

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

First draft: October 2013

This draft: 20 June 2014

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 differ by the amount of slack in the economy. These results are robust to many alternative specifications. Most specifications also suggest that multipliers are not significantly higher when interest rates are near the zero lower bound. 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, Graham Elliott, Yuriy Gorodnichenko, Jim Hamilton, Michael Owyang, Garey Ramey, Tatevik Sekhposyan, Harald Uhlig, seminar participants at Berkeley, Bocconi, EUI, EIEF, Rice, University of Arizona and participants at the Hallfest conference for very helpful suggestions. We thank Kate Vermann and Michelle Ramey for excellent research assistance.

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 et al. (2013), we exploit the fact

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

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. In Ramey and Zubairy (2014), 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. Most estimates of the multiplier are between 0.6 and 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 most of the differences in conclusions between our work and that of Auerbach and Gorodnichenko (2012) lie 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.

We also find little evidence that the multiplier is higher at the zero lower bound. The only case in which the multiplier noticeably exceeds unity in the zero lower bound state is when we exclude the rationing periods of WWII. However, the results are not as robust and the multiplier falls even in this sample when we allow for a slightly more general specification.

The paper proceeds as follows. We begin by discussing the data construction in Section 2. In Section 3 we introduce the econometric methodology. 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 2013 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 follow Gordon and Krenn (2010) by using various higher frequency series to interpolate existing annual series.³ In most cases, we use the proportional Denton procedure which results in series that average up to the annual series.

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.⁴

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 starting in 1889 to annual NIPA data starting in 1929. Following Gordon and Krenn (2010), we use monthly federal outlay series from the NBER Macrobistory database to interpolate annual government spending from 1889 to 1938. We use the 1954 quarterly NIPA data from 1939-1946

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.

4. We also checked the robustness of our results by using alternative series constructed by Christina Romer. See Romer (1999) for a discussion of her data.

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 include government transfers and are on a cash (rather than accrual) basis. Fortunately, the measurement errors are less of an issue for us because we identify the shocks using narrative methods.⁵

The unemployment series is constructed by interpolating Weir's (1992) annual unemployment series, adjusted for emergency worker employment.⁶ Before 1948 we use the monthly unemployment series available from the NBER Macroeconomy 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 are 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.

5. 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.

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

3 Econometric Methodology

We use Jordà's (2005) local projection method to estimate impulse responses and multipliers. Auerbach and Gorodnichenko (2013) were the first to use this technique to estimate state-dependent fiscal models, employing it in their analysis of OECD panel data.⁷ The Jordà method simply requires estimation of a series of regressions for each horizon h for each variable. The linear model looks as follows:

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

x is the variable of interest, 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 and government spending. In addition, y includes lags of the news variable to control for any serial correlation in the news variable. $\psi(L)$ is a polynomial of order 2.⁸ 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 projection 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:

$$(2) \quad x_{t+h} = I_{t-1} [\alpha_{A,h} + \psi_{A,h}(L)y_{t-1} + \beta_{A,h}shock_t] \\ + (1 - I_{t-1}) [\alpha_{B,h} + \psi_{B,h}(L)y_{t-1} + \beta_{B,h}shock_t] + \text{quartic trend} + \varepsilon_{t+h}.$$

7. Stock and Watson (2007) also explore the properties of this method for forecasting.

8. Both AIC and BIC favor 2 lags for the government spending and GDP equations. AIC also favors 2 lags for the GDP equation, but favors 4 lags for government spending. We decided to go with 2 lags for both since the results are similar to those with 4 lags.

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 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 misspecification 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, as the horizon increases, one loses observations from the end of the sample. Third, the impulse responses sometimes display 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.

Finally, in order to avoid the bias of converting elasticities using sample averages, highlighted by Ramey and Zubairy (2014), 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:

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

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

$$\frac{G_{t+h} - G_{t-1}}{Y_{t-1}} \approx (\ln G_{t+h} - \ln G_{t-1}) \cdot \frac{G_{t-1}}{Y_{t-1}}$$

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.⁹ 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.¹⁰ The results were very similar. A second concern is that cyclicity 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. This specification follows the method of Gordon and Krenn (2010), both in the way that potential GDP is constructed and in its use in the denominator. 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

9. Based on private correspondence with Danny Shoag, Alan Auerbach and Yuriy Gorodnichenko.

10. See the data appendix for details on how this was constructed.

results. The later sections conduct robustness checks and analyzes in detail why our results are different from Auerbach and Gorodnichenko (2012).

