Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data

Valerie A. Ramey
University of California, San Diego and NBER

Sarah Zubairy
Texas A&M University

Preliminary version: 4 October 2013

Abstract

This paper investigates whether U.S. government spending multipliers differ according to two potentially important features of the economy: (1) the amount of slack and (2) whether interest rates are near the zero lower bound. We shed light on these questions by analyzing new quarterly historical U.S. data covering multiple large wars and deep recessions. We estimate a state-dependent model in which impulse responses and multipliers depend on the average dynamics of the economy in each state. We find no evidence that multipliers are different across states, whether defined by the amount of slack in the economy or whether interest rates are near the zero lower bound. We show that our results are robust to many alternative specifications. Our results imply that, contrary to recent conjecture, government spending multipliers were not necessarily higher than average during the Great Recession.

We are grateful to Alan Auerbach, Yuriy Gorodnichenko, Michael Owyang, Garey Ramey, Harald Uhlig, seminar participants at Berkeley, and participants at the Hallfest conference for very helpful suggestions. We thank Kate Vermann and Michelle Ramey for excellent research assistance. This paper is part of an ongoing research project on state dependent government spending multipliers by Michael Owyang, Valerie Ramey, and Sarah Zubairy.

1 Introduction

What is the multiplier on government purchases? The answer to this question continues to be an important part of the public policy debate in the face of lingering high unemployment and the need for eventual fiscal consolidations. The majority of estimates based on aggregate data over the post-WWII period find modest multipliers, often below unity. If multipliers are indeed this low, they suggest that increases in government purchases are unlikely to stimulate private activity and that fiscal consolidations that involve spending decreases are unlikely to do much harm.

Most of the estimates are based on *averages* for a particular country over a particular historical period. Because there is no scope for controlled, randomized trials on countries, all estimates of aggregate government multipliers are necessarily dependent on historical happenstance. Theory tells us that details such as the persistence of spending changes, how they are financed, how monetary policy reacts, and the tightness of the labor market can significantly affect the magnitude of the multiplier. Unfortunately, the data do not present us with clean natural experiments that can answer these questions. While the recent U.S. stimulus package was purely deficit financed and undertaken during a period of high unemployment and accommodative monetary policy, it was enacted in response to a weak economy and hence any aggregate estimates are subject to simultaneous equations bias.

During the last several years, the literature has begun to explore whether estimates of government spending multipliers vary depending on circumstances. One strand of this literature considers the possibility that multipliers are different during recessions (e.g. Auerbach and Gorodnichenko (2012), Bachmann and Sims (2012), Baum et al. (2012), Mittnik and Semmler (2012), Semmler and Semmler (2013), Auerbach and Gorodnichenko (2013) and Fazzari et al. (2013)). Another strand of the literature considers how monetary policy affects government spending multipliers. New Keynesian DSGE models show that when interest rates are stuck at the zero lower bound, multipliers can be higher than in normal times (e.g. Cogan et al. (2010), Christiano et al. (2011), Coenen and et al. (2012)).

This paper contributes to the empirical literature by investigating whether government spending multipliers differ according to two potentially important features of the economy: (1) the amount of slack in the economy and (2) whether interest rates are near the zero lower bound.[1] Extending the initial analysis in Owyang, Ramey, Zubairy (2013a), we exploit the

1. We will use government "spending" and "purchases" interchangeably. It should be noted that our multipliers apply only to government purchases, not government transfers.

fact that the entire 20th Century contains potentially richer information than the post-WWII data that has been the focus of most of the recent research. We create a new quarterly data set for the U.S. extending back to 1889. This sample includes episodes of huge variations in government spending, wide fluctuations in unemployment, prolonged periods near the zero lower bound of interest rates, and a variety of tax responses.

In addition to our extended sample, our methodology also differs from that used by most of the literature. First, rather than estimating regime-switching models, we estimate our state-dependent models using Jordà's (2005) local projection method. This method offers a simple solution to some of the thorny issues that arise in computing impulse responses in regime-switching models. Second, we depart from the standard SVAR literature that first estimates elasticities and then converts them to multipliers using an *ex post* conversion factor. We show that this approach can lead to biases in the estimates of multipliers. We instead define our variables as in Hall (2009) and Barro and Redlick (2011).

We find no evidence that the multiplier on government purchases is higher during high unemployment states or when interest rates are near the zero lower bound. Most estimates of the multiplier are between 0.7 and 1.1. We perform extensive robustness checks with respect to our measures of state, sample periods, identification method and the behavior of taxes and find little change in the estimates. We demonstrate that the differences in conclusions between our work and that of Auerbach and Gorodnichenko (2012) lies in the construction of impulse response functions on which the multipliers are based. In contrast to linear models, where the calculation of impulse response functions is a straightforward undertaking, there is no "right" way to construct impulse response functions in nonlinear models.

The paper proceeds as follows. We begin by discussing the data construction in Section 2. In Section 3 we introduce the econometric methodology and discuss issues with calculating impulse responses. In Section 4, we present our measures of slack and then present estimates of a model in which multipliers are allowed to vary according to the amount of slack in the economy. We first present baseline results using our new data and methodology. We then conduct various robustness checks and then explain why our results are different from those in the literature. We also explore possible explanations for our results, such as the behavior of taxes. Section 5 tests theories that predict that multipliers should be greater when interest rates are at the zero lower bound and the final section concludes.

2 Data Description

A key contribution of the paper is the construction of a new data set that spans historical periods that involve potentially informative movements of the key variables. In particular, we construct quarterly data from 1889 through 2011 for the U.S.. We choose to estimate our model using quarterly data rather than annual data because agents often react quickly to news about government spending and the state of the economy can change abruptly.[2] The historical series include real GDP, the GDP deflator, government purchases, federal government receipts, population, the unemployment rate, interest rates, and defense news.

The data appendix contains full details, but we highlight some of the features of the data here. For the post-WWII sample, we use available published quarterly series. For the earlier periods, we use various higher frequency series to interpolate existing annual series. In most cases, we use the proportional Denton procedure which results in series that average up to the annual series.[3]

The annual real GDP data combine the series from *Historical Statistics of the U.S.* (Carter et al. (2006)) for 1889 through 1928 and the NIPA data from 1929 to the present. The annual data are interpolated with Balke and Gordon's (1986) quarterly real GNP series for 1889-1938 and with quarterly NIPA nominal GNP data adjusted using the CPI, for 1939-1946. We use similar procedures to create the GDP deflator.

Real government spending is derived by dividing nominal government purchases by the GDP deflator. Government purchases include all federal, state, and local purchases, but exclude transfer payments. We splice Kendrick's (1961) annual series from 1889 to 1928 to annual NIPA data starting in 1929. Following Gordon and Krenn (2010), we use monthly federal outlay series from the NBER Macrohistory database to interpolate annual government spending from 1889 to 1938. We use the 1954 quarterly NIPA data from 1939-1946 to interpolate the modern series. We follow a similar procedure for federal receipts.

Figure 1 shows the logarithm of real per capita government purchases and GDP. We include vertical lines indicating major military events, such as WWI, WWII and the Korean War. It is clear from the graph that both series are quite noisy in the pre-1939 period. This behavior stems from the interpolator series, especially in the case of government spending. Part of this behavior owes to the fact that the monthly data used for interpolation includes

2. For example, the unemployment rate fell from over 10 percent to 5 percent between mid-1941 and mid-1942.

3. Gordon and Krenn (2010) use similar methods to construct quarterly data back to 1919. We constructed our own series rather than using theirs in order to include WWI in our analysis.

government transfers and part is due to the vagaries of government budget accounting. For this reason, our government spending series should ***not*** be used to identify shocks using standard Choleski decompositions (such as the method of Blanchard and Perotti (2002)). Fortunately, the measurement errors are less of an issue for us because we identify the shocks using narrative methods.[4]

The unemployment series is constructed by interpolating Weir's (1992) annual unemployment series, adjusted for emergency worker employment.[5] Before 1948 we use the monthly unemployment series available from the NBER Macrohistory database back to April 1929 to interpolate. Before 1929, we interpolate Weir's (1992) annual unemployment series using business cycle dates and the additive version of Denton's method. Our comparison of the series produced using this method with the actual quarterly series in the post-WWII period reveal that they were surprisingly close.

Because it is important to identify a shock that is not only exogenous to the state of the economy but is also unanticipated, we use narrative methods to extend Ramey's (2011) defense news series. This news series focuses on changes in government spending that are linked to political and military events, since these changes are most likely to be independent of the state of the economy. Moreover, changes in defense spending are anticipated long before they actually show up in the NIPA accounts. For a benchmark neoclassical model, the key effect of government spending is through the wealth effect. Thus, the news series is constructed as changes in the expected present discounted value of government spending. The particular form of the variable used as the shock is the nominal value divided by one-quarter lag of nominal GDP. We display this series in later sections when we construct the states so that one can see the juxtaposition.

3 Econometric Methodology and Issues

In this section, we first describe the methodology we employ. We then discuss two important econometric issues that have arisen in the literature, and show how the methodology we use overcomes those problems in a straightforward way.

4. Because our shock is constructed independently from news sources and we regress both government spending and GDP on the shock and use the ratio of coefficients, our method is much less sensitive to measurement error in any of the series. See the appendix of Ramey (2011) and footnote 14 of Mertens and Ravn (2013) for a discussion.

5. Because we use the unemployment series to measure slack, we follow the traditional method and include emergency workers in the unemployment rate.

3.1 Model Estimation using Local Projections

Auerbach and Gorodnichenko (2013) were the first to recognize the advantages of using Jordà's (2005) local projection technique to estimate state-dependent models and calculate impulse responses, applying this method in their analysis of state-dependent multipliers in OECD panel data.[6] The Jordà method simply requires estimation of a series of regressions for each horizon h for each variable. The linear model looks as follows:

$$(1) \qquad x_{t+h} = \alpha_h + \psi_h(L)y_{t-1} + \beta_h shock_t + quartic\ trend + \varepsilon_{t+h}, \text{for h = 0, 1, 2, ...}$$

x is the variable of interest (discussed in detail in Section 3.3), y is a vector of control variables, $\psi_h(L)$ is a polynomial in the lag operator, and *shock* is the identified shock. The shock is identified as the defense news variable scaled by lagged nominal GDP. Our vector of baseline control variables, y, contains logs of real per capita GDP, government spending and federal revenues, and $\psi(L)$ is a polynomial of order 4.[7] As discussed in Francis and Ramey (2009), it is potentially important to include a quadratic trend in the post-WWII period because of slow-moving demographics. Since our current sample is twice as long, we include a quartic trend. The coefficient β_h gives the response of x at time $t + h$ to the shock at time t. Thus, one constructs the impulse responses as a sequence of the β_h's estimated in a series of single regressions for each horizon. This method stands in contrast to the standard method of estimating the parameters of the VAR for horizon 0 and then using them to iterate forward to construct the impulse response functions.

The local project method is easily adapted to estimating a state-dependent model. For the model that allows state-dependence, we estimate a set of regressions for each horizon h as follows:

$$(2) \qquad \begin{aligned} x_{t+h} \;=\; & I_{t-1}\left[\alpha_{A,h} + \psi_{A,h}(L)y_{t-1} + \beta_{A,h} shock_t\right] \\ & +(1 - I_{t-1})\left[\alpha_{B,h} + \psi_{B,h}(L)y_{t-1} + \beta_{B,h} shock_t\right] + quartic\ trend + \varepsilon_{t+h}. \end{aligned}$$

I is a dummy variable that indicates the state of the economy when the shock hits. We allow all of the coefficients of the model (other than deterministic trends) to vary according

6. Stock and Watson (2007) also explore the properties of this method for forecasting.
7. Note here in departure from Owyang et al. (2013a), we additionally use log of tax revenues as a control variable.

to the state of the economy. Thus, we are allowing the forecast of x_{t+h} to differ according to the state of the economy when the shock hit. The only complication associated with the Jordà method is the serial correlation in the error terms induced by the successive leading of the dependent variable. Thus, we use the Newey-West correction for our standard errors (Newey and West (1987)).

Apart from the advantages specific to estimating state-dependent multipliers that we will discuss in the next two sections, the Jordà method has the advantage that it does not constrain the shape of the impulse response function, so it is less sensitive to mis-specification of the SVAR. Second, it does not require that all variables enter all equations, so one can use a more parsimonious specification. A third advantage is that the left-hand-side variables do not have to be in the same form as the right-hand-side variables. As we will explain below, this is an important advantage over a standard SVAR in this particular context.

The Jordà method does not uniformly dominate the standard SVAR method for calculating impulse responses, though. First, because it does not impose any restrictions that link the impulse responses at h and $h + 1$, the estimates are often erratic because of the loss of efficiency. Second, it sometimes displays oscillations at longer horizons. Ramey (2012) compares impulse responses estimated using Jordà's method to both a standard VAR and a dynamic simulation (such as the one used by Romer and Romer (2010)), based on military news shocks. The results are qualitatively similar for the first 16 quarters, though the responses using the Jordà method tend to be more erratic. However, at longer horizons, the Jordà method tends to produce statistically significant oscillations not observed in the other two methods. Since we are interested in the shorter-run responses, the long-run estimates are not a concern for us.

