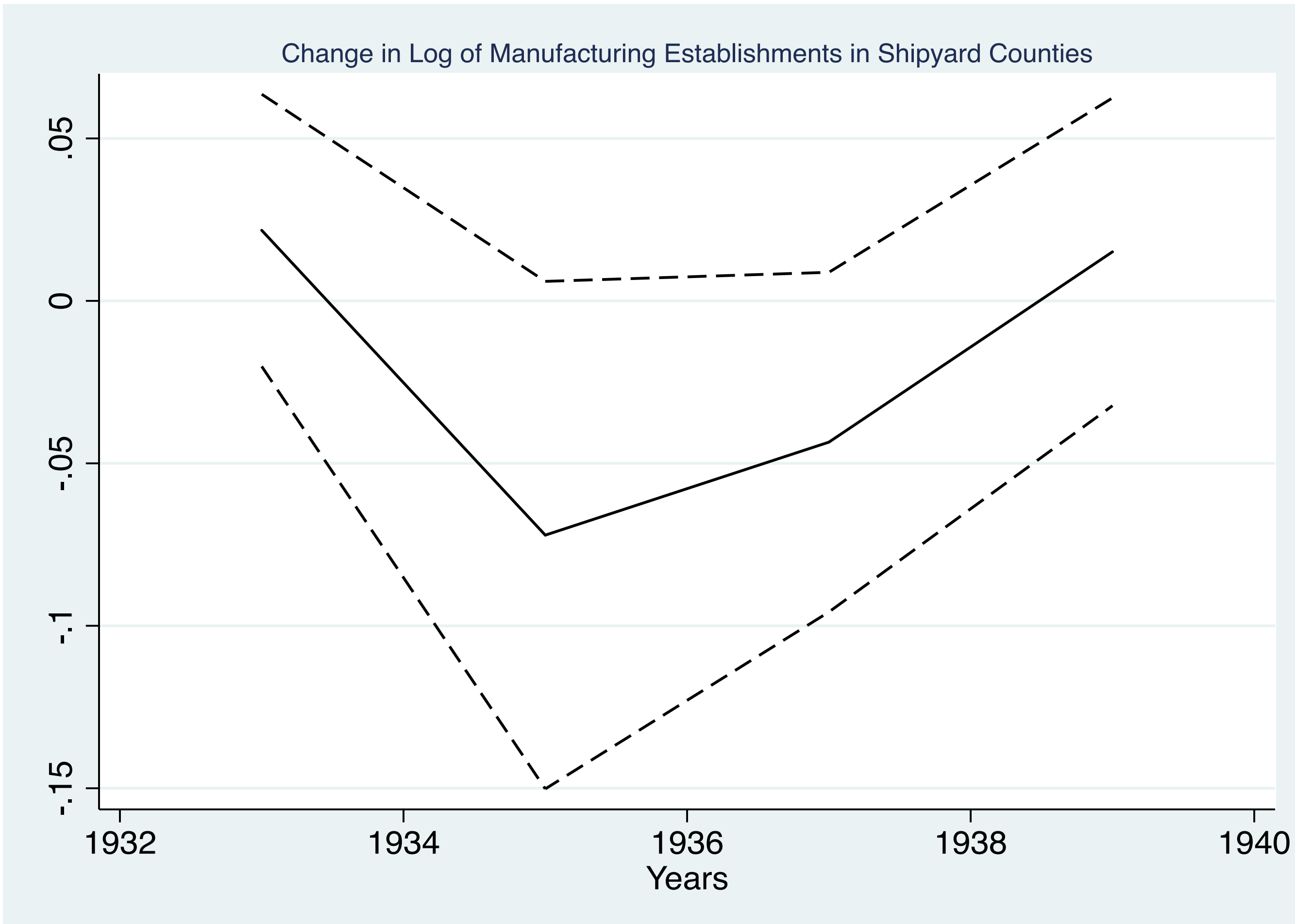


Figure plots coefficients on $Shipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $Shipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is "industrialized," on a coast, and its urban percentage in 1930.

Figure 5.15: Additional Growth in Number of Manufacturing Establishments Associated with Shipyard Border Counties

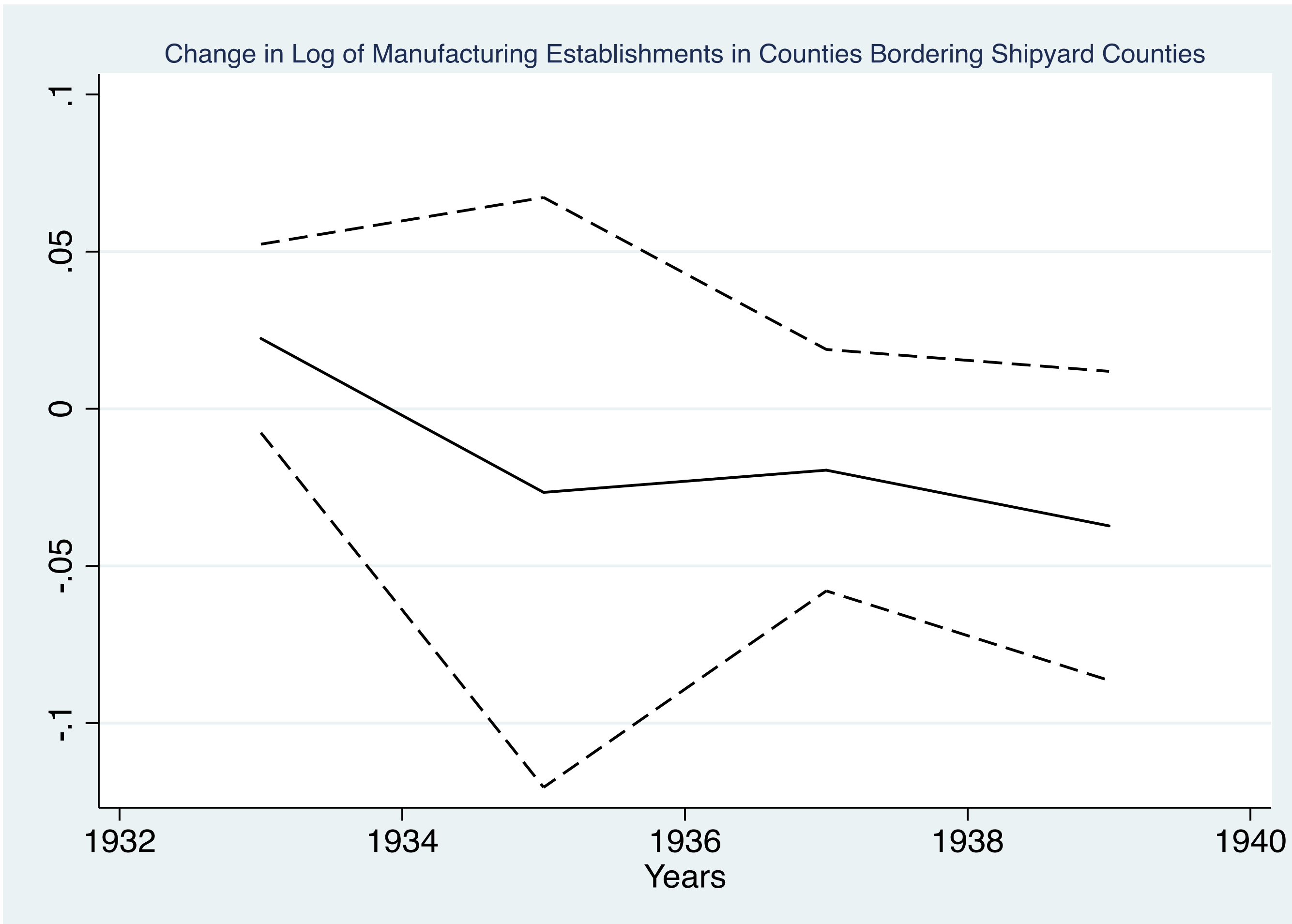


Figure plots coefficients on $BordersShipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $BordersShipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is "industrialized," on a coast, and its urban percentage in 1930.