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.¹¹ 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. Michailat (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, Michailat (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 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 the U.S. Federal Reserve's choice of that value as a threshold in its recent policy announcements. 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

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

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.

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 relevance of the news variables.¹² 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 above 10. The exception is the slack state in the post-WWII period, where the F-statistic is barely above 2. 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.¹³

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

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

13. 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 two horizons: the ratio of the cumulative responses of GDP and government spending through two years, and the cumulative responses through four years.¹⁴ The integral multipliers address the relevant policy question 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 0.7 to 0.9. The estimates are not statistically different from one at the five percent significance level.¹⁵

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 2 and 4 year multipliers are slightly lower during the high unemployment state. In no case do we find a statistically different multiplier across states.

One might worry that the multipliers are different at other horizons. Figure 4 shows the cumulative multiplier for each horizon from impact to 5 years out. The top graph shows the linear model multipliers and the bottom graph shows the state dependent multipliers. In the linear case, the cumulative multiplier in the first year is above one but then falls. The reason for the higher initial multipliers after a news shock is given by Ramey (2011):

14. To further clarify, 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.

15. The standard errors are computed by estimating all of the regressions jointly as one panel regression and using the Driscoll and Kraay (1998) standard errors to account for correlations in the error terms. We implement this estimation in Stata, using *xtscc* followed by *nlcom*, which uses the delta method to analyze functions of parameters.

output responds immediately to news about future government spending increases. Since output rises more quickly than government spending, the calculated multiplier looks large. When we instead use Blanchard-Perotti SVAR shocks (which do not account for news), we find that multipliers are close to zero in the first year. The bottom graph shows that whatever the values, the multipliers in the high unemployment state are below or equal to those in the low unemployment state.

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

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.

We chose our threshold unemployment rate as 6.5 percent in the baseline specification, but it might be a concern that we are not allowing for the possibility of state-dependence that might arise only for a higher degree of slack in the economy. The third panel in Table 3 shows that when we choose the threshold for the unemployment rate to be higher than 8 percent, our main results are still preserved where there is no evidence of difference in multipliers across states. Notably, the 2 year integral multiplier is 0.78 in high unemployment state and 0.77 in low unemployment state. In addition, for the four year integral multiplier

the high unemployment multiplier with the more severe amount of slack in the economy is 0.76 and slightly higher at 0.96 for the low unemployment state.

We also allow for 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 5, 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 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 6 shows the function $F(z)$ along with the NBER recessions for our full sample. Results in the bottom panel 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, 1947-2013, 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 -5 and the four year integral multiplier is large

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.

and positive taking a value of 25! 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.

Most of the literature that finds evidence of state-dependence in spending multipliers also identifies government spending shocks differently. The commonly used approach is to employ Blanchard and Perotti (2002) identification scheme which is based on the assumption that within quarter government spending does not contemporaneously respond to macroeconomic variables.¹⁸ Table 3 shows that when we identify the government spending shock using Blanchard and Perotti (2002) identification scheme for our full sample, we still find multipliers less than one in the linear and state-dependent case. Also, the multipliers are not statistically different across the high and low unemployment states.

Another potential concern is the role of rationing in World War II. If consumers are constrained in what they can buy, then the traditional Keynesian multiplier might not work to full effect. While most of the World War II period is characterized as a non-slack state by our definition, some of the initial news shocks hit during the slack state and rationing could potentially be depressing the multiplier during the subsequent periods. Official rationing did not start until 1942, but other restrictions were imposed on the U.S. economy starting in the second half of 1941. For example, the Office of Production Management asked for a cutback in automobile production. Gordon and Krenn (2010) carefully document various other capacity constraints that occurred starting in the second half of 1941. In order to determine whether our results are sensitive to the rationing period, we exclude observations on the dependent variables in our regression for the period 1941q3 through 1945q4.