3.2 Estimating Impulse Responses in Nonlinear Models

Government spending multipliers are usually estimated by comparing impulse response functions of output and government spending, since these capture the dynamics of the responses to an initial shock. As Koop et al. (1996) point out, constructing impulse responses in nonlinear models is far from straightforward since many complexities arise when one moves from linear to nonlinear systems. In a linear model, the impulse responses are invariant to history, proportional to the size of the shock, and symmetric in positive and negative shocks. In a nonlinear model, the response can depend differentially on the magnitude and sign of the shock, as well as on the history of previous shocks. If one estimates the parameters of a nonlinear model and then iterates on those parameters to construct impulse responses, assumptions on how the economy transitions from state-to-state, as well as the feedback of the shocks to the state, are key components of the constructed responses.

Auerbach and Gorodnichenko (2012) calculate their baseline impulse responses under the assumption that the economy stays in its current state for the 20 quarters over which they compute their multiplier. This assumption turns the problem into a linear one, but it is potentially inconsistent with the data and the multipliers actually estimated. We will discuss this in more detail later when we compare our results to theirs.

The Jordà method uses a completely different method for constructing impulse responses. In that method, the impulse responses at each horizon are estimated directly by regressing x_{t+h} on the shock in period t and lagged values of other control variables. Since separate regressions are estimated for each horizon h, no iteration is involved. The estimated parameters depend on the average behavior of the economy in the historical sample between t and $t + h$, given the shock, the initial state, and the control variables. As an example, consider the forecast of x_{t+6} when $shock_t = 1$ and the economy starts in State A. The parameter estimates on the control variables incorporate the average tendency of the economy to evolve between states even without government spending shocks. Thus, if the duration of State A is typically short relative to State B, the six quarter ahead forecast will take this into account, and will incorporate the natural tendency of the economy to exit State A quickly. The estimate of $\beta_{A,6}$ tells us that six quarters hence x will be $\beta_{A,6}$ units higher than if the shock had been 0. $\beta_{B,6}$ answers the same question for the case when the economy starts in State B. Both of these parameter estimates also incorporate the average tendency of the government spending shock to push the economy from one state to the other by quarter 6. Thus, the estimates incorporate both the natural transitions and endogenous transitions from state to state that occur on average in the data. As we will show in a later section, it is the contrast between this aspect of the impulse responses produced by the Jordà method with the assumptions used by Auerbach and Gorodnichenko (2012) that explains the difference in our results.

3.3 Pitfalls in Converting Elasticities to Multipliers

We now highlight a potential problem that affects multipliers computed not only from nonlinear VARs but also from all of the standard linear SVARS used in the literature. The usual practice in the literature is to use the log of variables, such as real GDP, government spending, and taxes. However, the estimated impulse response functions do not directly reveal the government spending multiplier because the estimated elasticities must be converted to dollar equivalents. Virtually all analyses using VAR methods obtain the spending multiplier by using an *ex post* conversion factor based on the sample average of the ratio of GDP to government spending, Y/G.

We first noticed a potential problem with this method when we extended our sample back in time. In the post-WWII sample, Y/G varies between 4 and 7, with a mean of 5. In our full sample from 1889-2011, Y/G varies from 2 to 24 and with a mean of 8. We realized that we could estimate the same elasticity of output with respect to government spending, but derive much higher multipliers simply because the mean of Y/G was so much higher.

To determine whether using *ex post* conversion factors can lead to inflated multipliers, we conducted a test based on the following point made by Ramey (2013). If the multiplier exceeds one, then it must be the case that private spending $Y - G$ rises when G rises. Thus, one can compare the multipliers estimated the standard way to the response of real private spending to see if there is a contradiction.

To conduct this test, we first estimate a trivariate SVAR with military news, log real per capita government spending, and log real GDP, using four lags and quartic trend, on data from 1889 - 2011. The estimated elasticity is around 0.23 (based on the ratio of the peak of response of $ln(Y)$ to the peak of the response of $ln(G)$). We then multiply the estimated elasticity by the average of Y/G for the full sample, and obtain an implied multiplier of 1.84. To conduct the comparison, we next estimate a model in which we substitute the log of real private spending for log real GDP, and compute the impulse response functions (using the standard method). These responses show that private spending *falls* when government spending rises. Thus, these results imply a multiplier that is less than unity. It appears that the practice of backing out multipliers using *ex post* conversion factors can lead to upward biased multiplier estimates.

To avoid this bias, we follow Hall (2009) and Barro and Redlick (2011) and convert GDP and government spending changes to the same units *before* the estimation. In particular, our x variables on the left-hand-side of equation 2 are defined as $(Y_{t+h} - Y_{t-1})/Y_{t-1}$ and $(G_{t+h} - G_{t-1})/Y_{t-1}$. The first variable can be rewritten as:

$$\frac{Y_{t+h} - Y_{t-1}}{Y_{t-1}} \approx (lnY_{t+h} - lnY_{t-1})$$

and hence is analogous to the standard VAR specification. The second variable can be rewritten as:

$$\frac{G_{t+h} - G_{t-1}}{Y_{t-1}} \approx (lnG_{t+h} - lnG_{t-1}) \, . \frac{G_{t-1}}{Y_{t-1}}$$

Thus, this variable converts the percent changes to dollar changes using the value of G/Y at each point in time, rather than using sample averages. This means that the coefficients from the Y equations are in the same units as those from the G equations, which is required for constructing multipliers.

It would be difficult to use this transformation in a standard SVAR, since all the variables on the left and right must be of the same form. It is easy to use it in the Jordà framework since the variables on the right side of the equation are control variables that do not have to be the same as the left-hand-side variables.

There are some potential econometric concerns about the Hall-Barro-Redlick transformation, however.[8] One concern is that measurement error in Y_{t-1} induces biases because it appears in the denominator of both the news variable and the spending and GDP variables. We explored the impact of this potential bias by re-estimating everything with potential nominal GDP in the denominator of the news variable.[9] The results were very similar. A second concern is that cyclicality of Y_{t-1} can lead to biases in the state-dependent estimates. To explore this potential problem, we re-estimated our model substituting potential GDP for Y_{t-1} in the denominator of the government spending variable, the GDP variable, and the news variable. Again, our results were little changed. Thus, we did not detect any biases in our estimates from using this transformation.

4 Multipliers During Times of Slack

We now conduct an extended analysis of whether the state of slack in the economy affects government spending multipliers. We begin our analysis of multipliers during times of slack by first noting the gap in the theoretical literature. Section 4.2 discusses our measure of slack and shows graphs of the data and periods of slack. Section 4.3 presents the main results. The later sections conduct robustness checks and analyzes in detail why our results are different from Auerbach and Gorodnichenko (2012).

8. Based on private correspondence with Danny Shoag, Alan Auerbach and Yuriy Gorodnichenko.
9. See the data appendix for details on how this was constructed.

4.1 Theoretical Literature

The original Keynesian notion that government spending is a more powerful stimulus during times of high unemployment and low resource utilization permeates undergraduate textbooks and policy debates. Surprisingly, there is no modern DSGE model that produces this most-Keynesian of ideas.

Numerous papers explore theoretically the possibility of state-dependent multipliers that depend on the debt-to-GDP ratio, the condition of the financial system, and exchange rate regimes.[10] Other than the zero lower bound papers, which make a distinct argument that we will discuss below, there is only one paper of which we are aware that analyzes a rigorous model that produces fiscal multipliers that are higher during times of high unemployment. Michaillat (2014) develops a search and matching model and shows that the multiplier on one particular type of government spending doubles as the unemployment rate rises from 5 percent to 8 percent. In particular, he analyzes government spending on public employment. However, Michaillat (2014) does not model the original Keynesian notion that arbitrary government spending can stimulate private employment. Thus, there is still a gap between Keynes' original notion and modern theories.

4.2 Measurement of Slack States

There are various potential measures of slack, such as output gaps, the unemployment rate, or capacity utilization. Based on data availability and the fact that it is is generally accepted as a key measure of underutilized resources, we use the unemployment rate as our indicator of slack. We define an economy to be in a slack state when the unemployment rate is above some threshold. For our baseline results, we use 6.5 as the threshold based on Federal Reserve Chairman Bernanke's recent announcement about policy. We also conduct various robustness checks using time varying thresholds.

Note that our use of the unemployment rate to define the state is different from using NBER recessions or Auerbach and Gorodnichenko's (2012) moving average of GDP growth. The latter two measures, which are highly correlated, indicate periods in which the economy is moving from its peak to its trough. A typical recession encompasses periods in which unemployment is *rising* from its low point to its high point, and hence is not an indicator of a state of slack. Only half of the quarters that are official recessions are also periods of high unemployment.

10. See Corsetti et al. (2012) for a brief survey of this literature.

Figure 2 shows the unemployment rate and the military spending news shocks. The largest military spending news shocks are distributed across periods with a variety of unemployment rates. For example, the largest news shocks about WWI and the Korean War occurred when the unemployment rate was below the threshold. In contrast, the initial large news shocks about WWII occurred when the unemployment rate was still very high.

Because our method for estimation can be interpreted as an instrumental variables regression, it is important to gauge the relevance of the news variable as an instrument. The upper panel of Table 1 shows the F-tests for the exclusion of the news variables.[11] The table shows these for the full sample as well as the post-WWII sample, and splits each of these according to whether the unemployment rate is above 6.5 percent. According to Staiger and Stock (1997), a first-stage F-statistic below 10 can indicate that the instrument may have low relevance. For the full historical and post-WWII samples, the statistics are all around 10 or above. The F-statistics fall when the samples are split between slack and non-slack quarters. While they are less than the safety threshold of 10, most are substantially higher than the typical macro F-statistic. The exception is the slack state in the post-WWII period, where the F-statistic is barely above 1. This low statistic indicates that not much can be learned about the difference in multipliers between slack and non-slack states in the post-WWII period. The F-statistics during slack states are much higher for the full historical sample. This difference supports our initial conjecture that the post-WWII sample was not sufficiently rich to be able to distinguish multipliers across states using the military news instrument.[12]

4.3 Baseline Results for Slack States

We now present the main results of our analysis using the full historical sample and the local projections method. We first consider results from the linear model, which assumes that multipliers are invariant to the state of the economy. The top panel of Figure 3 shows the responses of government spending and output to a military news shock in the linear model using the U.S. data. The bands are 95 percent confidence bands and are based on

11. The F-tests are based on regressions of log real per capita government spending on four lags of its own value, four lags of log real per capita GDP, four lags of log real per capita federal receipts, and the current value and four lags of the news variable (scaled by the previous quarter's nominal GDP), as well as a quartic trend.

12. In contrast, the Blanchard and Perotti (2002) identification scheme is almost guaranteed to produce shocks with high F-statistics since the shock is identified as the part of current government spending not explained by the other lagged variables in the SVAR. However, this type of shock is much more sensitive to measurement error and is subject to the critique that it is likely to have been anticipated.

Newey-West standard errors that account for the serial correlation induced in regressions when the horizon $h > 0$. After a shock to news, output and government spending begin to rise and peak at around 12 quarters.

In the linear model, the multipliers are derived from the estimated β_h's from the set of Y and G regressions. We compute multipliers over three horizons: the ratio of the peak of the GDP response to the peak of the government spending response, the ratio of the cumulative responses through two years, and the cumulative responses through four years.[13] The integral multipliers are the closest to answering the relevant question for policy because they measure the cumulative GDP gain relative to cumulative government spending during a given period. We focus on the two-year and four-year horizons which are most relevant for short-run stimulus policy. As indicated in the first column of the top panel of Table 2, the implied multipliers are below one and range from 0.8 to 0.9. The estimates are not statistically different from one at the five percent significance level.[14]

The main question addressed in this paper is whether the multipliers are state-dependent. The impulse response functions and multipliers in the state-dependent case are derived from the estimated $\beta_{A,h}$ and $\beta_{B,h}$ for Y and G in equation 2. The bottom panel of Figure 3 shows the responses when we estimate the state-dependent model where we distinguish between periods with and without slack in the economy. Similar to many pre-existing studies (e.g. Auerbach and Gorodnichenko (2012)), we find that output responds more robustly during high unemployment states. Note that government spending also has a stronger response during those high slack periods. Consequently, the larger output response during the high unemployment state does not imply a larger government spending multiplier. In fact, as shown in the second and third columns of Table 2, the implied multipliers are slightly lower during the high unemployment state. In no case do we find a statistically different multiplier across states.

To summarize, in our linear model we find multipliers that are less than 1 in all cases. Considering state dependence, we find no evidence of larger multipliers in the periods of slack, and multipliers vary between 0.8 and 1.

These results also answer a question that has plagued those who include WWII in the

13. To further clarify, the peak multiplier is given as $\frac{max_{i=1...20}\{\Delta Y_i\}}{max_{i=1...20}\{\Delta G_i\}}$ and the cumulative multipliers are constructed as $\frac{\sum_{i=1}^{M}\Delta Y_i}{\sum_{i=1}^{M}\Delta G_i}$ for $M = 8$ and 16, where Δ denotes the difference between the path conditional on the shock versus no shock.