Figure 5.16: Additional Growth in Average Employees per Establishment Associated with Shipyard Counties

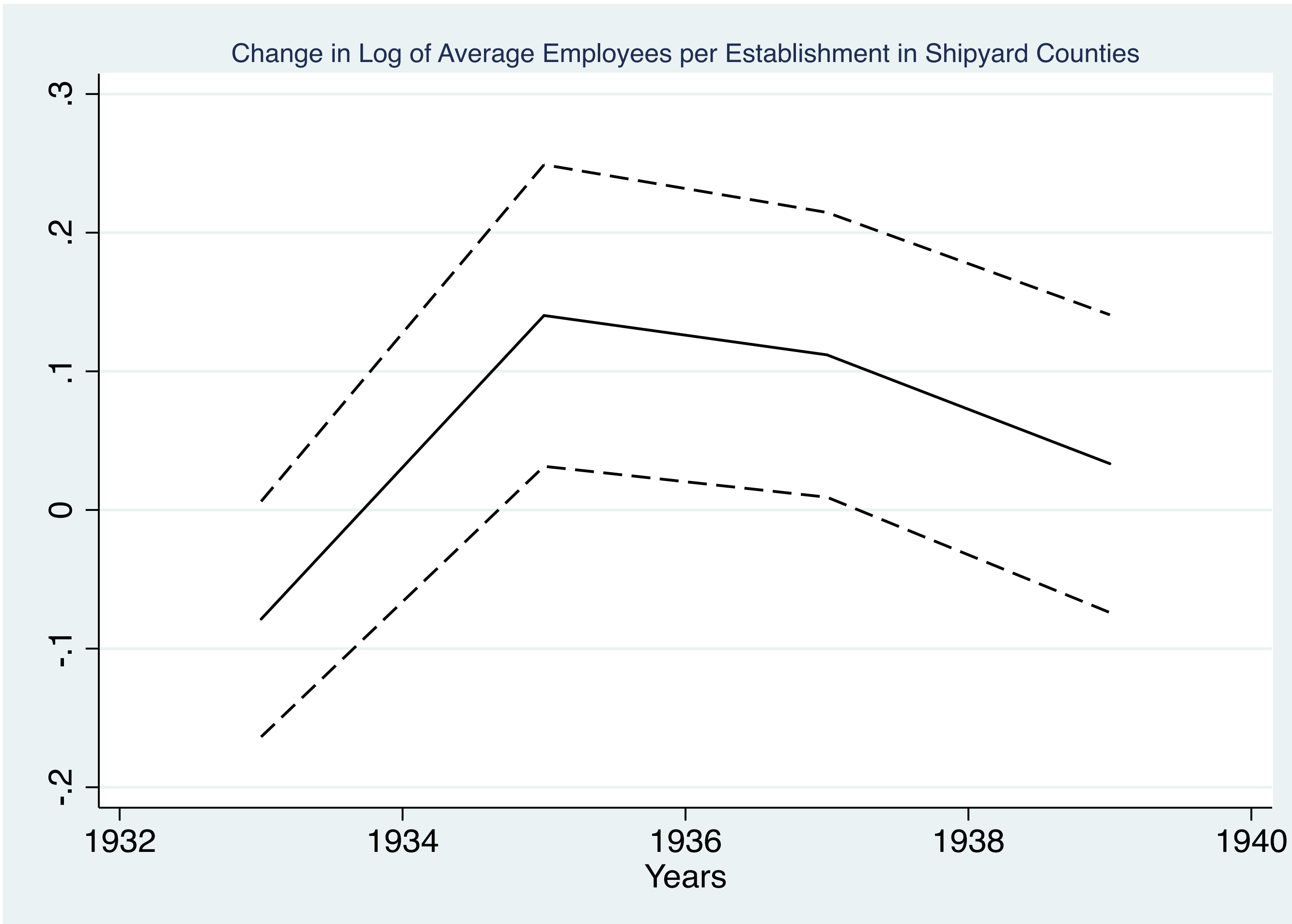


Figure plots coefficients on $Shipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $Shipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is "industrialized," on a coast, and its urban percentage in 1930.

Figure 5.17: Additional Growth in Average Employees per Establishment Associated with Shipyard Border Counties