When we exclude World War II from the sample, the multipliers are indeed a little higher overall. In the linear case, they are 1 and 1.22 for the two and four year integral multipliers, respectively. As the second to last panel in Table 3 shows, the multipliers are larger in the slack and non-slack state as well, relative to our baseline. The two year integral multiplier is 0.77 in high unemployment and 0.99 in low unemployment. Similarly, for the four year integral, the multiplier is again slightly higher in the non-slack state versus

18. However, this type of shock is much more sensitive to potential measurement errors given our historical construction of quarterly government spending and GDP series and is subject to the critique that it is likely to have been anticipated.

high unemployment period. Therefore, our baseline results are robust to the exclusion of World War II from the sample, where we find no significant difference between slack and non-slack states.

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 (2012, 2013). In this section, we explain the main source of 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 centered moving average of normalized real GDP growth as their measure of the state. They also include four lags of the 7-quarter moving average growth rate as exogenous regressors in their model.¹⁹ They construct their baseline impulse responses based on two assumptions: (1) the economy remains in an extreme recession or expansion state for at least 5 years; and (2) changes in government spending do not impact the state of the economy.²⁰ They find multipliers during recessions that are well above two and these results have been cited by those advocating stimulus during the Great Recession (e.g. DeLong and Summers (2012))

To understand the difference between our results and theirs, we begin by taking 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, logs of variables, estimated government spending shocks from their STVAR model, and inclusion of four lagged values of the centered 7-quarter moving average of output growth as controls. $F(z)$ is the indicator of the state as a function of the moving average of output (z). It varies between a maximum of one (extreme recession) and zero (extreme expansion).

Figure 7 shows the linear responses in the top panel.²¹ 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. The state-dependent responses shown in the

19. 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.

20. These results are shown in Table 1 and Figure 2 of their paper.

21. We have multiplied the log output response by 5.6 following AG-12. They use a nonstandard measure of government purchases as their measure of G . As a result their Y/G ratio used to convert multipliers is higher than the usual one based on total government purchases.

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 contrast to AG-12 who find that output rises robustly and continuously throughout the 5 years in the recession state but quickly falls toward zero and becomes negative in the expansion state.

The first panel of Table 4 compares AG-12's cumulative 5-year multipliers to those we estimated by the Jordà method.²² AG find multipliers of 2.24 for recessions and -0.33 for expansions whereas the Jordà method estimates multipliers of 0.84 in recessions and -0.59 in expansions. Thus, the Jordà multipliers are much lower in recessions and the gap between states is not nearly so large. One should keep in mind, however, that the Jordà estimates are not very precise.²³ This lack of precision likely reflects the fact that the post-WWII data does not have as much variation of government spending across states. We also report the two year multiplier for comparability with our earlier estimates. At this horizon, the multipliers in both states are similar and are very low.

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 identified shocks. Rather, it is likely due to the method for estimating the impulse responses. 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.²⁴

22. Following Blanchard and Perotti (2002), AG-12 also report "peak multipliers," calculated as the ratio of the peak of the output response to the *initial* shock to government spending. Since this ratio does not take into account the response of government spending to its own shock, it is not really a multiplier.

23. We were not able to estimate the standard errors of the multipliers because the `xtscc` command in Stata reported that the variance matrix was nonsymmetric or highly singular.

24. For instance, Caggiano et al. (2014) employ the STVAR approach of AG-12 for a shorter sample, but compute impulse response functions using the generalized impulse response approach advocated by Koop et al. (1996), and find that the spending multipliers in recessions are not statistically larger than in expansions. They only find evidence of nonlinearities when focusing on deep recessions versus strong expansionary periods.

In the Jordà 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. The parameter estimates on the control variables incorporate the average tendency of the economy to evolve between states. Thus, if the duration of State A is typically short relative to State B, the h quarter ahead forecast will take this into account. Similarly, the estimate of the coefficient on the shock includes the effects of the shock on the future state of the economy. Thus, the estimates incorporate both the natural transitions and endogenous transitions from state to state that occur on average in the data.