14. The variance-covariance matrix is computed by estimating all of the regressions as one panel regression and using the Newey-West procedure to adjust the standard errors. We implement this estimation in Stata, using *newey* followed by *nlcom*, which uses the delta method to analyze functions of parameters.

analysis. Some researchers have rightly worried that wartime rationing might have depressed multipliers starting in 1942. However, the unemployment rate had fallen below the 6.5 percent threshold by the second quarter 1942, when rationing began to be applied in earnest. All of the periods with significant rationing are classified as non-slack states in our analysis, so if anything, rationing would serve to lower the multiplier in the non-slack state relative to the slack state. Yet we still find no differences across states.[15]

4.4 Robustness

Our baseline results suggest that there is no difference in multipliers across slack states. These results are potentially sensitive to the numerous specification choices we made that were not guided by theory. Thus, in this section we explore the sensitivity of our findings to these choices.

We first investigate the impact of using a different interpolation method for the data. Recall that our underlying interpolators were quite volatile and led to a lot of jumps in the early data. To investigate the impact, we create alternative data that uses linear interpolation of the annual data in the pre-WWII period. When we re-estimate the model, we find slightly lower multipliers on average, and no difference in multipliers across states of the economy. These results are shown in the first panel of Table 3.

We included the quartic trend in our equations to capture low frequency demographics, such as the Baby Boom. Since times series estimates can be sensitive to trends, we investigate the impact of omitting the quartic trend in our model. As shown in the second panel of Table 3, this specification produces slightly lower multipliers, but no difference in multipliers across states of the economy.

Next we consider a time-varying threshold, where we consider deviations from trend for Hodrick-Prescott filtered unemployment rate with a very high smoothing parameter of $\lambda = 1,000,000$. This definition of threshold results in about 40 percent of the observations being above the threshold . As shown in Figure 4, this threshold also suggests prolonged periods of slack both in the late 1890s and during the 1930s. There is substantial evidence that the "natural rate" of unemployment displayed an inverted U-shape in the post-WWII period, and this time-varying threshold also helps account for this. Using this time-varying threshold, we find results in line with our baseline findings, with multipliers typically less

15. The results are very similar if we redefine the unemployment threshold to be 7 percent rather than 6.5 percent in order to classify 1942q1 to be non-slack.

than one for the state-dependent case and no significant difference between the multipliers in the low unemployment state and the high unemployment state (see Table 3).

Next, we consider a threshold based on the moving average of output growth, as in Auerbach and Gorodnichenko (2012). We construct a smooth transition threshold, where we replace the dummy variable I_{t-1} in equation (2) with the function $F(z_t)$, where z is the normalized 7-quarter centered moving average of output growth.[16] Figure 5 shows the function $F(z)$ along with the NBER recessions for our full sample. Results in the bottom part of Table 3 show that when we use this weighting function for recessionary regimes in our specification to construct state-dependent multipliers, we still get multipliers less than one for U.S. across both recession and expansion regimes, and do not find any evidence of higher multipliers in expansions versus recessions.

Another point of departure with the pre-existing literature is the fact that most of the papers employ a shorter data sample that spans the post World War II period. As a robustness check we limit our sample to this period, 1948-2011, and employ the Jorda local projection method on this data set. In this shorter sub-sample too, about 30 percent of the observations are above our baseline threshold for unemployment rate, signifying state of slack.[17] As shown in the fourth panel of Table 3, in the linear case, the multipliers for U.S. are still smaller than 1. Looking at state-dependent multipliers, we find that the multiplier in the high unemployment state jumps around. The two year integral multiplier in high unemployment state is large and negative at -2 and the four year integral multiplier is large and positive taking a value of 18! Since the military news variable has very low instrument relevance during slack periods in the post-WWII period, the impulse responses in this state are very imprecisely estimated. Also, rather counter-intuitively in this sub-sample, output has a negative response to the news shock in the high unemployment state, and the government spending response also becomes negative after 2-3 years. Thus, it is hard to take these state-dependent multipliers for the sub-sample seriously.

4.5 Comparison to Auerbach-Gorodnichenko (2012, 2013)

Our finding that multipliers do not differ across slack states stands in contrast to two studies by Auerbach and Gorodnichenko (2013, 2014). In this section, we explain the source of

16. We use the same definition of $F(z)$ as Auerbach and Gorodnichenko (2012), which is given by $F(z_{t-1}) = \frac{\exp(-\gamma z_{t-1})}{1+\exp(-\gamma z_{t-1})}$ and set $\gamma = 3$, in order to ensure that $F(z)$ is greater than 0.8 close to 30 percent of the time for the U.S., which lines up with the total duration of recessions during our full sample starting in 1889.

17. When conducting this sub-sample analysis we change our baseline specification to use a quadratic trend.

the different results.

We first compare our results to Auerbach and Gorodnichenko (2012) (AG-12). They use a smooth transition VAR (STVAR) model, post-WWII data, the Blanchard-Perotti identification method, and a function of the 7-quarter moving average of normalized real GDP growth as their measure of the state. They also include the current value and three lags of the actual 7-quarter moving average growth rate as exogenous regressors in their model.[18] They construct their baseline impulse responses by iterating on the estimated parameters assuming that the economy remains in its initial state for at least 20 quarters.

The following exercise reveals the the main source of the difference in our results. We take only one step away from what AG-12 did by using all details of their analysis except the method for estimating and constructing the impulse responses. In particular, we apply the Jordà method to their post-WWII data, using their exact definition of states, identification of shocks, use of logs of variables, and inclusion of current and lagged values of the centered 7-quarter moving average of output growth as a control. As discussed earlier, the Jordà method directly provides the estimates of the impulse responses.

Figure 6 shows the linear responses in the top panel.[19] The government spending response looks similar to the linear case in AG-12, though the GDP response is more erratic and the standard error bands are much wider. However, the state-dependent responses shown in the lower panel look very different. The Jordà method produces impulse responses in which the response of government spending to a shock is higher in a recession than in an expansion, similar to our earlier results, but in opposition to those of AG-12. The response of output differs little across states, in sharp contrast to AG-12 who find that output rises robustly throughout the 20 quarters in the recession state but quickly reverts to 0 or becomes negative in the expansion state.

Table 4 reproduces AG-12's multipliers in the first panel and those estimated by the Jordà method in the second panel. Following AG-12, we look at both the peak multipliers and the 20 quarter integral multipliers.[20] Consider first the linear case. Whereas AG-12

18. The published paper does not discuss these additional terms, but the initial working paper version includes these terms in one equation and the codes posted for the published paper include them. Our replications indicate that they are included.

19. Note that the online files indicate that AG-12 use only government consumption as their measure of G. As a result their Y/G ratio used to convert multipliers is 5.6, which is higher than the usual one based on total government purchases.

20. As noted in the table, we recomputed their peak multipliers because they did not follow standard practice of comparing the peak of output to the peak of government spending. Instead, they compared the peak of output to the initial government spending shock, which is not really a multiplier.

obtained multipliers of 0.68 for the peak and 0.57 for the 20-quarter integral, we find higher multipliers, around one in both cases. The results differ markedly for the state dependent case. AG find multipliers of 2.2 to 2.5 for recessions, and between -0.33 and 0.28 for expansions. We find multipliers of 0.9 to 1.1 in recessions and 0.5 to 1 in expansions.

Thus, the difference between our results and those of AG-12 is not due to the sample period, the definition of slack, the definition of the left-hand-side variables, or the identification method, but rather is due to the method for calculating the impulse responses. AG-12 calculate their baseline impulse responses under the assumption that the economy stays in its current state for at least 20 quarters. This assumption turns the problem into a linear one, greatly simplifying the construction of impulse responses, but it is potentially problematic for two reasons. First, as Figure 5 shows, during the post-WWII period the recession states are much shorter than 20 quarters in duration; the mean duration is 3.3 quarters. This inconsistency of their assumption with the data means that the multipliers they calculate for recessions are based on impulse responses that do not represent any episode ever experienced in their sample. Second, the assumption implies that a positive shock to government spending during a recession does not help the economy escape the recession. Yet this very assumption implies multipliers of 2.2 to 2.5 in a recession state. Thus, the crucial assumption used to calculate the multipliers - that government spending does not help the economy escape a recession - is at odds with the result that government spending has powerful stimulus effects on output during recessions.

To determine how this assumption affects their results, we compute alternative impulse responses using their data and their STVAR parameter estimates. As discussed earlier, impulse responses in nonlinear models depend on the history of shocks, the size of the shocks and sign of the shocks. One approach is to use Koop et al. (1996)'s generalized impulse response functions (GIRFs), which computes impulse responses conditional on histories of shocks and then integrates over all observed histories. We are investigating this method in ongoing work (Owyang et al. (2013b)). In the current paper, we conduct a specific experiment to illustrate our point. We consider the difference in multipliers for a government spending shock that hits in 1991q1 versus one that hits in 1993q1. Why this period? According to the AG-12 definition, the U.S. economy was in a recession state for seven quarters in 1990 through 1991. Thus, this AG-defined recession was longer than average. Once the economy entered the expansion, it stayed there until 2001q3. Thus, this episode constitutes a perfect experiment because it featured a longer-than-average recession (when there may have been time to enact a stimulus package during the recession), and did not have a back-to-back recession or a shortened expansion as did some other periods in the sample.

The NBER peak of the business cycle was in August 1990. We can then use AG-12's parameter estimates to ask the following question. Suppose that the fiscal authorities recognized the recession quickly and implemented a stimulus package in 1991q1. What would the impact be relative to increasing government spending during an expansion? We set the initial shock as 0.8 percent of government purchases, so that the cumulative increase in government spending over the 20 quarters is about 14 percent. This is about half of the size of the 2009 stimulus package.[21] We construct the impulse responses conditional on the initial state of the economy as well as on the actual sequence of GDP shocks and tax shocks over the 20 quarters, and we allow feedback of the shocks to the state. We then run the same experiment for 1993q1, which was the start of a long expansion.

The third panel of Table 4 shows the results from this exercise. We find multipliers that are much closer to those from our Jordà method estimates than to those constructed by AG-12. The multipliers are somewhat higher during the recessionary quarter of 1991q1 than in the expansionary quarter of 1993q1, but the differences are relatively small. Thus, the results and conclusions change significantly once we allow feedback.[22]

The results obviously depend on the historical period as well as on the particular size and sign of the shock. However, one can easily conduct a variety of these types of experiments for other historical periods. This experiment illustrates an advantage of AG-12's STVAR method over the Jordà method. The Jordà method produces multipliers based on averages over the estimation sample, whereas the STVAR estimates can in principle be used to conduct a variety of experiments.

As discussed earlier, the second Auerbach-Gorodnichenko paper (AG-13) also applies the Jordà method to a panel of OECD countries; in fact, AG-13 were the ones who first realized the potential of this method for state-dependent models. Thus, a key question is why they find higher multipliers during recessions even with this method. There is, of course, the obvious difference in time period and country sample. We believe, however, the most likely reason for the difference is in two details of how they calculate multipliers. First,

21. However, the 2009 stimulus package consisted of mostly transfers rather than government purchases.

22. AG-12's Figure 3 shows a time series of multipliers, which are described as having been calculated allowing the government spending shocks to feed back to the state of the economy. They find multipliers near 1.5 for 1991, falling to about 0.5 by 1993. We were unable to determine the assumptions under which these multipliers were computed, and hence are unable to shed light on why our recession multipliers are lower than theirs.

following the standard practice, they estimate everything in logarithms and then use the *ex post* conversion factor based on average Y/G during their sample to convert elasticities to multipliers. Second, they depart from the rest of the literature by comparing the path or peak of output to the *impact* of government spending rather than to the peak or integral of the path of government spending. This is a big difference because the effects of a shock to government usually build up for several quarters. This is also not the type of multiplier policy makers are interested in because it does not count the average cumulative cost of government spending associated with the path.

We now demonstrate the difference that these two details make by applying *their* method for calculating multipliers to *our* U.S. historical data. First, we estimate our baseline model in logs and use the sample average of Y/G of 8 to convert the elasticities. This results in 2-year integral multipliers of 1.85 in the high unemployment state and 1.72 in the low unemployment state. Thus, this change in method *doubles* the constructed multiplier relative to our baseline method, but does not lead to a difference across states. Second, still using the log specification and the conversion factor, we instead calculate multipliers by dividing the average response of output over the 2-year horizon by the the *initial* shock to government spending. In this case, we calculate 2-year multipliers of 7.27 for the high unemployment state and 5.24 for the low unemployment state. Thus, this method not only produces multipliers that are huge in both states, but also induces a large difference between states. As our figures make clear, this difference shows up because this calculation does not take into account the fact that government spending rises more robustly after an initial shock during a recession. Thus, it is clear that even using the same estimation method and same method for computing impulse responses, details of the calculations of multipliers can make a big difference.