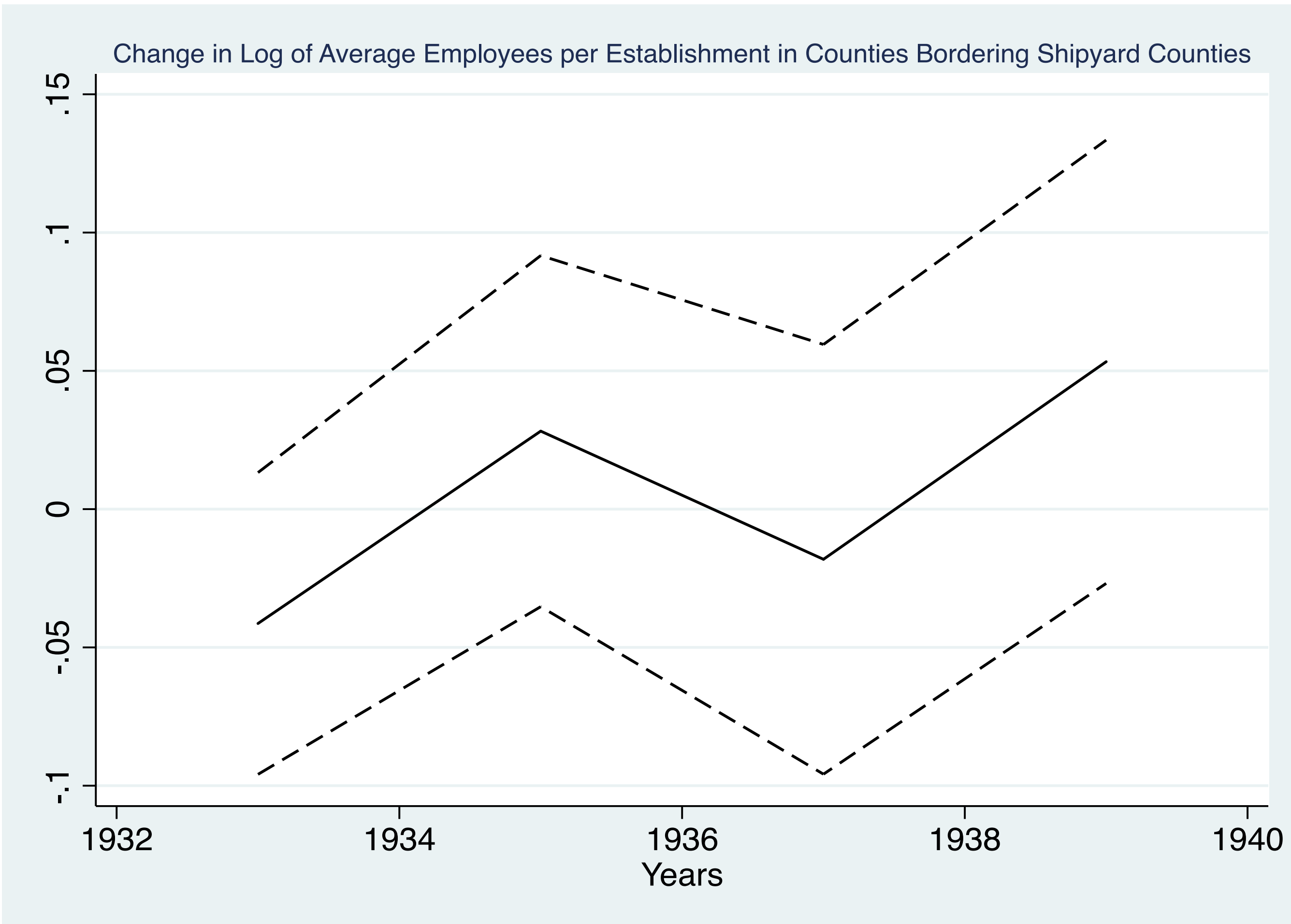


Figure plots coefficients on $BordersShipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $BordersShipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is "industrialized," on a coast, and its urban percentage in 1930.

Figure 5.18: Additional Growth in Labor Productivity in Manufacturing Firms Associated with Shipyard Counties

Figure plots coefficients on $Shipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $Shipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is "industrialized," on a coast, and its urban percentage in 1930.

Figure 5.19: Additional Growth in Labor Productivity in Manufacturing Firms Associated with Shipyard Border Counties

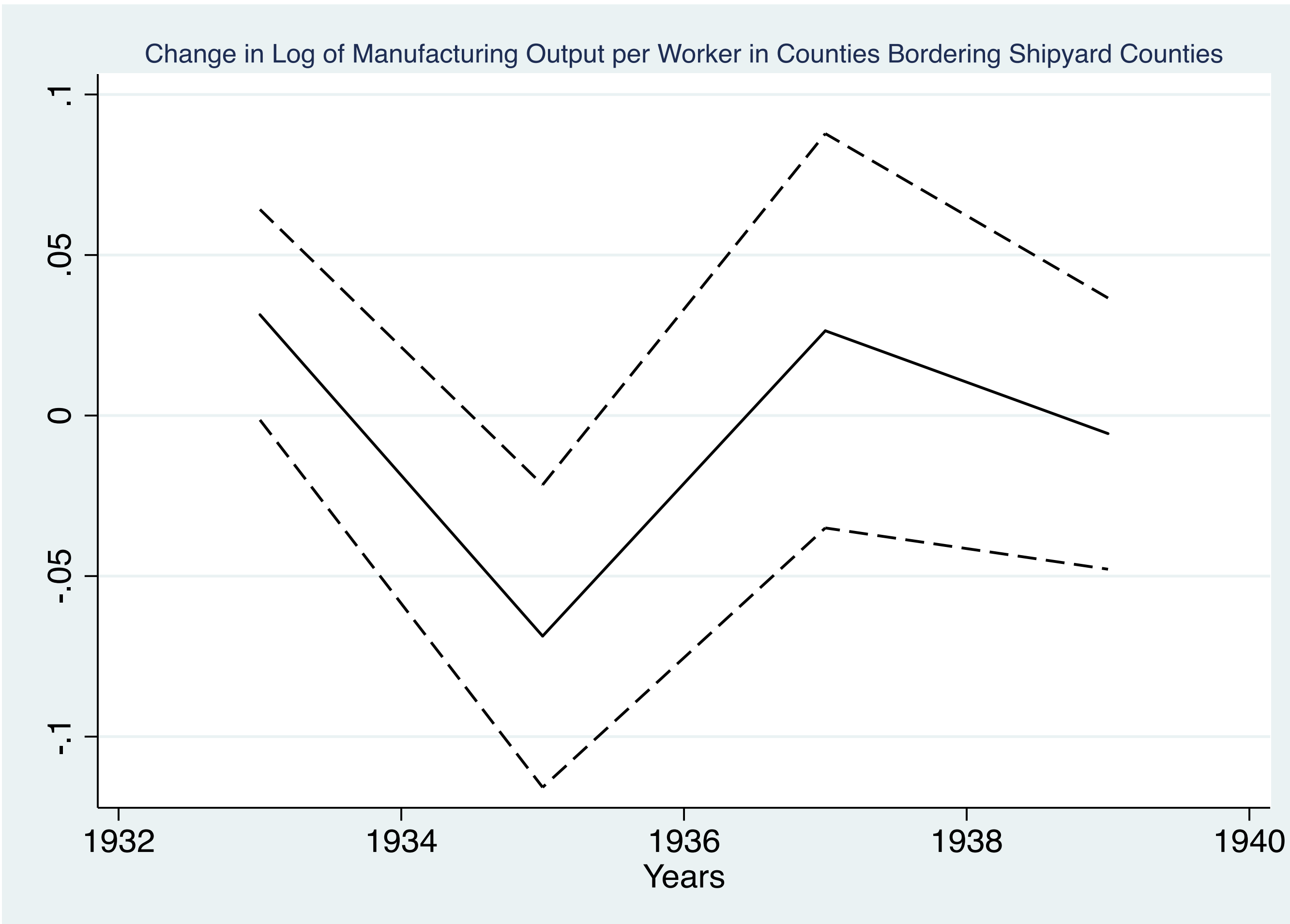
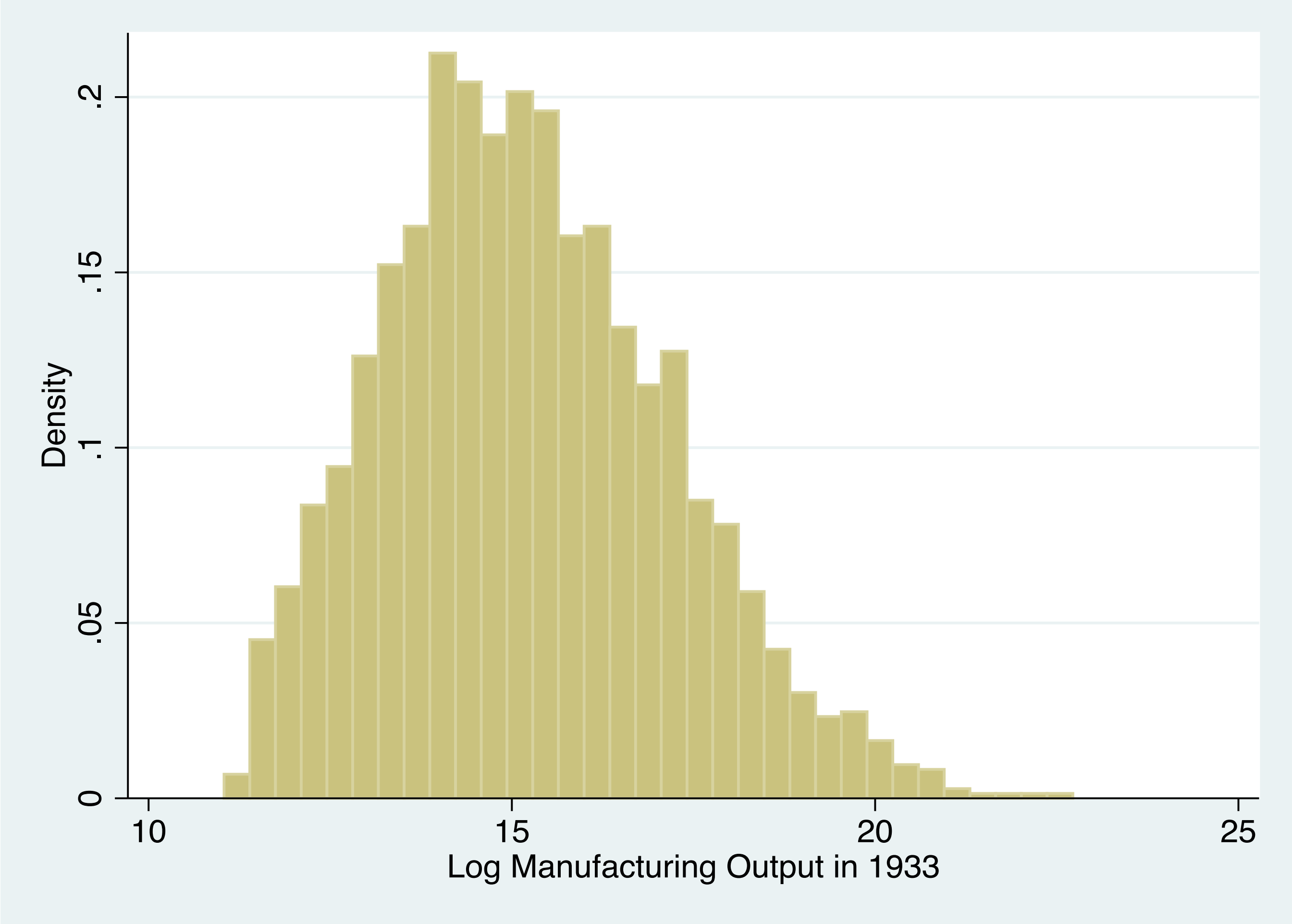