In contrast, 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. As Figure 6 shows, during the post-WWII period the recession states are much shorter than 20 quarters in duration; the mean duration is only 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. Moreover, the assumptions imply that a positive shock to government spending during a recession does not help the economy escape the recession. AG-12 relax this second assumption in a second experiment, and we will describe those results below.

To see the importance of these assumptions, we conduct several experiments. In these experiments, we compute alternative impulse responses by iterating on AG's STVAR parameter estimates under different assumptions about the dynamics of the state of the economy. Since the economy is never literally in an extreme recession or expansion, we focus on the average of "severe" recessions and "severe" expansions, which we define as the few quarters in which $F(z)$ is above 0.95 or below 0.05, respectively. The few quarters of severe recession occur during the 1974-75, 1981-82, and 2008-09 recessions.²⁵

The second panel of Table 4 reports these experiments. For reference, the first line of the table shows that AG's baseline 5-year multipliers do not change much when we change $F(z)$ by a small amount. The second line shows the multiplier calculated assuming a constant

25. The mean value of AG's $F(z)$ indicator is 0.81 during NBER recessions and 0.42 during NBER expansions.

state and no feedback, but looking at the 2-year integral multiplier. This calculation requires less drastic assumptions because it only assumes that the state remains constant for 2 years rather than 5 years. Here, the multipliers in severe recessions are not as high and those in severe expansions are not as low, so the difference across states falls from 2.47 to 1.46.

The next experiment, "Actual State Dynamics," assumes that instead of staying constant at an extreme value, the state indicator $F(z)$ is equal to its actual value at each point in time. In practice the experiment is conducted as follows. We first calculate the paths of government spending and output, assuming that the shocks to the government spending, tax, and output equations take their estimated values. This essentially reproduces the actual path of the economy for all variables including $F(z)$. We then calculate an alternative path of government spending, taxes, and output assuming that there is an additional one-time positive shock in the current period to government spending, equal to one standard deviation of the estimated government spending shock (equal to 1.3 percent of government spending). We allow the shock to change the path of spending, taxes, and output, but not the state of the economy, $F(z)$, relative to its actual path. The difference between this simulation and the actual values of the variables forms the impulse response functions. Despite the lack of feedback, this experiment is different from AG's baseline experiment because it allows the state of the economy to experience its natural dynamics (i.e. $F(z)$ is allowed to vary between its extremes as it actually does). The third and fourth lines of the lower panel show the multipliers for this experiment. The five-year multiplier is estimated to be 1.4 in severe recessions and 0.2 in severe expansions. The difference in multipliers across states is 1.2, which is less than AG's baseline. The gap is even less for the two-year multipliers. Thus, allowing the state of the economy to follow its subsequent natural dynamics reduces the gap in multipliers across states.

AG-12 relax one of their baseline assumptions in a second experiment by allowing partial feedback of government spending on the state of the economy, but otherwise not allowing the state to change. They are not able to allow full feedback, though, because of the nature of their state variable.²⁶ The fifth and sixth lines of the second panel of the table show the average of their multipliers in severe recessions and expansions. Their experiment also compresses the difference across states. In severe recessions the five-year multiplier is 1.36 and the two-year multiplier is only 1.01.

26. Their state variable is a function of a *centered* moving average of GDP growth. Thus, future values of GDP enter the current state. This formulation makes it impossible to allow full feedback in a logically consistent manner.

The final two lines of the table show our experiment in which we allow both the natural dynamics of the economy and partial feedback from the government spending shock. The experiment is the same as the "Actual State Dynamics, No Feedback" except that it also allows the $F(z)$ indicator to change from its actual path in reaction to current changes of output resulting from the government spending shock. As shown in the table, the five-year multiplier during severe recessions is calculated to be 1.07 and during severe expansions is 0.14. The difference is 0.9. The results are similar for the two-year multiplier.

Thus, even when we use AG's STVAR parameter estimates, we can get very different estimates for the multiplier. Differing assumptions made about transitions between states and the feedback of government spending to the state lead to very different estimates of multipliers. When we compute multipliers allowing for the natural dynamics of the economy we find a much smaller gap across states than AG-12. The gap we do find is not because the multiplier is so high during recessions, but because it is so low during expansions.