4.6 The Behavior of Taxes

All of our regressions include taxes as controls, but our analysis so far has ignored the responses of taxes. Romer and Romer's (2010) estimates of tax effects indicate very significant negative multipliers on taxes, on the order of -2 to -3. Thus, it is important for us to consider how the increases in government spending are financed in order to interpret our multiplier results.

To analyze how taxes and deficits behave, we re-estimate our basic model augmented to include deficits, and with tax rates substituted for real tax revenues so that we can distinguish increases in revenues caused by rising output versus rising rates. Tax rates are computed as the ratio of federal receipts to nominal GDP, and represent average tax rates rather than marginal tax rates. The deficit is the real total deficit. We include four lags of the logs of these two new variables along with GDP and government spending as controls all of the regressions. We estimate this system for the full sample using the Jordà method.

Figure 7 shows the results from the linear case. The responses of government spending and GDP are almost identical to the baseline case. The bottom panels show that both average tax rates and the deficit increase, with deficits increasing more rapidly at the beginning and tax rates rising more slowly and more persistently. Taking the ratio of the path of deficits to government spending, we estimate that most of the increase in government spending during the first year is financed by deficits. The deficit fraction of government spending hits a peak at quarter 4, when 70 percent of that quarter's increase in government spending is financed by deficits. It stays high for about a year and then begins to decline.

From a theoretical perspective, the fact that tax rates respond more slowly than spending has significant implications for the multiplier. If all taxes are lump-sum taxes, news about a future increase in the present discounted value of government spending leads to an immediate jump in hours and output because of the negative wealth effect. In a neoclassical model, the effect is the same whether the taxes are levied concurrently or in the future. In contrast, the need to raise revenues through distortionary taxation can change incentives significantly. As Baxter and King (1993) show, if government spending is financed with current increases in tax rates, the multiplier can become negative in a neoclassical model.

The situation changes considerably when tax rates are slow to adjust, but agents anticipate higher future tax rates. To see this, consider the case of labor income taxes and a forward-looking household:

$$1 = \beta E_t \left[\frac{u_{n,t+1}}{u_{n,t}} \frac{(1-\tau_t)w_t}{(1-\tau_{t+1})w_{t+1}} (1+r_t) \right] \tag{3}$$

where u_n is the marginal utility of leisure, τ is the tax on labor income, w is the real wage rate, r is the real interest rate, and E_t is the expectation based on period t information. In expectation, the household should vary the growth rate of leisure inversely with the growth rate of after-tax real wages. This means that if τ_{t+1} is expected to rise relative to τ_t, households have an incentive to substitute their labor to the present (when it is taxed less) and their leisure to the future.

It is easy to show in a standard neoclassical model that the delayed response of taxes, such as we observe in the estimated impulse responses, results in a multiplier that is higher in the short-run but lower in the long-run relative to the lump-sum tax case. We have also conducted this experiment in the Gali et al. (2007) model where 50 percent of the households are rule-of-thumb consumers. We found the same effect in that model as well. Drautzburg and Uhlig (2011) analyze an extension of the Smets-Wouters model and also find that the timing of distortionary taxes is very important for the size of the multiplier. Given the impulse response of tax rates, and with these theoretical results in mind, it is very possible that our estimated multipliers are greater than we would expect if taxation were lump-sum.

Nevertheless, our finding that multipliers do not vary across states could be due to differential financing patterns. To determine whether this is the case, Figure 8 shows the state-dependent results. As we showed before, both government spending and GDP rise more if a news shock hits during a slack state, even after adjusting the initial size of the shock. The bottom panels show that tax rates and deficits also rise more during recessions, but there are other interesting differences in the patterns. When we study the ratio of the deficit to government spending at each point in time along the path, we find that more of government spending is financed with deficits when a shock hits during a slack state. For example, at quarter five the ratio of the deficit to government spending is 86 percent if a shock hits during a slack state but only 29 percent if the shock hits during a non-slack state. Comparing the peak of the deficit to the peak of the government spending response, the ratio is 0.6 for the slack state and 0.4 for the no slack state. Thus, on average short-run government spending is financed more with deficits if the shock hits during a slack state. This would imply that the multiplier should be greater during times of slack. In fact, our estimates imply that it is not.

5 Multipliers at the Zero Lower Bound

We now investigate whether government spending multipliers differ when government interest rates are near the zero lower bound or are being held constant to accommodate fiscal policy. As we will discuss shortly, New Keynesian models suggest that government spending multipliers will be higher when the economy is at the zero lower bound. Very few papers have attempted to test the predictions of the theory empirically. As far as we know, only two examples exist. Ramey (2011) estimates her model for the U.S. over the sub-sample from 1939 through 1951 and shows that the multiplier is no higher during that sample. Crafts and Mills (2012) construct defense news shocks for the U.K. and estimate multipliers on quarterly data from 1922 through 1938. They find multipliers below unity even when interest rates were near zero.[23]

5.1 Theoretical Background

Several recent papers have analyzed the effects of fiscal policy in New Keynesian models when the zero lower bound on interest rates prevents nominal interest rates from responding according to the Taylor rule. For example, Eggertsson and Woodford (2003), Eggertsson (2011), and Christiano et al. (2011) show that the government spending multiplier can be much larger if interest rates are at the zero lower bound.[24] The intuition for why the zero lower bound can raise the multiplier is as follows. A deficit-financed increase in government spending leads expectations of inflation to rise. When nominal interest rates are held constant, this increase in expected inflation drives the real interest rate down, spurring the economy. Christiano et al. (2011) show that if interest rates are held constant for 12 quarters and government spending goes up during this time, the multiplier peaks at 2.3. Fernández-Villaverde et al. (2012) take into account the inherent nonlinearities at the ZLB in their analysis, but still find that the government spending multiplier can be three times greater at the zero lower bound. Cogan et al. (2010) and Coenen and et al. (2012) also show in a suite of policy models that monetary accommodation and its duration have consequences for the size of the government spending multiplier. Cogan et al. (2010) find modest consequences, whereas Coenen and et al. (2012) document that the multiplier doubles in most models as they move from no monetary accommodation to two years of monetary accommodation.[25]

In all of the theoretical models, it is not the zero lower bound per se, but rather the fact that nominal interest rates stay constant rather than following the Taylor rule that amplifies the stimulative effects of government spending. Thus, to assess whether multipliers are

23. Bruckner and Tuladhar (2013) focus on local not aggregate multipliers for Japan, and find that the effects of local spending are larger in the ZLB period, but only modestly.

24. The relationship between government spending multipliers and the degree of monetary accommodation, even outside zero lower bound has been explored by many others, including Davig and Leeper (2011) and Zubairy (2014).

25. They point out how across the different models under consideration, the effects of fiscal shocks can be sensitive to assumptions about the duration of the liquidity trap, the degree wage and price rigidities, and the persistence of the spending shock.

greater in these situations we can include periods in which the nominal interest rate is relatively constant despite dramatic fluctuations in government spending.

5.2 Defining States by Monetary Policy

The bottom panel of Figure 9 shows the behavior of three-month Treasury Bill rates from 1920 through the present, where the shortened sample is based on data availability. This interest rate was near zero during much of the 1930s and 1940s, as well as starting again in the fourth quarter of 2008. To indicate the degree to which interest rates were pegged (either by design or the zero lower bound), we compare the behavior of actual interest rates to that prescribed by the Taylor rule. We use the standard Taylor rule formulation:

(4) nominal interest rate = 1 + 1.5 year-over-year inflation rate + 0.5 output gap

Figure 10 shows the behavior of inflation and the output gap, which were quite volatile during the early period. Figure 11 compares the behavior of actual interest rates to the Taylor rule. This graph makes clear that there were large deviations of interest rates from those prescribed by the Taylor rule briefly during the early 1920s and in a sustained way during most of the 1930s and 1940s.

For our baseline, we define ZLB or extended monetary accommodation times to be 1932q2 - 1951q1 and 2008q4 - 2011q4 (the end of our sample). We do not include the early 1920s since the episode was so brief and theory tells us that the multiplier depends on the (expected) length of the spell. Also, while the deviation from the Taylor rule widens starting in 1930, we do not include early 1930s in our ZLB state. This is because the T-bill rate was fluctuating during this period, potentially responding to the state of the economy, and was as high as 2.5 percent in 1932q1 before falling to 0.5 percent in 1932q2 and staying low from then onwards. We will call these periods "ZLB states" for short, recognizing that they also include periods of monetary accommodation of fiscal policy. We end the early spell in 1951q1 because the Treasury Accord, which gave the Fed more autonomy, was signed in March 1951.

The top panel of Figure 9 shows the behavior of our defense news series over the states defined this way. The main shocks during these states occur after the start of WWII and at the start of the Korean War (in June 1950). There is essentially no information gained from the 1930s, unfortunately.[26] Table 5 shows the F-tests for the periods split into ZLB periods

26. An advantage of the Crafts and Mills (2012) analysis of UK data is that it has more defense news

and normal periods. The F-test is rather low for the ZLB period, around, so instrument relevance is a concern.

5.3 Results

To determine whether multipliers are different in ZLB states, we estimate our baseline state-dependent model, but now allowing the state to be defined by monetary policy. We consider our full sample spanning 1889-2011.[27] Figure 12 shows the impulse responses. The results suggest that government spending responds more slowly, but more persistently during ZLB states than in normal states. The difference in GDP responses follow this pattern, but in a muted way. Table 6 shows the multipliers in each state, computed the three different ways. For the peak multiplier, we find that the multiplier is slightly higher in the normal state than the ZLB state, though both are less than unity. For the two-year integral, the multipliers are very similar. Oddly, for the 4-year integral, the multiplier is very high at 1.6 in the normal state, but less than one in the ZLB state. Thus, in no case do we find evidence of significantly higher multipliers during periods at the zero lower bound or constant interest rates.

A major concern is that an important part of the ZLB state was during the rationing period of WWII. If rationing depressed multipliers, and all of the rationing occurred during the ZLB state, then this could explain why we find no differences across periods. To determine whether our results are sensitive to the rationing period, we reclassify 1942q1 through 1945q3 to the "non-ZLB" period. Table 7 shows that this change reduces the 1.6 multiplier (calculated with the 4-year integral) in the normal state to 1, but does not affect the multiplier in the ZLB state. Thus, our finding of no difference in multipliers across ZLB and non-ZLB states cannot be attributed to WWII rationing.

Another robustness check we ran was to simply define the ZLB state as periods where the T-bill rate was less than or equal to 50 basis points. Since our data for the 3-month T-bill rate starts in 1920, we conducted this robustness check over a shorter sub-sample than our baseline exercise, spanning 1920-2011. As shown in the lower panel of Table 7, this re-definition of the monetary state results in multipliers close to or slightly above 1 in the non-ZLB state and lower in the ZLB state. Thus again, we do not find any evidence of

shocks during the 1930s.

27. Even though the 3 month T-bill rate was not available before 1920, we still consider the earlier period and call it a non-ZLB state, based on narrative evidence and data on commercial paper rate for which monthly data is available starting 1857.

higher multipliers in the ZLB state.

These results are contrary to the predictions of the New Keynesian model. While the key to those predictions lie in the behavior of inflationary expectations, we think it unlikely that expectations would have behaved in a way to change the results. Thus, both our results and those of Crafts and Mills (2012) for the UK are inconsistent with the predictions of the New Keynesiam model at the ZLB.[28]

6 Conclusion

In this paper, we have investigated whether government spending multipliers vary depending on the state of the economy. In order to maximize the amount of variation in the data, we constructed new historical quarterly data spanning more than 120 years in the U.S. We considered two possible indicators of the state of the economy: the amount of slack, as measured by the unemployment rate, and whether interest rates were being held constant close to the zero lower bound. Using a more robust method for estimating state-dependent impulse responses and better ways of calculating multipliers from them, we provided numerous estimates of multipliers across different specifications.

Our results can be summarized as follows. We find no evidence of significant differences in multipliers when the U.S. economy is experiencing substantial slack as measured by the unemployment rate. Most multipliers are slightly below unity with a few slightly above unity. Our numerous robustness checks suggest that our results are not sensitive to variations in our specification. We also conducted a detailed analysis of why our results are so different from those of Auerbach and Gorodnichenko (2012). We found the key source of differences are the specialized assumptions Auerbach and Gorodnichenko (2012) use to calculate their impulse response functions.

In our analysis of multipliers in zero lower bound interest rate states, we found that multipliers were never higher at the zero lower bound. Thus, we found no support for the predictions of recent New Keynesian models.

Of course, our results come with many caveats. As discussed in the introduction, we are forced to use data determined by the vagaries of history so we do not have a controlled experiment. Because we use news about future military spending as our identified shock, our results do not inform us about the size of multipliers on transfer payments or infras-

28. In separate work, Wieland (2013) tests the New Keynesian prediction that negative supply shocks are expansionary at the ZLB, and also finds evidence contrary to that prediction.

tructure spending. The results might also be affected by other aspects of war-time periods. Moreover, because the episodes we studied were characterized by certain paths of taxes, the results are not immediately applicable to the case of deficit-financed stimulus packages or fiscal consolidations. Finally, because we study only the U.S., our results are not necessarily inconsistent with analyses of multipliers during recent fiscal consolidations in European economies, such as the work of Blanchard and Leigh (2013).