Figure plots coefficients on $BordersShipyard_{1934}$ dummy variable (corresponding to the year 1933) and coefficients on interaction terms between $BordersShipyard_{1934}$ dummy and dummy variables for 1935, 1937, and 1939. Figure contains 95% confidence bands estimated with standard errors clustered at the state level. Regression includes controls for state fixed effects, whether county is "industrialized," on a coast, and its urban percentage in 1930.

Figure 5.20: Distribution of Log Manufacturing Output across Counties in 1933



The figure plots the distribution of log manufacturing output across counties in 1933.

Chapter 6

Conclusion

This dissertation has had as its objective, the evaluation of the possibility that government purchases may have state dependent effects. That is, it has strived to answer the question of whether the output multiplier on purchases is greater when the economy is, as a whole, relatively weaker. This notion dates back at least as far as Keynes (1936).

In the first chapter, I showed that when one considers a specification in which control variables are expressed in log differences, monetary policy is controlled for, and the threshold level of the unemployment rate is estimated via the least squares technique of Hansen (2000), local projection impulse response estimation methods suggest that the multiplier is near two when the unemployment rate is relatively high and is well below one when it is relatively low. The "bad" state multiplier is significantly greater than one (ignoring the uncertainty inherent in estimating the threshold level of the unemployment rate), while the "good" state multiplier is significantly below one. What is more, the two multipliers are statistically different as well. These results provide support for the idea that there is state dependence in the government spending multiplier, agreeing with the results of Auerbach and Gorodnichenko (2012b) and disagreeing with those of Ramey and Zubairy (2014). Extensions suggest that there is also state dependence in the effects of tax changes on output and that the effects of spending changes are driven mostly by a positive response of private investment and durable goods consumption.

The second chapter, however, questions the robustness of these results to reasonable variations in the empirical specification. I consider variation across eight dimensions of specification choices. I find that there are certain specification choices (such as specifying nonstationary control variables as log deviations from a deterministic time trend) that are more likely to deliver "extreme" multiplier estimates that have no basis in any reputable theory and do not seem realistic. There are some choices that systematically lead to a higher or lower multiplier estimate, such as which macroeconomic variable to use to define good and bad states of the economy. This chapter illustrates the perils attendant to estimating the government spending multiplier, an inherently complex computation, where there are so few observations in the aggregate time series. It bolsters the argument for taking seriously the effects at local level, where identification can be sharper and the number of observations larger.

The local (county) level is the setting for the third chapter of this dissertation. I consider the effects on county manufacturing and retail sales of the Vinson-Trammell Act of 1934, passage of which was plausibly exogenous with respect to the economic health of the counties hosting shipyards. This is because the primary motivation behind this legislation was to counter a rising military threat from Japan. Note also that this was a time of extreme slack in the United States economy. My results show that manufacturing output grew six percentage points faster annually in counties hosting shipyards in the latter half of the 1930s than did like counties that did not have shipyards. Retail sales, likewise, grew three to four percentage points faster and this retail sales growth spilled over into neighboring counties as well. Even at the household level, a positive effect on consumption is visible. Although it is apparent that counties that received the spending were affected positively, it is less clear how counties that merely paid for the spending responded. This makes it difficult to translate the estimated local government spending multiplier to an aggregate government spending multiplier.

In sum, these results support the idea that government purchases have larger effects on output when there are more underutilized resources, but that support is not ironclad. There is a tension in that aggregate studies more directly measure the object of interest (that is, the government spending multiplier in good times and in bad), but they do so very noisily, and there will always be reason to quibble with identification schemes or sample periods, changes in which may substantially alter the conclusions. Disaggregated studies get around these problems, but it can be very difficult to generalize the local results. Thus, as ever, more research is needed, especially given the importance of this question to policy makers who must every so often make a determination as to whether an increase in purchases may be helpful in stimulating activity.

Bibliography

Acconcia, Antonio, Giancarlo Corsetti, and Saverio Simonelli. 2014. "Mafia and Public Spending: Evidence on the Fiscal Multiplier from a Quasi-Experiment." *The American Economic Review* 104 (7):2185–2209.

Alesina, Alberto and Silvia Ardagna. 2013. "The Design of Fiscal Adjustments." *Tax Policy and the Economy* 27:19–68.

Alloza, Mario. 2014. "Is Fiscal Policy More Effective in Uncertain Times or During Recessions?" Working Paper, University College London.

Angrist, Joshua D., Oscar Jordà, and Guido Kuersteiner. 2013. "Semiparametric Estimates of Monetary Policy Effects: String Theory Revisted." NBER Working Paper 19355.

Auerbach, Alan J. and Yuriy Gorodnichenko. 2012a. "Fiscal Multipliers in Recession and Expansion." In *Fiscal Policy after the Financial Crisis*, edited by Alberto Alesina and Francesco Giavazzi. University of Chicago Press, 63–98.

———. 2012b. "Measuring the Output Responses to Fiscal Policy." *American Economic Journal: Economic Policy* 4 (2):1–27.

———. 2013. "Output Spillovers from Fiscal Policy." *The American Economic Review* 103 (3):141–146.

———. 2014. "Fiscal Multipliers in Japan." NBER Working Paper 19911.

Bachmann, Rudiger and Eric R. Sims. 2012. "Confidence and the Transmission of Government Spending Shocks." *Journal of Monetary Economics* 59 (3):235–249.

Barro, Robert J. and Charles J. Redlick. 2011. "Macroeconomic Effects from Government Purchases and Taxes." *The Quarterly Journal of Economics* 126 (1):51–102.

Baxter, Marianne and Robert G. King. 1993. "Fiscal Policy in General Equilibrium." *The American Economic Review* 83 (3):315–334.

Beaudry, Paul and Gary Koop. 1993. "Do Recessions Permanently Change Output?" *Journal of Monetary Economics* 31 (2):149–163.

Beveridge, Stephen and Charles R. Nelson. 1981. "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'." *Journal of Monetary Economics* 7 (2):151–174.

Biolsi, Christopher. 2015. "Asymmetric Effects of Government Purchases over the Business Cycle." Working Paper, University of Houston.

Blanchard, Olivier and Daniel Leigh. 2013. "Growth Forecast Errors and Fiscal Multipliers." *The American Economic Review* 103 (3):117–120.

Blanchard, Olivier and Roberto Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *The Quarterly Journal of Economics* 117 (4):1329–1368.

Bognanni, Mark. 2013. "An Empirical Analysis of Time-Varying Fiscal Multipliers." Working Paper, University of Pennsylvania.

Bruckner, Markus and Anita Tuladhar. 2014. "Local Government Spending Multipliers and Financial Distress: Evidence from Japanese Prefectures." *The Economic Journal* 124 (581):1279–1316.

Caggiano, Giovanni, Efrem Castelnuovo, Valentina Colombo, and Gabriela Nodari. 2014. "Estimating Fiscal Multipliers: News from a Nonlinear World." Melbourne Institute Working Paper Series Working Paper No. 26/14.

Calvo, Guillermo A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12 (3):383–398.

Candelon, Bertrand and Lenard Lieb. 2013. "Fiscal Policy in Good and Bad Times." *Journal of Economic Dynamics & Control* 37 (12):2679–2694.

Canova, Fabio and Evi Pappa. 2007. "Price Differentials in Monetary Unions: The Role of Fiscal Shocks." *The Economic Journal* 117 (520):713–737.

Canzoneri, Matthew, Fabrice Collard, Harris Dellas, and Behzad Diba. 2012. "Fiscal Multipliers in Recessions." Working Paper, Georgetown University and University of Bern.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. 2014. "Housing Booms, Labor Market Outcomes, and Educational Attainment." Working Paper, University of Chicago.

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* 4 (3):118–145.

Christiano, Lawrence, Martin Eichenbaum, and Sergio Rebelo. 2011. "When is the Government Spending Multiplier Large?" *Journal of Political Economy* 119 (1):78–121.

Christiano, Lawrence J. 1992. "Searching for a Break in GNP." *Journal of Business & Economic Statistics* 10 (3):237–250.

Clark, Peter K. 1987. "The Cyclical Component of U.S. Economic Activity." *The Quarterly Journal of Economics* 102 (4):797–814.

Clemens, Jeffrey and Stephen Miran. 2012. "Fiscal Policy Multipliers on Subnational Government Spending." *American Economic Journal: Economic Policy* 4 (2):46–68.

Coenen, Gunter, Christopher J. Erceg, Charles Freedman, Davide Furceri, Michael Kumhof, René Lalonde, Douglas Laxton, Jesper Lindé, Annabelle Mourougane, Dirk Muir, Susanna Mursula, Carlos de Resende, John Roberts, Wern er Roeger, Stephen Snudden, Mathias Trabandt, and Jan In't Veld. 2012. "Effects of Fiscal

Stimulus in Structural Models." *American Economic Journal: Macroeconomics* 4 (1):22–68.

Cogley, Timothy and James M. Nason. 1995. "Effects of the Hodrick-Prescott Filter on Trend and Difference Stationary Time Series: Implications for Business Cycle Research." *Journal of Economic Dynamics and Control* 19 (1-2):253–278.

Cohen, Lauren, Joshua Coval, and Christopher Malloy. 2011. "Do Powerful Politicians Cause Corporate Downsizing?" *Journal of Political Economy* 119 (6):1015–1060.

Cook, James F. 2004. *Carl Vinson: Patriarch of the Armed Forces.* Macon, GA: Mercer University Press.

Crafts, Nicholas and Terence C. Mills. 2013. "Rearmament to the Rescue? New Estimates of the Impact of 'Keynesian' Policies in 1930s' Britain." *The Journal of Economic History* 73 (4):1077–1104.

Davig, Troy and Eric M. Leeper. 2011. "Monetary-Fiscal Policy Interactions and Fiscal Stimulus." *European Economic Review* 55 (2):211–227.

Denes, Matthew, Gauti B. Eggertsson, and Sophia Gilbukh. 2013. "Deficits, Public Debt Dynamics and Tax and Spending Multipliers." *The Economic Journal* 123 (566):F133–F163.

Diebold, Francis X. and Abdelhak S. Senhadji. 1996. "The Uncertain Unit Root in Real GNP: Comment." *The American Economic Review* 86 (5):1291–1298.

Dixit, Avinash K. and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *The American Economic Review* 67 (3):297–308.

Durlauf, Steven N., Salvador Navarro, and David A. Rivers. 2014. "Model Uncertainty and the Effect of Shall-Issue Right-to-Carry Laws on Crime." Working Paper, University of Wisconsin at Madison and University of Western Ontario.

Eggertsson, Gauti B. 2010. "What Fiscal Policy is Effective at Zero Interest Rates?" *NBER Macroeconomics Annual* 25:59–112.

Elliott, Graham, Thomas J. Rothenberg, and James H. Stock. 1996. "Efficient Tests for an Autoregressive Unit Root." *Econometrica* 64 (4):813–836.

Engemann, Kristie M., Michael T. Owyang, and Sarah Zubairy. 2008. "A Primer on the Empirical Identification of Government Spending Shocks." *Federal Reserve Bank of St. Louis Review* 90 (2):117–132.

Farhi, Emmanuel and Iván Werning. 2013. "Fiscal Multipliers: Liquidity Traps and Currency Unions." Working Paper, Harvard University and Massachusetts Institute of Technology.

Fazzari, Steven M., James Morley, and Irina Panovska. 2013. "State-Dependent Effects of Fiscal Policy." Forthcoming in *Studies in Nonlinear Dynamics and Econometrics.*

Fischer, Gerald J. 1946. "Cost of War Built Vessels from Inception, October 25, 1936 to June 30, 1946." Records of the Office of the Historian, Box 35. Records of the USMC (National Archives.

Fishback, Price, Werner Troesken, Trevor Kollman, Michael Haines, Paul Rhode, and Melissa Thomasson. 2011a. "Information and the Impact of Climate and Weather on Mortality Rates during the Great Depression." In *The Economics of Climate Change*, edited by Gary D. Libecap and Richard H. Steckel. University of Chicago Press, 131–168.

———. 2011b. "Weather, Demography, Economy, and the New Deal at the County Level, 1930-1940."

Fishback, Price V. and Joseph A. Cullen. 2013. "Second World War Spending and Local Economic Activity in U.S. Counties, 1939-58." *The Economic History Review* 66 (4):975–992.

Fishback, Price V., William C. Horrace, and Shawn Kantor. 2005. "Did New Deal Grant Programs Stimulate Local Economies? A Study of Federal Grants and Retail Sales during the Great Depression." *The Journal of Economic History* 65 (1):36–71.

Fishback, Price V. and Valentina Kachanovskaya. 2010. "In Search of the Multiplier for Federal Spending in the States during the Great Depression." NBER Working Paper 16561.

Fisher, Jonas D. M. and Ryan Peters. 2010. "Using Stock Returns to Identify Government Spending Shocks." *The Economic Journal* 120 (544):414–436.

Gabaix, Xavier. 2011. "The Granular Origins of Aggregate Fluctuations." *Econometrica* 79 (3):733–772.

Galí, Jordi, J. David López-Salido, and Javier Vallés. 2007. "Understanding the Effects of Government Spending on Consumption." *Journal of the European Economic Association* 5 (1):227–270.