References

Auerbach, Alan and Yuriy Gorodnichenko, 2012. "Measuring the Output Responses to Fiscal Policy." *American Economic Journal: Macroeconomics* 4(2): 1–27.

Auerbach, Alan and Yuriy Gorodnichenko, 2013. "Fiscal Multipliers in Recession and Expansion." In *Fiscal Policy After the Financial Crisis*, edited by Alberto Alesian and Francesco Giavazzi, pp. 63–98. University of Chicago Press.

Bachmann, Rudiger and Eric R. Sims, 2012. "Confidence and the transmission of government spending shocks." *Journal of Monetary Economics* 59: 235–249.

Balke, Nathan and Robert J. Gordon, 1986. "Appendix B Data Tables." In *The American Business Cycle: Continuity and Change*, edited by Robert J. Gordon, p. http://www.nber.org/data/abc/. University of Chicago Press.

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

Baum, Anja, Marcos Poplawski-Ribeiro, and Anke Weber, 2012. "Fiscal Multipliers and the State of the Economy." Unpublished paper, IMF.

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

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

Blanchard, Olivier J. and Daniel Leigh, 2013. "Growth Forecast Errors and Fiscal Multipliers." NBER Working Paper 18779.

Bruckner, Marcus and Anita Tuladhar, 2013. "Local Government Spending Multipliers and Financial Distress: Evidence from Japanese Prefectures." *Economic Journal,* forthcoming.

Carter, Susan B., Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, Gavin Wright, and Richard Sutch, 2006. "Table Ca9-19 : Gross domestic product: 1790-2002 [Continuous annual series]." In *Historical Statistics of the United States Millennial Edition Online*. Cambridge University Press.

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

Coenen, Günter and et al., 2012. "Effects of Fiscal Stimulus in Structural Models." *American Economic Journal: Macroeconomics* 4(1): 22–68.

Cogan, John F., Tobias Cwik, John B. Taylor, and Volker Wieland, 2010. "New Keynesian versus Old Keynesian Government Spending Multipliers." *Journal of Economic Dynamics and Control* 34: 281–295.

Conference Board, The, 1945. *The Economic Almanac 1945-46*. New York, New York: National Industrial Conference Board, Inc.

Corsetti, Giancarlo, Andre Meier, and Gernot Mueller, 2012. "Government Spending Mutlipliers." *Economic Policy* .

Crafts, Nicholas and Terence C. Mills, 2012. "Fiscal Policy in a Depressed Economy: Was There a 'Free Lunch' in 1930s' Britain?" Unpublished paper, University of Warwick.

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

Drautzburg, Thorsten and Harald Uhlig, 2011. "Fiscal Stimulus and Distortionary Taxation." NBER Working Papers 17111, National Bureau of Economic Research.

Eggertsson, Gauti B., 2011. "What Fiscal Policy is Effective at Zero Interest Rates?" *NBER Macroeconomics Annual* 25: 59–112.

Eggertsson, Gauti B. and Michael Woodford, 2003. "The Zero Bound on Interest Rates and Optimal Monetary Policy." *Brookings Papers on Economic Activity* 34(1): 139–235.

Fazzari, Steven M., James Morley, and Irina Panovska, 2013. "State Dependent Effects of Fiscal Policy." Unpublished paper, Australian School of Business.

Fernández-Villaverde, Jesús, Grey Gordon, Pablo A. Guerrón-Quintana, and Juan Rubio-Ramírez, 2012. "Nonlinear Adventures at the Zero Lower Bound." NBER Working Papers 18058, National Bureau of Economic Research.

Francis, Neville and Valerie A. Ramey, 2009. "Measures of per Capita Hours and Their Implications for the Technology-Hours Debate." *Journal of Money, Credit, and Banking* 41(6): 1071–1097.

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

Gordon, Robert J. and Robert Krenn, 2010. "The End of the Great Depression: VAR Insight on the Roles of Monetary and Fiscal Policy." Working paper 16380, National Bureau of Economic Research.

Hall, Robert E., 2009. "By How Much Does GDP Rise If the Government Buys More Output?" *Brookings Papers on Economic Activity* Fall: 183–236.

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

Kendrick, John D., 1961. *Productivity Trends in the United States*. Princeton New Jersey: Princeton University Press for the National Bureau of Economic Research.

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

Mertens, Karel and Morten O. Ravn, 2013. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." *American Economic Review* 103.

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

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

Newey, Whitney K and Kenneth D West, 1987. "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55(3): 703–08.

Owyang, Michael T., Valerie A. Ramey, and Sarah Zubairy, 2013a. "Are Government Spending Multipliers Greater During Times of Slack? Evidence from 20th Century Historical Data." *American Economic Review* 103(2): 129–34.

Owyang, Michael T., Valerie A. Ramey, and Sarah Zubairy, 2013b. "Estimating Fiscal Multipliers in State-Dependent Models." Work-in-progress.

Ramey, Valerie A., 2011. "Identifying Government Spending Shocks: It's All in the Timing." *Quarterly Journal of Economic* 126(1): 51–102.

Ramey, Valerie A., 2012. "Comment on Roads to Prosperity or Bridges to Nowhere? Theory and Evidence on the Impact of Public Infrastructure Investment." In *NBER Macroeconomics Annual 2012, Volume 27*, NBER Chapters, pp. 147–153. National Bureau of Economic Research.

Ramey, Valerie A., 2013. "Government Spending and Private Activity." In *Fiscal Policy After the Financial Crisis*, edited by Alberto Alesian and Francesco Giavazzi, pp. 19–55. University of Chicago Press.

Romer, Christina D. and David H. Romer, 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." *American Economic Review* 100(3): 763–801.

Semmler, Willi and Andre Semmler, 2013. "The Macroeconomics of the Fiscal Consolidation in the European Union." Working paper.

Staiger, Douglas and James H. Stock, 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65(3): 557–586.

Stock, James H. and Mark Watson, 2007. "Why Has US Inflation Become Harder to Forecast?" *Journal of Money, Credit, and Banking* 39(1): 3–33.

Weir, David R., 1992. "A Century of U.S. Unemployment, 1890-1990: Revised Estiamtes and Evidence for Stabilization." In *Research in Economic History*, edited by Roger L. Ransom, pp. 301–346. JAI Press.

Wieland, Johannes, 2013. "Are Negative Supply Shocks Expansionary at the Zero Lower Bound?" Working paper, University of California, San Diego.

Zubairy, Sarah, 2014. "On Fiscal Multipliers: Estimates from a Medium-Scale DSGE Model." *International Economic Review,* forthcoming.

Data Appendix

GDP and GDP deflator:

1947 - 2011: Quarterly data on chain-weighted real GDP, nominal GDP, and GDP deflator from BEA NIPA (downloaded from FRED, December 20, 2012 revision.)

1889 - 1946: Annual data from 1929 - 1946 from BEA NIPA (downloaded from FRED, December 20, 2012 version). For 1889 - 1928, series Ca9 and Ca13 from Table Ca9-19 in Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition, Carter et al. (2006). These series are based on the work of Kuznets, Kenrick, Gallman and Balke-Gordon.

1939 - 1946: We used seasonally adjusted quarterly nominal data on GNP from National Income, 1954 Edition, A Supplement to the Survey of Current Business and seasonally unadjusted CPI (all items, all urban consumers) from FRED.

1889 - 1938: Quarterly data on real GNP and GNP deflator. Source: Balke and Gordon (1986). Data available at: http://www.nber.org/data/abc/

Data adjustment: For 1939-1946, we used a simplified version of the procedure used by Valerie Ramey, "Identifying Government Spending Shocks: It's All in the Timing", Quarterly Journal of Economics, February 2011. We used the quarterly nominal GNP series published in *National Income, 1954 Edition, A Supplement to the Survey of Current Business* to interpolate the the modern NIPA annual nominal GDP series, and the quarterly averages of the CPI to interpolate the NIPA annual GDP price deflator using the proportional Denton method. We took the ratio to construct real GDP to use as a second round interpolator. We spliced this quarterly real GDP series to the Balke-Gordon quarterly real GNP series from 1889 - 1938 and used the combined series to interpolate the annual real GDP series (described above) using the proportional Denton method. This method insures that all quarterly real GDP series average to the annual series. We used the Balke-Gordon deflator to interpolate the annual deflator series from 1889 - 1938 and combined it with the CPI-interpolated series from 1939-1946. Finally, we linked the earlier series to the modern quarterly NIPA series from 1947 to the present.

Potential GDP:

Real potential GDP was constructed by splicing the February 2013 CBO estimates of real potential GDP from 1949 to the present with an estimated cubic trend through real GDP from 1889-1950, excluding 1930 - 1946 from the estimation. To derive nominal potential GDP, we multiplied real potential GDP by the actual price level. To derive the output gap for the Taylor rule, we used the difference between log actual real GDP and log potential.

Government Spending:

1947 - 2011: Quarterly data on nominal "Government Consumption Expenditures and Gross Investment," BEA Table 1.1.5, line 21, December 20, 2012 version.

1889 - 1946: NIPA annual nominal data from 1929 - 1946 (BEA Table 1.1.5, line 21) is spliced to annual data from 1889-1928, Source: Kendrick (1961) Table A-II.

1939 - 1946: Quarterly data on nominal government spending National Income, 1954 Edition, A Supplement to the Survey of Current Business is used to interpolate the modern annual NIPA values.

1889 - 1938: Monthly data on federal budget expenditures. Source: NBER MacroHistory Database.
http://www.nber.org/databases/macrohistory/contents/chapter15.html

m15005a U.S. Federal Budget Expenditures, Total 01/1879-09/1915
m15005b U.S. Federal Budget Expenditures, Total 11/1914-06/1933
m15005c U.S. Federal Budget Expenditures, Total 01/1932-12/1938

Data adjustment: The monthly series are spliced together (using a 12-month average at the overlap year) and seasonally adjusted in Eviews using X-12. This series includes not just government expenditures but also transfer payments, and so the monthly interpolator series is distorted by large transfer payments in different quarters. Thus, rather than using the series directly, we use it as a monthly interpolator for the annual series which excludes transfers. Following Gordon and Krenn (2010), to find these quarters, we calculated the monthly log change in the interpolator, and whenever a monthly change of +40 percent or more was followed by a monthly change of approximately the same amount with a negative sign (and also symmetrically negative followed by positive), we replaced that particular observation by the average of the preceding and succeeding months. These instances occurred for the following months: 1904:5, 1922:11, 1931:2, 1931:12, 1932:7, 1934:01, 1936:06, and 1937:06. In addition, the first quarter of 1917 was adjusted. The jump in spending was so dramatic in 1917q2 that the interpolated series showed a decline in spending in 1917q1 even though the underlying expenditure series showed an increase of 16 percent in that quarter relative to the previous one. Thus, we replaced the value of 1917q1 with a value 16 percent higher than the previous quarter. Note that our use of the proportional Denton method creates a bumpier series than an alternative that uses the additive Denton method. However, the additive Denton method leads to series that behave very strangely around large buildups and builddowns of government spending, so we did not use it. On the other hand, the alternative series gave very similar results for the multiplier.

Population:

1890-2011: Annual population data, based on July of each year, were taken from Historical Statistics of the United States Millennial Edition Online, Carter et al. (2006) We used total population, including armed forces overseas for all periods where available (during WWI and 1930 and after); otherwise we used the resident population. For 1952 through the present we used the monthly series available on the Federal Reserve Bank of St. Louis FRED database, "POP."

Data adjustment: For 1890 through 1951, we linearly interpolated the annual data to obtain monthly series so that the annual value was assigned to July. We then took the averages of monthly values to obtain quarterly series. We did the same to convert the monthly FRED data from 1952 to the present.

Tax Revenues:

1947-2012: Quarterly data on nominal "Federal Government Current receipts," BEA Table 3.2, line 1, December 20, 2012 version.

1879-1946: Monthly data on federal budget receipts. Source: NBER Macro-History Database
http://www.nber.org/databases/macrohistory/contents/chapter15.html

> m15004a U.S. Federal Budget Receipts, Total 01/1879-06/1933 m15004b U.S. Federal Budget Receipts, Total 07/1930-06/1940 m15004c U.S. Federal Budget Receipts, Total 07/1939-12/1962

Annual data on federal receipts. Source: Historical Statistics - fiscal year basis (e.g. fiscal year 1890 starts July 1, 1889)

Data adjustment: The monthly series are strung together (with the most recent series used for overlap periods) and seasonally adjusted in Eviews using X-12. The annual series is interpolated using the monthly data with the Denton proportional method.

Unemployment rate:

1948-2011: Monthly civilian unemployment rate. Source: Federal Reserve Bank of St. Louis FRED database, UNRATE
http://research.stlouisfed.org/fred2/series/UNRATE

Data adjustment: Quarterly series is constructed as the average of the three months.

1890-1947: Annual civilian unemployment rate. Source: Weir (1992). We adjusted the Weir series from 1933-1943 to include emergency workers from Conference Board (1945).