Giavazzi, Francesco and Marco Pagano. 1990. "Can Severe Fiscal Contractions be Expansionary? Tales of Two Small European Countries." *NBER Macroeconomics Annual* 5:75–122.

Gilchrist, Simon, Vladimir Yankov, and Egon Zakrajšek. 2009. "Credit Market Shocks and Economic Fluctuations: Evidence from Corporate Bond and Stock Markets." *Journal of Monetary Economics* 56 (4):471–493.

Gordon, Robert J. and Robert Krenn. 2014. "The End of the Great Depression 1939-41: Fiscal Multipliers, Capacity Constraints, and Policy Contributions." Working Paper, Northwestern University and U.S. Airways.

Gorodnichenko, Yuriy. 2014. "Discussion of: 'Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data'."

Gospodinov, Nikolay, Ana Maria Herrera, and Elena Pesavento. 2013. "Unit Roots, Cointegration, and Pretesting in VAR Models." In *VAR Models in Macroeconomics - New Developments and Applications: Essays in Honor of Christopher A. Sims*, edited by Thomas B. Fomby, Lutz Kilian, and Anthony Murphy, chap. 9. Emerald Group Publishing Limited, 81–115.

Gupta, Ashmita. 2015. "Effect of Trade Liberalization on Educational Attainment: Evidence from Indian Tariff Reforms." Working Paper, University of Houston.

Hall, Robert E. 2009. "By How Much Does GDP Rise if the Government Buys More Output?" *Brookings Papers on Economic Activity* 2009:2:183–231.

Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57 (2):357–384.

Hansen, Bruce E. 1996. "Inference when a Nuisance Parameter is not Identified under the Null Hypothesis." *Econometrica* 64 (3):413–430.

———. 2000. "Sample Splitting and Threshold Estimation." *Econometrica* 68 (3):575–603.

Hausman, Joshua K. 2013. "Fiscal Policy and Economic Recovery: The Case of the 1936 Veterans' Bonus." Berkeley Economic History Lab Working Paper WP2013-06.

Hodrick, Robert J. and Edward C. Prescott. 1997. "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit, and Banking* 29 (1):1–16.

Hooker, Mark A. 1996. "How Do Changes in Military Spending Affect the Economy? Evidence from State-Level Data." *New England Economic Review* 7 (2):1–16.

Ilzetzki, Ethan, Enrique G. Mendoza, and Carlos A. Végh. 2013. "How Big (Small?) are Fiscal Multipliers?" *Journal of Monetary Economics* 60 (2):239–254.

Interuniversity Consortium for Political and Social Research. 2005. "Historical, Demographic, Economic, and Social Data: The United States, 1790-1970." URL `http://doi.org/10.3886/ICPSR00003.v1`.

Jordà, Oscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *The American Economic Review* 95 (1):161–182.

Jordà, Oscar and Alan M. Taylor. 2013. "The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy." NBER Working Paper 19414.

Kehrig, Matthias. 2015. "The Cyclical Nature of the Productivity Distribution." Working Paper, University of Texas.

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest, and Money*. London, UK: The Macmillan Press.

Kim, Chang-Jin, James Morley, and Jeremy Piger. 2005. "Nonlinearity and the Permanent Effects of Recessions." *Journal of Applied Econometrics* 20 (2):291–309.

Kim, Chang-Jin and Christian J. Murray. 2002. "Permanent and Transitory Components of Recessions." *Empirical Economics* 27 (2):163–183.

Kim, Chang-Jin and Charles R. Nelson. 1999. "Friedman's Plucking Model of Business Fluctuations: Tests and Estimates of Permanent and Transitory Components." *Journal of Money, Credit, and Banking* 31 (3, Part 1):317–334.

King, Thomas B. and James Morley. 2007. "In Search of the Natural Rate of Unemployment." *Journal of Monetary Economics* 54 (2):550–564.

Koop, Gary, Hashem Pesaran, and Simon M. Potter. 1996. "Impulse Response Analysis in Nonlinear Multivariate Models." *Journal of Econometrics* 74 (1):119–147.

Kormilitsina, Anna and Sarah Zubairy. 2013. "Propagation Mechanisms for Government Spending Shocks: A Bayesian Comparison." Working Paper, Southern Methodist University and Texas A&M University.

Kuhn, Florian and Chacko George. 2014. "Business Cycle Implications of Capacity Constraints under Demand Shocks." Working Paper, University of Texas and Federal Deposit Insurance Corporation.

Lane, Frederic C. 1951. *Ships for Victory. A History of Shipbuilding under the U.S. Maritime Commission in World War II.* Baltimore, MD: The Johns Hopkins Press.

Mankiw, N. Gregory and Matthew Weinzierl. 2011. "An Exploration of Optimal Stabilization Policy." *Brookings Papers on Economic Activity* 2011:1:209–272.

McGrattan, Ellen R. and Lee E. Ohanian. 2010. "Does Neoclassical Theory Account for the Effects of Big Fiscal Shocks? Evidence from World War II." *International Economic Review* 51 (2):509–532.

Michaillat, Pascal. 2014. "A Theory of Countercyclical Government Multiplier." *American Economic Journal: Macroeconomics* 6 (1):190–217.

Mittnik, Stefan and Willi Semmler. 2012. "Regime Dependence of the Fiscal Multiplier." *Journal of Economic Behavior & Organization* 83 (3):502–522.

Morley, James and Jeremy Piger. 2012. "The Asymmetric Business Cycle." *The Review of Economics and Statistics* 94 (1):208–221.

Mountford, Andrew and Harald Uhlig. 2009. "What are the Effects of Fiscal Policy Shocks?" *Journal of Applied Econometrics* 24 (6):960–992.

Murray, Christian J. 2003. "Cyclical Properties of Baxter-King Filtered Time Series." *The Review of Economics and Statistics* 85 (2):472–476.

Murray, Christian J. and Charles R. Nelson. 2000. "The Uncertain Trend in U.S. GDP." *Journal of Monetary Economics* 46 (1):79–95.

———. 2002. "The Great Depression and Output Persistence." *Journal of Money, Credit, and Banking* 34 (4):1090–1098.

———. 2004. "The Great Depression and Output Persistence: A Reply to Papell and Prodan." *Journal of Money, Credit, and Banking* 36 (No. 3 Part 1):429–432.

Nakamura, Emi and Jón Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *The American Economic Review* 104 (3):753–792.

Nelson, Charles R. 2006. "The Beveridge-Nelson Decomposition in Retrospect and Prospect." Working Paper, University of Washington.

Nelson, Charles R. and Heejoon Kang. 1981. "Spurious Periodicity in Inappropriately Detrended Time Series." *Econometrica* 49 (3):741–751.

Nelson, Charles R. and Charles I. Plosser. 1982. "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications." *Journal of Monetary Economics* 10 (2):139–162.

Newey, Whitney K. and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3):703–708.

Ng, Serena and Pierre Perron. 1995. "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag." *Journal of the American Statistical Association* 90 (429):268–281.

Ostergaard, Charlotte, Bent E. Sørensen, and Oved Yosha. 2002. "Consumption and Aggregate Constraints: Evidence from U.S. States and Canadian Provinces." *Journal of Political Economy* 110 (3):634–645.

Owyang, Michael T., Valerie A. Ramey, and Sarah Zubairy. 2013. "Are Government Spending Multipliers Greater During Periods of Slack? Evidence from 20th Century Historical Data." *The American Economic Review* 103 (3):129–134.

Ozer-Balli, Hatice and Bent Sørensen. 2013. "Interaction Effects in Econometrics." *Empirical Economics* 45 (1):583–603.

Papell, David H. and Ruxandra Prodan. 2004. "The Uncertain Unit Root in U.S. Real GDP: Evidence with Restricted and Unrestricted Structural Change." *Journal of Money, Credit, and Banking* 36 (No. 3 Part 1):423–427.

Perotti, Roberto. 1999. "Fiscal Policy in Good Times and Bad." *The Quarterly Journal of Economics* 114 (4):1399–1436.

Perron, Pierre. 1989. "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis." *Econometrica* 57 (6):1361–1401.

Ramey, Valerie A. 2011a. “Can Government Purchases Stimulate the Economy?” *Journal of Economic Literature* 49 (3):673–685.

———. 2011b. “Identifying Government Spending Shocks: It’s All in the Timing.” *The Quarterly Journal of Economics* 126 (1):1–50.

———. 2012. “Government Spending and Private Activity.” In *Fiscal Policy after the Financial Crisis*, edited by Alberto Alesina and Francesco Giavazzi. University of Chicago Press, 19–55.

———. 2014. “Defense News Shocks, 1889-2013: Estimates Based on News Sources.” Working Paper, University of California, San Diego.

Ramey, Valerie A. and Matthew D. Shapiro. 1998. “Costly Capital Reallocation and the Effects of Government Spending.” *Carnegie-Rochester Conference Series on Public Policy* 48:145–194.

Ramey, Valerie A. and Sarah Zubairy. 2014. “Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data.” NBER Working Paper 20719.

Riera-Crichton, Daniel, Carlos A. Végh, and Guillermo Vuletin. 2014. “Fiscal Multipliers in Recessions and Expansions: Does It Matter Whether Government Spending is Increasing or Decreasing?” World Bank Policy Research Working Paper 6993.

Romer, Christina D. and David H. Romer. 2010. “The Macroeconomic Effects of Tax

Changes: Estimates Based on a New Measure of Fiscal Shocks." *The American Economic Review* 100 (3):763–801.

———. 2014. "Transfer Payments and the Macroeconomy: The Effects of Social Security Benefit Changes, 1952-1991." NBER Working Paper 20087.

Rossi, Barbara and Sarah Zubairy. 2011. "What is the Importance of Monetary and Fiscal Shocks in Explaining U.S. Macroeconomic Fluctuations?" *Journal of Money, Credit, and Banking* 43 (6):1247–1270.

Rotemberg, Julio J. and Michael Woodford. 1992. "Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity." *Journal of Political Economy* 100 (6):1153–1207.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. "Integrated Public Use Microdata Series, Volume 5.0." Machine-readable database.

Sala-I-Martin, Xavier X. 1997. "I Just Ran Four Million Regressions." NBER Working Paper 6252.

Schmidt, Julia. 2013. "Country Risk Premia, Endogenous Collateral Constraints and Non-linearities: A Threshold VAR Approach." Working Paper, Graduate Institute of International and Development Studies.

Serrato, Juan Carlos Suárez and Philippe Wingender. 2014. "Estimating Local Fiscal Multipliers." Working Paper, Stanford Institute for Economic Policy Research and International Monetary Fund.

Shoag, Daniel. 2010. "The Impact of Government Spending Shocks: Evidence on the Multiplier from State Pension Plan Returns." Working Paper, Harvard University.

———. 2013. "Using State Pension Shocks to Estimate Fiscal Multipliers since the Great Recession." *The American Economic Review* 103 (3):121–124.

Sims, Eric and Jonathan Wolff. 2013. "The Output and Welfare Effects of Fiscal Shocks over the Business Cycle." NBER Working Paper 19749.

Sizova, Natalia. 2015. "Efficient Tests for Long-Run Predictability: Do Long-Run Relations Convey Extra Information?" Working Paper, Rice University.

Tagkalakis, Athanasios. 2008. "The Effects of Fiscal Policy on Consumption in Recessions and Expansions." *Journal of Public Economics* 92 (5-6):1486–1508.

Thornton, Rebecca Achee and Peter Thompson. 2001. "Learning from Experience and Learning from Others: An Exploration of Learning and Spillovers in Wartime Shipbuilding." *The American Economic Review* 91 (5):1350–1368.

United States Department of Commerce, Bureau of the Census. 1992. "CONTIGUOUS COUNTY FILE, 1991 [United States]." URL `http://doi.org/10.3886/ICPSR09835.v1`. ICPSR09835-v1, Ann Arbor, MI.

———. 2012. "County and City Data Book [United States] Consolidated File: County Data, 1947-1977." URL `http://doi.org/10.3886/ICPSR07736.v2`. ICPSR07736-v2, Ann Arbor, MI.

U.S. Dept. of Labor. BLS. Cost of Living Division, U.S. Dept. of Agriculture. BHE. Economics Division, U.S. Natural Resources Committee. Consumption Research Staff. Industrial Section, U.S. Central Statistical Board, and U.S. WPA. 2009. "Study of Consumer Purchases in the United States, 1935-1936." URL `http://doi.org/10.3886/ICPSR08908.v3`. ICPSR08908-v3, Ann Arbor, MI.

Wieland, Johannes F. 2012. "Fiscal Multipliers at the Zero Lower Bound: International Theory and Evidence." Working Paper, University of California, Berkeley.

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

Yang, Weonho, Jan Fidrmuc, and Sugata Ghosh. 2014. "Using Military Build-Ups to Capture Fiscal Shocks: A Reassessment." CESifo Working Paper No. 4689.

Zivot, Eric and Donald W. K. Andrews. 1992. "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis." *Journal of Business and Economic Statistics* 10 (3):251–270.

Zubairy, Sarah. 2014. "On Fiscal Multipliers: Estimates from a Medium Scale DSGE Model." *International Economic Review* 55 (1):169–195.

Appendix A

The Nakamura and Steinsson (2014) Model

The theoretical foundations for this study borrow heavily from the model developed by Nakamura and Steinsson (2014), who construct an environment in which two economies ("home" and "foreign" in their terminology) are linked in a monetary and fiscal union. The objective of their model is the calculation of an "Open Economy Relative Multiplier," which they define as the effect of a relative spending increase in one region relative to another on relative output. They also translate their findings on open economy relative multipliers to the "Closed Economy Aggregate Multiplier" often estimated in macroeconomic research. Households and firms exhibit the same behavior in both regions. Households maximize lifetime utility over consumption and labor supply, as given by

$$E_0 \sum_{t=0}^{\infty} \beta^t u(C_t, L_t(x)), \tag{A.1}$$

and in this formulation, β is the household's subjective discount factor and $L_t(x)$ reflects that each household supplies a differentiated kind of labor, indexed by x. C_t represents a composite consumption good made up of goods produced in both the home and foreign regions. The home and foreign goods themselves are also composite goods of a large number of differentiated goods produced in each region, aggregated as in Dixit and Stiglitz (1977). There is open trade in goods between the two regions, but labor is immobile across regions. There is no capital accumulation in the baseline model. The authors consider two types of preference specifications for their households, one that is separable in consumption and labor, specified as

$$u(C_t, L_t(x)) = \frac{C_t^{1-\sigma^{-1}}}{1-\sigma^{-1}} - \chi \frac{L_t(x)^{1+\nu^{-1}}}{1+\nu^{-1}}, \tag{A.2}$$

where σ denotes the intertemporal elasticity of substitution and ν denotes the Frisch labor elasticity. The second specification is nonseparable in consumption and labor and takes the form

$$u(C_t, L_t(x)) = \frac{(C_t - \chi L_t(x)^{1+\nu^{-1}}/(1+\nu^{-1}))^{1-\sigma^{-1}}}{1-\sigma^{-1}}, \tag{A.3}$$

and Nakamura and Steinsson (2014) demonstrate that the specification of preferences makes a considerable difference in the size of their open economy relative multiplier. I will discuss the intuition for why this is further below. Households choose consumption (including over home and foreign goods and over the differentiated goods within each category)[1] and labor supply subject to a flow budget constraint where