1890-1929: NBER-based monthly recession indicators. Source: Federal Reserve Bank of St. Louis FRED database, USREC
http://research.stlouisfed.org/fred2/series/USREC.

1930-1946: Monthly civilian unemployment rate (including emergency workers). Source: NBER MacroHistory Database
http://www.nber.org/databases/macrohistory/contents/chapter08.html

m08292a U.S. Unemployment Rate, Seasonally Adjusted 04/1929-06/1942 m08292a U.S. Unemployment Rate, Seasonally Adjusted 01/1940, 03/1940-12/1946

1947: Monthly civilian unemployment rate (including emergency workers, seasonally adjusted) Source: Geoffrey Moore, Business Cycle Indicators, Volume II, NBER p. 122

Data adjustment: Monthly NBER recession data are used to interpolate annual data using the Denton interpolation from 1890-1929. For 1930-1947 onwards we use the monthly unemployment rate series to interpolate annual data using the Denton proportional interpolation.

Interest rate:

*1934-2011:*Monthly 3 month Treasury bill. Source: Federal Reserve Bank of St. Louis FRED database, TB3MS
http://research.stlouisfed.org/fred2/series/TB3MS.

1920-1933: Monthly 3 month Treasury bill. Source: NBER MacroHistory Database
http://www.nber.org/databases/macrohistory/contents/chapter13.html

> m13029a U.S. Yields On Short-Term United States Securties, Three-Six Month Treasury Notes and Certificates, Three Month Treasury 01/1920-03/1934 m13029b U.S. Yields On Short-Term United States Securities, Three-Six Month Treasury Notes and Certificates, Three Month Treasury 01/1931-11/1969

Data adjustment: Quarterly series is constructed as the average of the three months.

Table 1. Tests of Instrument Relevance Across States of Slack

	F-statistic	Number of observations
1891:1 - 2011:4 - All	9.98	484
1891:1 - 2011:4 - Slack	7.38	172
1891:1 - 2011:4 - No slack	7.46	312
1948:1 - 2011:4 - All	19.01	256
1948:1 - 2011:4 - Slack	0.97	74
1948:1 - 2011:4 - No slack	15.73	182

Note: "Slack" is when the unemployment rate exceeds 6.5 percent. The F-tests are the joint significance of news variables in a regression of log real per capita government spending on its own four lags, four lags of log real per capita GDP, four lags of log real per capita federal receipts, current and four lags of news (scaled by lagged GDP), and a deterministic time trend (quartic in the full sample, quadratic in the post-WWII sample).

Table 2. Estimated Multipliers Across States of Slack

	Linear Model	High Unemployment	Low Unemployment	P-value for difference in multipliers across states
U.S.				
Peak	0.92	0.82	1.15	
	(0.462)	(0.351)	(0.696)	0.645
2 year integral	0.78	0.79	0.87	
	(0.118)	(0.131)	(0.184)	0.758
4 year integral	0.87	0.80	1.11	
	(0.109)	(0.095)	(0.181)	0.209

Note: The values in brackets under the multipliers give the standard errors.

Table 3. Robustness Checks of Multipliers Across States of Slack

	Linear Model	High Unemployment	Low Unemployment
Using linearly interpolated data			
Peak	0.86	0.77	1.04
2 year integral	0.70	0.74	0.66
4 year integral	0.80	0.73	0.89
Omitting the quartic trend			
Peak	0.76	0.65	0.87
2 year integral	0.63	0.60	0.68
4 year integral	0.70	0.63	0.79
HP filtered time-varying threshold (with $\lambda = 10^6$)			
Peak	0.92	0.87	1.08
2 year integral	0.78	0.89	0.82
4 year integral	0.87	0.82	0.96
Subsample: 1948-2011			
Peak	0.94	4.14	1.10
2 year integral	0.58	-2.34	0.78
4 year integral	0.79	18.51	1.05

	Linear	Recession	Expansion
7 qtr. moving avg. output growth, $F(z)$			
Peak	0.92	0.47	0.95
2 year integral	0.78	0.41	0.88
4 year integral	0.87	0.32	0.93

Table 4. Comparison to Auerbach-Gorodnichenko (2012)

	Linear Model	Recession	Expansion
AG's Estimates			
Peak**	0.68	2.48	0.28
5 year integral	0.57	2.24	-0.33
Jordà Method Applied to AG Specification			
Peak	0.92	1.14	0.96
5 year integral	1.05	0.87	0.53
IRFs Allowing Full Feedback		1991q1	1993q1
Peak		1.25	0.55
5 year integral		0.89	0.42

**The peak estimates are different from those reported by AG because they compared the peak of the output response to the initial shock to G. We use the standard method and compare the peak of output to the peak of the government spending response.

Note: This table shows multipliers based on impulse responses calculated three different ways. The AG estimates are calculated assuming that the economy remains in the current state for at least 5 years. The Jordà estimates are based on average transitions in the data. The full feedback estimates use AG's parameters estimates, but calculate the impulse responses conditional on the history in two particular periods (1991q1 and 1993q1) and allow full feedback to the state of the economy.

Table 5. Tests of Instrument Relevance Across Monetary Policy Regimes

	F-statistic	Number of observations
1891:1 - 2011:4 - All	9.98	484
1891:1 - 2011:4 - ZLB	2.07	89
1891:1 - 2011:4 - Normal	18.22	395

Note: "ZLB" is when interest rates are near the zero lower bound or the Fed is being very accommodative of fiscal policy (1932q1 - 1951q1, 2008q4-2011q4). The F-tests are the joint significance of news variables in a regression of log real per capita government spending on its own four lags, four lags of log real per capita GDP, four lags of log real per capita federal receipts, current and four lags of news (scaled by lagged GDP), and a quartic time trend.

Table 6. Estimated Multipliers Across Monetary Policy Regimes

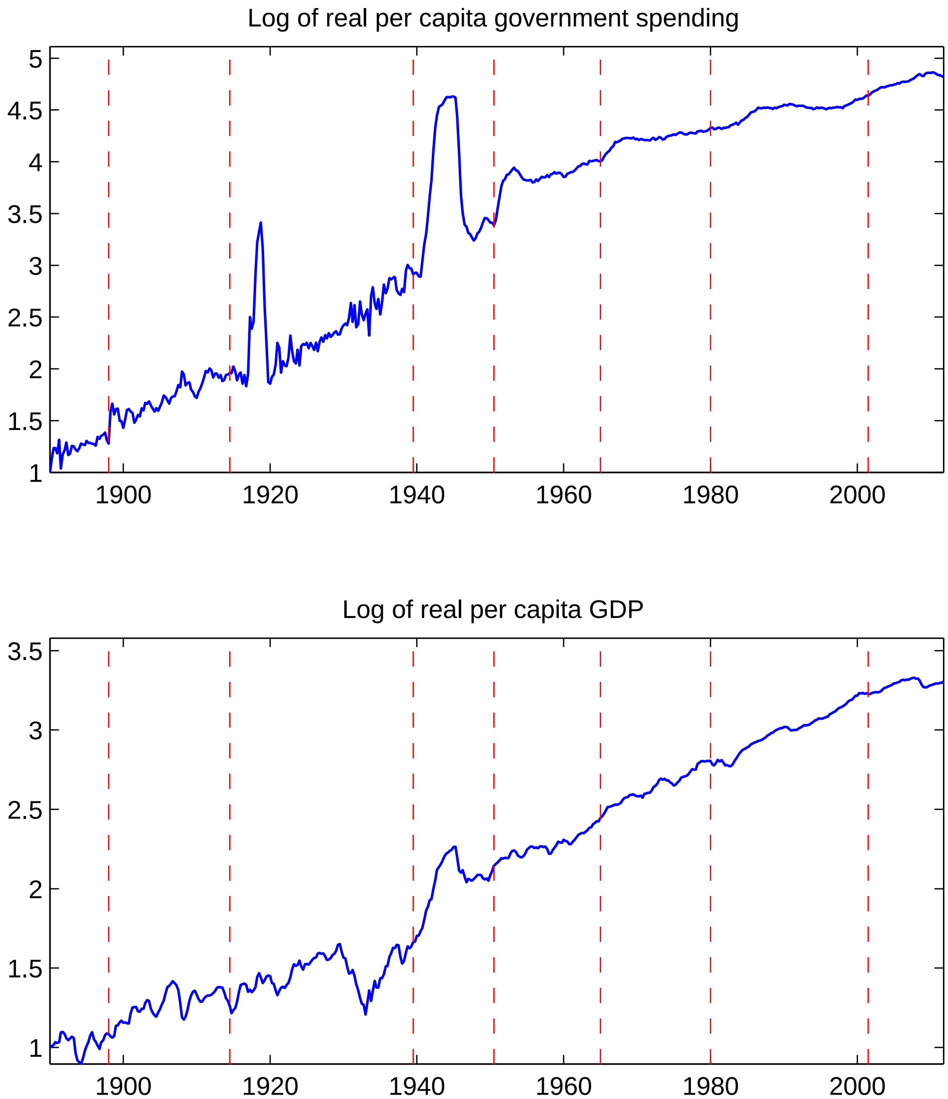
	Linear Model	Near Zero Lower Bound	Normal	P-value for difference in multipliers across states
Peak	0.92	0.71	0.80	
2 year integral	0.78	0.78	0.73	
	(0.118)	(0.172)	(0.130)	0.952
4 year integral	0.87	0.73	1.60	
	(0.109)	(0.113)	(0.304)	0.007

Note: The values in brackets under the multipliers give the standard errors.

Table 7. Robustness Checks of Multipliers Across Monetary Policy Regimes

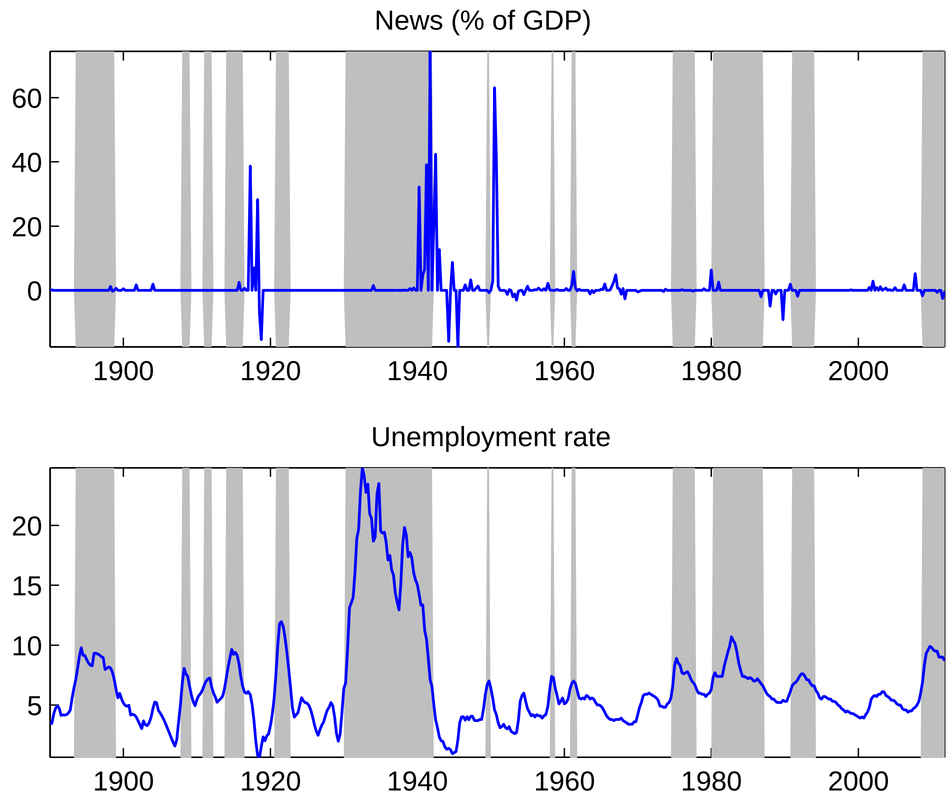
	Linear Model	Near Zero Lower Bound	Normal
Reclassification of 1942q1-1945q3 as non-ZLB			
Peak	0.92	0.67	0.92
2 year integral	0.78	0.69	0.66
4 year integral	0.87	0.70	0.97
Defining ZLB as T-bill rate$\leq$ 0.5 (1920-2011)			
Peak	0.81	0.79	1.30
2 year integral	0.72	0.70	0.91
4 year integral	0.79	0.73	1.06