[1] Details on how these choices are made can be found in their paper in Equations 10-13. Essentially, consumption of home and foreign goods depends on the elasticity of substitution between

income is made up of payoffs on state-contingent securities, labor income less labor income taxes, profits from the home composite good firm, and lump sum taxes or transfers. This results in a consumption Euler Equation

$$\frac{u_c(C_{t+j}, L_{t+j}(x))}{u_c(C_t, L_t(x))} = \frac{M_{t,t+j}}{\beta^j}\frac{P_{t+j}}{P_t}, \tag{A.4}$$

where M_t is the stochastic discount factor that prices the state-contingent securities and P_t is the composite price index. There is an analogous expression governing the consumption of the foreign household.[2] It also results in an intratemporal equilibrium condition that determines labor supply

$$\frac{u_l(C_t, L_t(x))}{u_c(C_t, L_t(x))} = (1-\tau)\frac{W_t(x)}{P_t}, \tag{A.5}$$

where τ is the distortionary tax rate on labor income and $W_t(x)$ is the nominal wage rate for the worker supplying labor of type x.

By combining the Euler Equations for the home and foreign households, one gets

$$\frac{u_c(C_t^*, L_t^*(x))}{u_c(C_t, L_t(x))} = \frac{P_t^*}{P_t}. \tag{A.6}$$

The model features a common fiscal authority that purchases (composite) final output from both the home and foreign regions according to exogenous AR(1) processes, with consumption of varieties within the home and foreign regions governed by the same parameters that govern private consumption of these varieties, that is, the relative prices and the elasticity of substitution among them, which is the same as for private households.

---

these composite goods and the relative prices, while within each category, consumption of the various differentiated varieties depends on the elasticity of substitution between the varieties and their relative prices.

[2] I will denote variables for the foreign household the same way that Nakamura and Steinsson (2014) do, with a star superscript.

The government in this model also includes a monetary authority that sets nominal interest rates according to an augmented Taylor rule, with the arguments entering the Taylor rule being aggregate inflation, the aggregate output gap, and aggregate government spending. Thus, monetary policy is common to both regions.

Firms employ labor to produce their differentiated product. There are a large number of firms within each industry x (which maps to the different kinds of labor supplied by households). Each firm must satisfy demand on the part of home and foreign households, as well as the fiscal authority, and it takes the wages in its industry as given. This leads to its profit maximizing labor demand, given by

$$W_t(x) = f_l(L_t(z))S_t(z), \tag{A.7}$$

where $f_l(L_t(z))$ is the marginal product for firm z in industry x, and $S_t(z)$ is its nominal marginal cost.

Nakamura and Steinsson (2014) consider a few different varieties of their model, in which the fiscal authority finances the exogenously given spending shock by raising lump sum taxes or labor income taxes; in which monetary policy is governed by the Taylor principle (in which real interest rates rise more than one-for-one with a rise in inflation) and in which it is not, and where firms face nominal rigidities of the Calvo (1983) type and where they do not. They find that the model with nonseparable preferences and sticky prices best fit their empirical findings.

In fact, it is the distinction between separable and nonseparable preferences that is most relevant to the discussion here. Consider the case of separable preferences.

Government demand for an individual variety of good in the home region is given by

$$g_{ht}(z) = G_{Ht} \left(\frac{p_{ht}(z)}{P_{Ht}} \right)^{-\theta}, \tag{A.8}$$

which is rising in G_{Ht}, government purchases from the home region in period t. A positive shock to purchases, therefore, raises government demand for every variety of good in the home region, which must be satisfied by all of the firms. This gives an increase in marginal cost for each firm, and a fraction $1 - \alpha$ of them can reoptimize the price of their good. This will, all else equal, push P_{Ht} and, consequently also P_t, higher. A look at the intertemporal Euler equation for the home household, Equation A.4, shows that consumption will then decline in the home region. The disparity between the home region's consumption decline and the foreign region's consumption decline will depend on the relative degree of home bias in consumption. With no investment in the model, this is likely to give a local output multiplier below one. This is because, when preferences are separable, a government spending shock in the home region raises prices in that region relative to the foreign region. Equation A.6 implies that consumption will decline in the home region relative to the foreign region. With nominal interest rates common across both regions, an upward shock to prices in the home region will make the real interest rate lower in the home region than in the foreign region in the short term, so one might expect that consumption would rise in the home region, as an upward shock to prices does in the model of Christiano, Eichenbaum, and Rebelo (2011), among others. In the long term, however, the shock to government spending in the home region will dissipate and the price ratio between home and foreign regions will return to its original level, absent any further shocks. Thus, from the perspective of an agent in the home region after the spending shock hits, the expected price path of home produced goods is negative, and the long term real interest rate (arguably the one that is more relevant for consumption decisions) is actually higher in the home region relative to the foreign region. This leads to a sharp decline in consumption.

When, however, preferences are nonseparable, consumption and labor supply are complementary. This can also be seen from the optimality conditions of the model. Consider the first order condition for maximizing the period utility function with respect to consumption.

$$u_c = (C_t - \frac{\chi L_t(x)^{1+\nu^{-1}}}{1+\nu^{-1}})^{-\sigma^{-1}} - \lambda_t P_t = 0, \tag{A.9}$$

where λ_t is the Lagrange multiplier. By taking the derivative of this expression with respect to labor supply, $L_t(x)$, one arrives at

$$u_{cl} = \sigma^{-1}(C_t - \frac{\chi L_t(x)^{1+\nu^{-1}}}{1+\nu^{-1}})^{-\sigma^{-1}-1}\chi L_t(x)^{\nu^{-1}}, \tag{A.10}$$

and this expression is unambiguously positive, as long as χ is not too large. That is, the more the agent works, the more she wants to consume as well. The rise in government purchases in the home region raises labor supply and consumption simultaneously. Nakamura and Steinsson (2014) give the examples of gasoline and meals away from home as goods that private agents will consume more of as they work more. This leads to a relatively high open economy relative multiplier and a high closed economy aggregate multiplier.

In this paper, in order to interpret the empirical regression results, I use this model to convert my local multiplier estimates to an aggregate multiplier. Following Nakamura and Steinsson (2014), I assume nonseparable preferences and sticky prices (the formulations that best match their empirical findings), and since the context of my study is mainly the Great Depression of the 1930s, when nominal interest rates were pinned near the Zero Lower Bound (ZLB), I will also assume a constant nominal interest rate policy. Also, since I observe that the higher spending was financed with higher taxes, I force the fiscal authority in the model to maintain a balanced budget in all periods.

I make one minor modification to the model. Given that the purpose of the spending that I study is to explicitly build up the navy, it may be that I want the fiscal authority in the model to have a lower (or at least) separate elasticity of substitution among the different varieties of goods compared to the households. I do this by specifying a new parameter θ_g which replaces θ in the government's demand for goods, Equation A.8.

A major contribution is that, unlike Nakamura and Steinsson (2014), I will be able to test empirically whether consumption rose with output, so as to verify that a model with nonseparable preferences is the appropriate one to use.