Figure 1. Government Spending and GDP



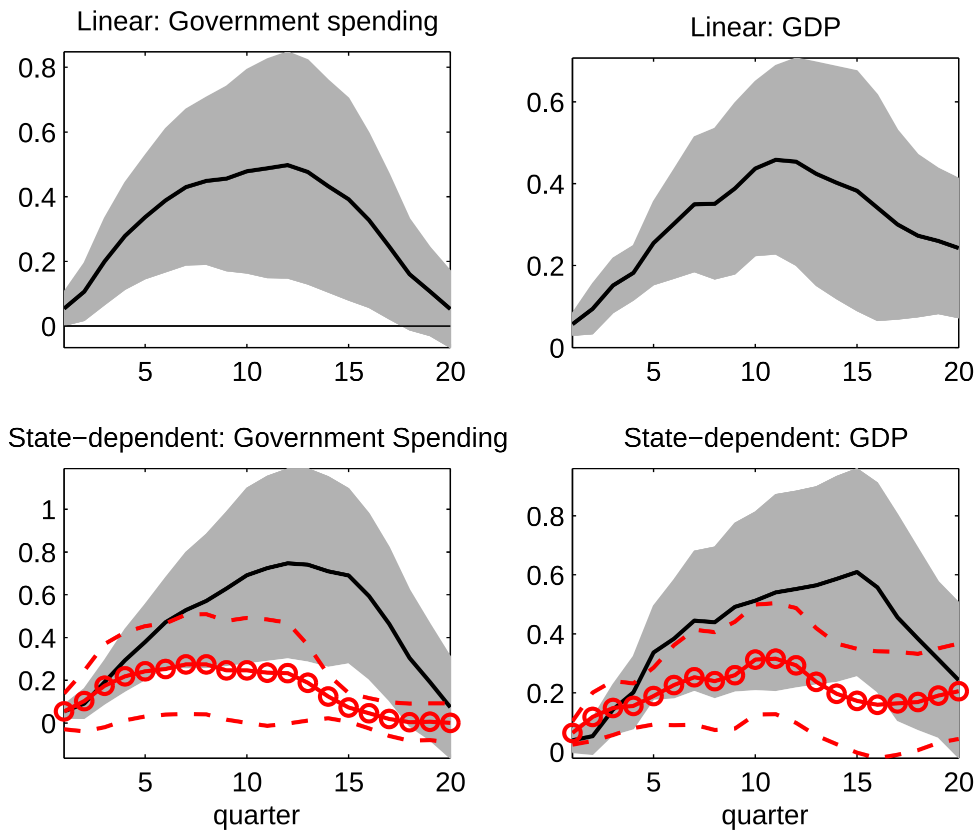
Note: The vertical lines indicate major military events: 1898q1(Spanish-American War), 1914q3 (WWI), 1939q3 (WWII), 1950q3 (Korean War), 1965q1 (Vietnam War), 1980q1 (Soviet invasion of Afghanistan), 2001q3 (9/11).

Figure 2. Military spending news and unemployment rate



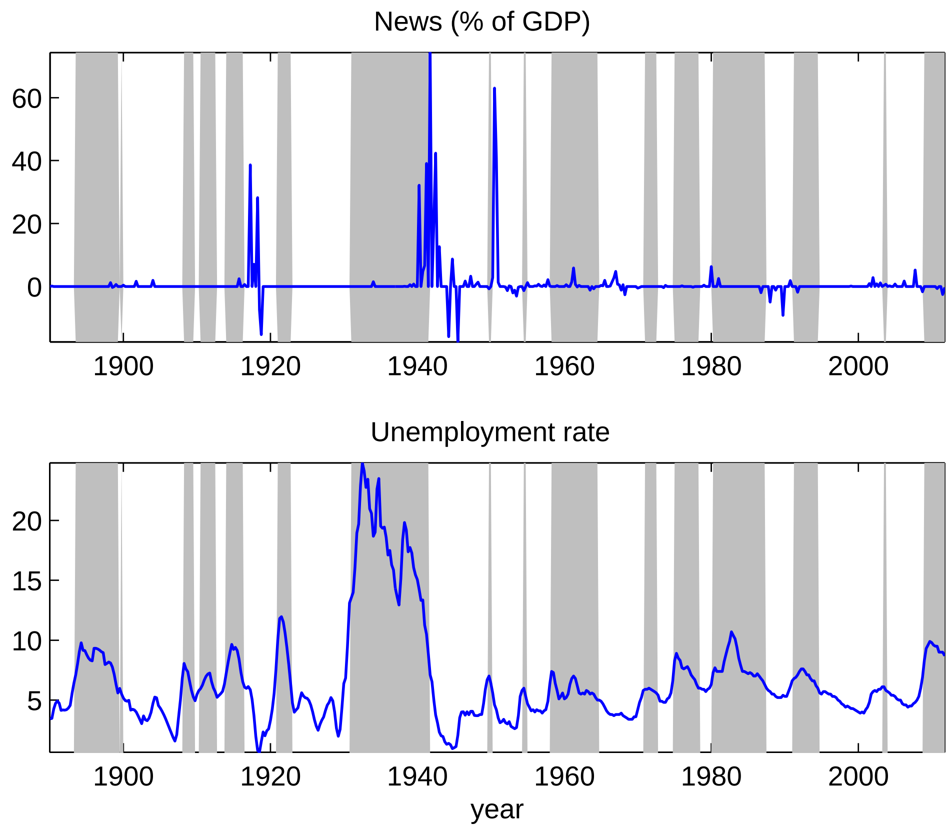
Note: Shaded areas indicate periods when the unemployment rate is above the threshold of 6.5%.

Figure 3. Government spending and GDP responses to a news shock across slack states



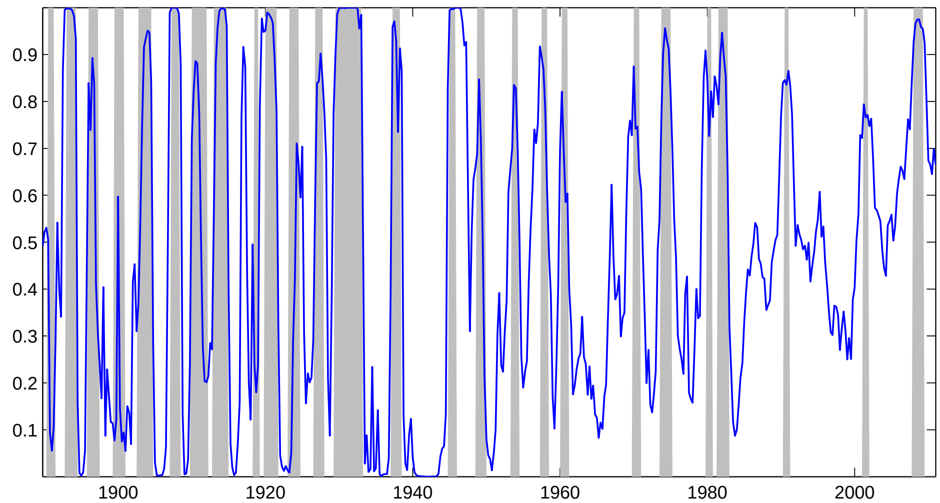
Note: Response of government spending and GDP to a news shock equal to 1% of GDP. The top panel shows the responses in the linear model. The bottom panel shows the state-dependent responses where the black solid lines are responses in the high unemployment state and the lines with red circles are responses in the low unemployment state. 95% confidence intervals are shown in all cases.

Figure 4. Robustness check: New threshold of unemployment rate based on time-varying trend



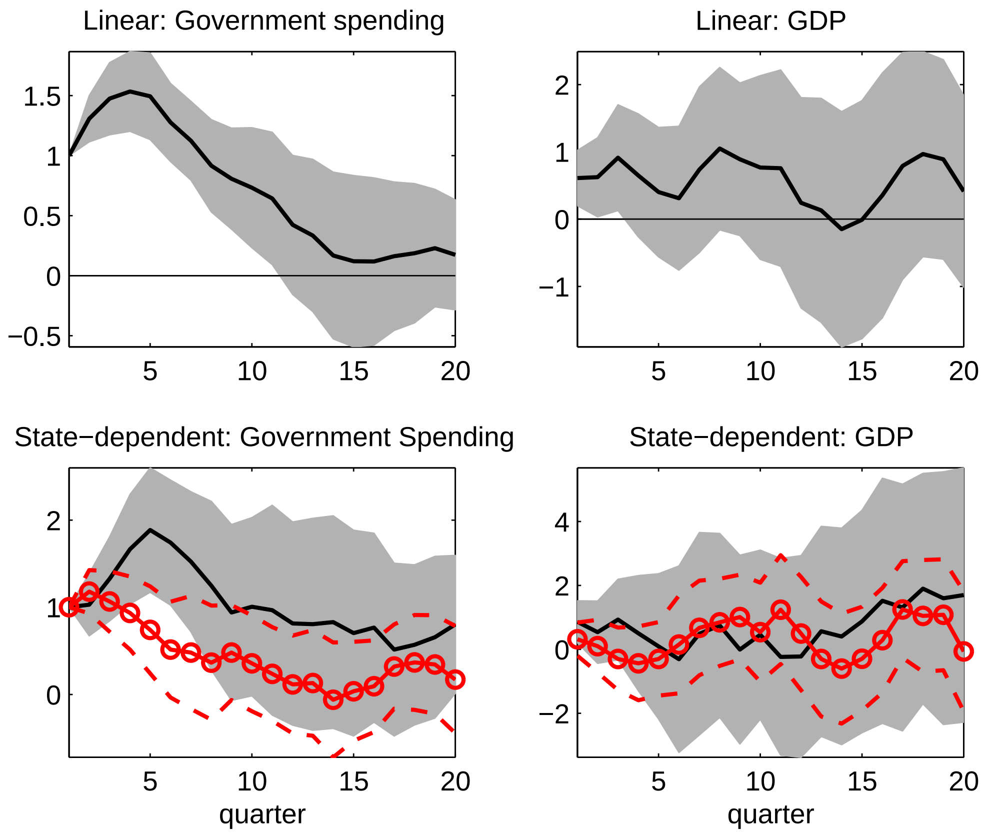
Note: Shaded areas indicate periods when the unemployment rate is above the time-varying trend based on HP filter with $\lambda = 10^6$.

Figure 5. Robustness check: New smooth transition threshold based on moving average of output growth



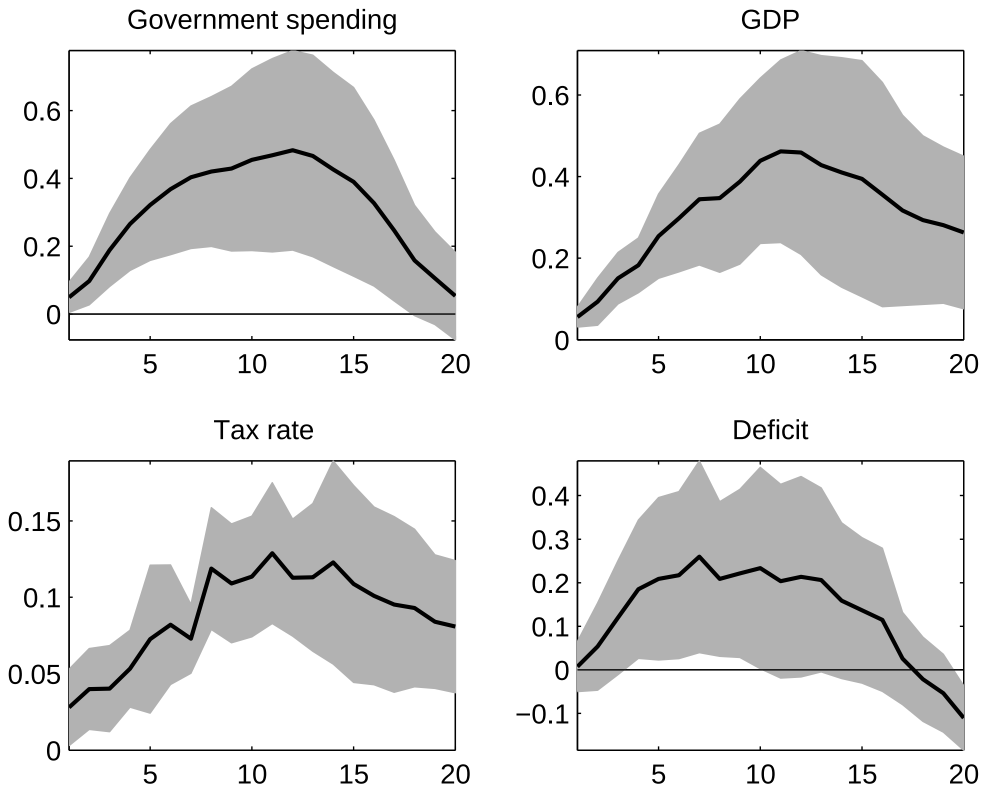
Note: The figures shows the weight on a recession regime, $F(z)$ and the shaded areas indicate recessions as defined by NBER.

Figure 6. Estimating Auerbach and Gorodnichenko (2012) with the Jorda method



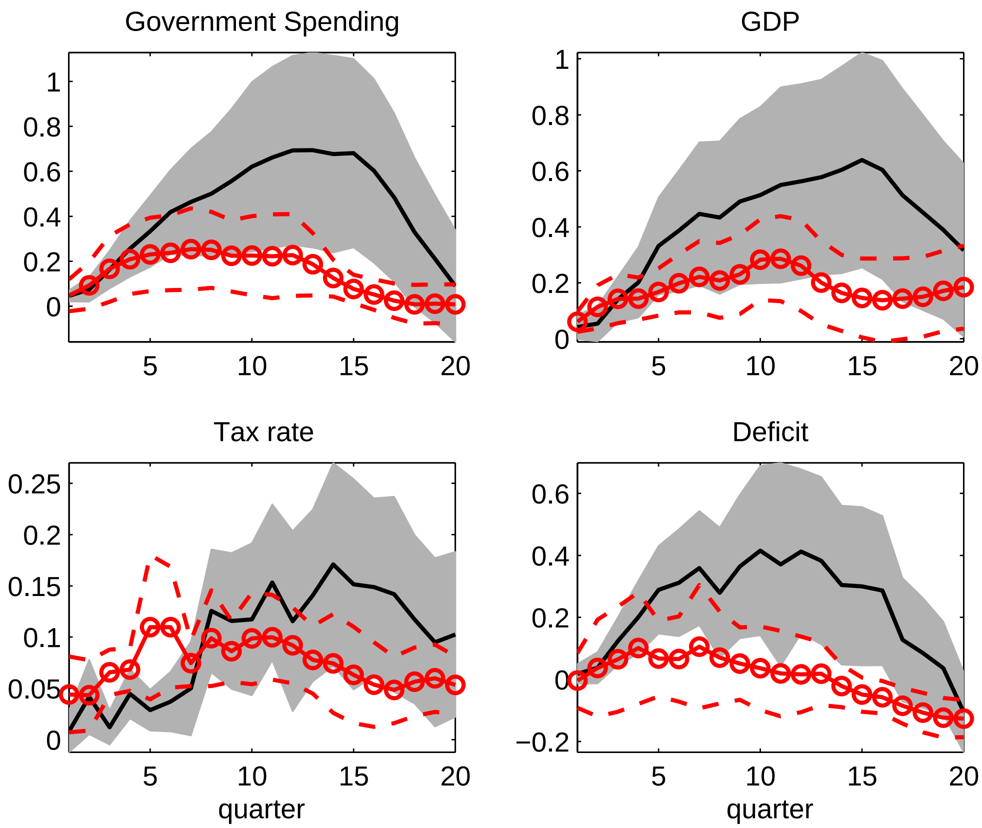
Note: Response of government spending and GDP to a government spending shock equal to 1% of GDP, with the same data, identification scheme and threshold definition as Auerbach and Gorodnichenko (2012). The top panel shows the responses in the linear model. The bottom panel shows the state-dependent responses where the black solid lines are responses in recession and the lines with red circles are responses in expansions. 95% confidence intervals are shown in all cases.

Figure 7. Responses of taxes and deficits



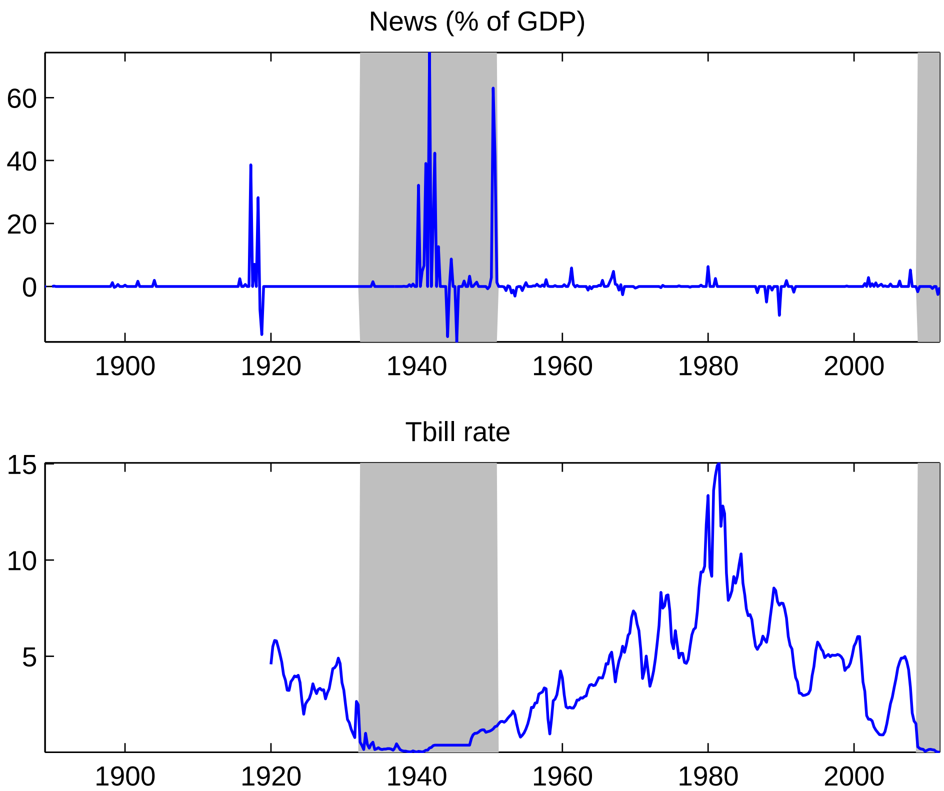
Note: These are responses for taxes and deficits in the linear model. The shaded areas indicate 95% confidence bands.

Figure 8. State-dependent responses of taxes and deficits: Considering slack state



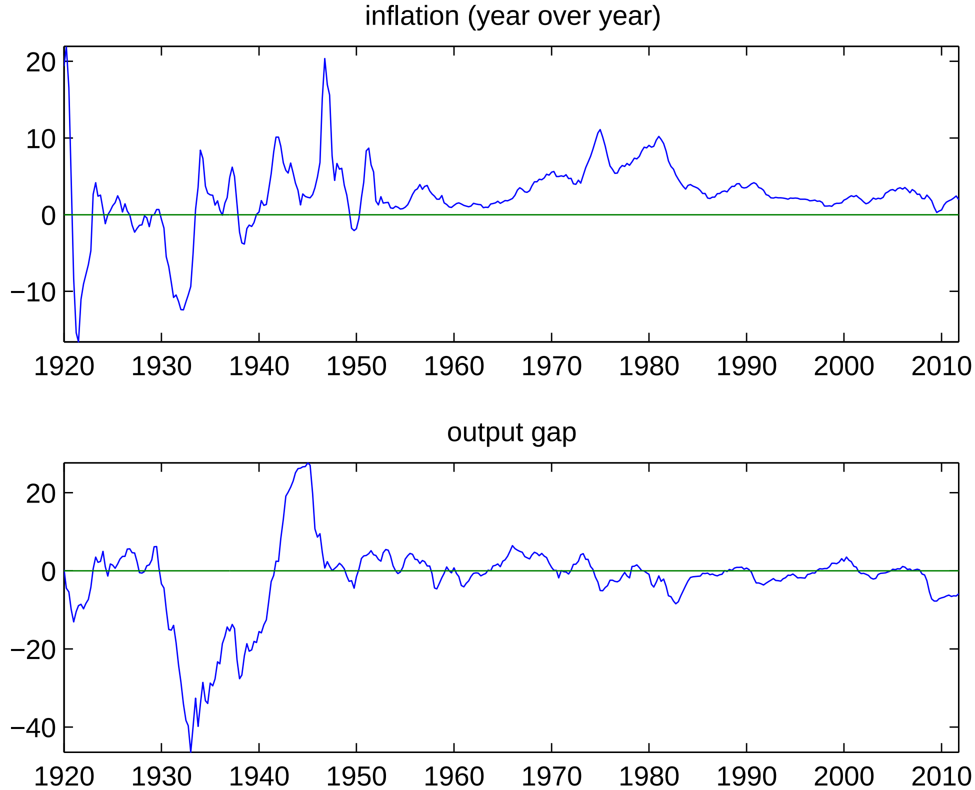
Note: These are state-dependent responses for taxes and deficits, where the black solid lines are responses in the high unemployment state and the lines with red circles are responses in the low unemployment state. 95% confidence intervals are also shown.

Figure 9. Military spending news and interest rate



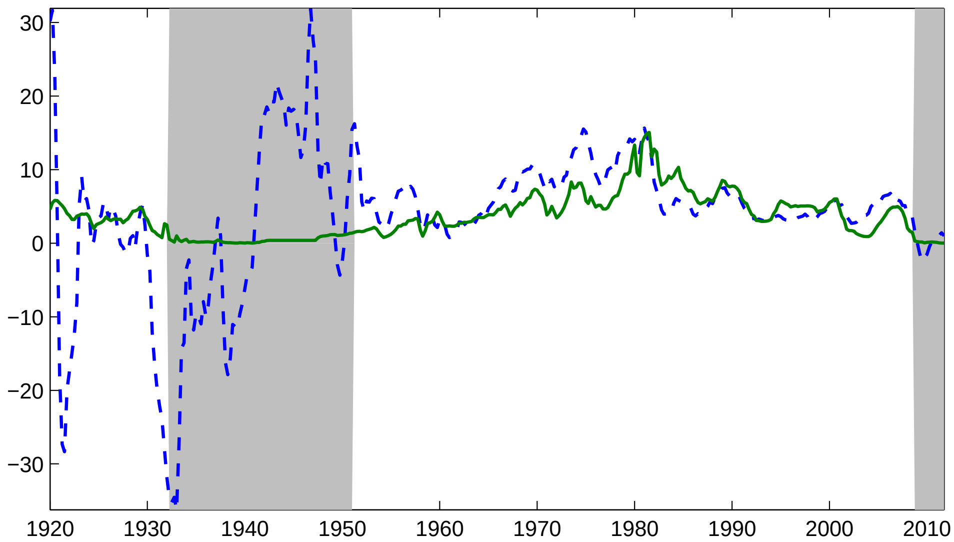
Note: Shaded areas indicate periods which we classify as the zero lower bound period for interest rate.

Figure 10. Inflation and output gap



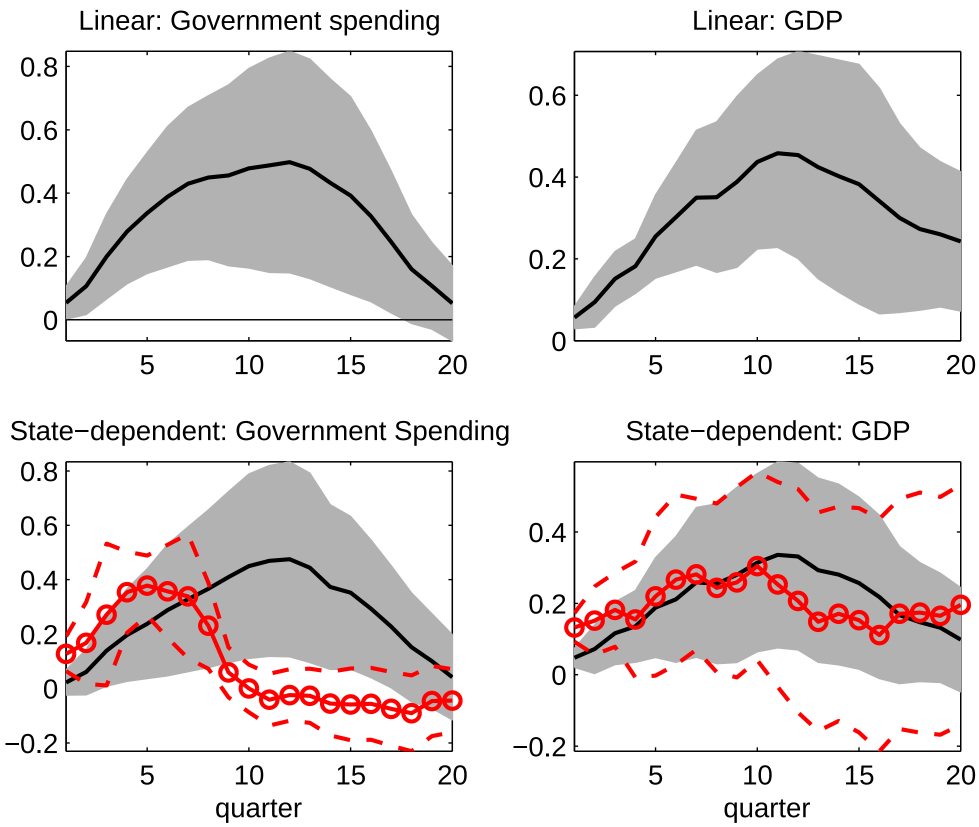
Note: The top panel shows the year-over-year GDP deflator inflation rate and the bottom panel shows the output gap, which is constructed as the percentage deviation between real GDP and potential GDP.

Figure 11. Taylor rule implied interest rate and the T-bill rate



Note: The solid line shows the data for the 3-month T-bill rate, and the dashed line shows the Taylor-rule implied nominal interest rate. Shaded areas indicate periods which we classify as the zero lower bound period for interest rate.

Figure 12. Government spending and GDP responses to a news shock: Considering zero lower bound



Note: Response of government spending and GDP to a news shock equal to 1% of GDP. The top panel shows the responses in the linear model. The bottom panel shows the state-dependent responses where the black solid lines are responses in the near zero-lower bound state and the lines with red circles are responses in the normal state. 95% confidence intervals are shown in all cases.