But why do we obtain different results from those of Auerbach-Gorodnichenko's second paper? 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 fiscal 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, 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. Ramey and Zubairy (2014) shows that this practice can lead to biased multiplier estimates. Second, they follow Blanchard and Perotti (2002) 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 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 7.8 to convert the elasticities. This results in 2-year integral multipliers of 1.52 in the high unemployment state and 1.44 in the low un-

employment state. Thus, this change in method close to *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 9.33 for the high unemployment state and 3.91 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

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 tax rates so that we can distinguish increases in revenues caused by rising output versus rising rates. Average tax rates are computed as the ratio of federal receipts to nominal GDP. The deficit is the real total deficit. We include two lags of the logs of these two new variables along with GDP and government spending as controls in all of the regressions, and we estimate this system for the full sample using the Jordà method.

Figure 8 shows the results from the linear case. The responses of government spending and GDP, as well as the multipliers, are almost identical to the baseline case. The middle panels show that both average tax rates and the deficit increase in response to news shock. Taking the ratio of the cumulative response of deficit to cumulative response of government spending at various horizons, we estimate a sharp rise in the share of government spending financed with the deficit during the first year. The deficit fraction of government spending then plateaus at 60 percent .

From a theoretical perspective, the fact that tax rates increase steadily during the first two years 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:

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

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 9 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.²⁷ 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 cumulative deficit to cumulative 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.²⁸ For example, at quarter seven the ratio of the cumulative deficit to cumulative government spending is 63 percent if a shock hits during a slack state but only 45 percent if the shock hits during a non-slack state. Thus, on average short-run government spending is financed more with deficits if the shock hits during a slack state. In addition, tax rates rise with a delay during the slack state relative to non-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, some 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.²⁹

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

27. The implied multipliers are very similar to the baseline case as well.

28. This is true with the exception of the second quarter. This can be explained by the fact that initially government spending and deficits rise slowly in response to a news shock and for the initial two quarters, deficits fall very slightly before rising.

29. 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.

(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.³⁰ 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.³¹

This view has been challenged by a series of new papers. For instance, Mertens and Ravn (2014) find that if the liquidity trap is caused by a self-fulfilling state of low confidence, rather than preference shocks, as in most of the afore-mentioned studies, then higher government spending has deflationary effects that consequently reduce the spending multiplier when the zero lower bound is binding. Aruoba and Schorfheide (2013) also analyze the effects of fiscal policy under non-fundamental shocks in a medium-scale DSGE model, and reach similar conclusions about the impact on fiscal multipliers. However, based on an estimated model on data from 1984-2007, they conclude that the empirical evidence for such an extended deflationary equilibrium is very weak.

Braun et al. (2013) also provide evidence that the properties of fiscal policy in the New Keynesian model at the zero lower bound are often not very different from the properties of fiscal policy away from the zero bound. Their results are based on parameterizations of the model that reproduce declines in output and inflation observed during the U.S. Great Recession or the U.S. Great Depression episode, including the expected duration of zero

30. 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).

31. 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 of wage and price rigidities, and the persistence of the spending shock.

interest rates. Kiley (2014) shows that replacing the standard sticky price assumption with the assumption of sticky information also results in smaller multipliers in a zero-lower bound environment. In particular, he shows that the spending multiplier is strictly below one and decreasing in the duration of the stimulus under sticky information, whereas it is strictly above one and increasing in the duration of the stimulus under sticky prices.

Thus, the literature now provides a number of plausible theories that predict both higher and lower multipliers at the zero lower bound. For this reason, it is useful to provide empirical evidence on this issue.

5.2 Defining States by Monetary Policy

The bottom panel of Figure 10 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:

$$(4) \quad \text{nominal interest rate} = 1 + 1.5 \text{ year-over-year inflation rate} + 0.5 \text{ output gap}$$

Figure 11 shows the behavior of inflation and the output gap, which were quite volatile during the early period.³² Figure 12 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.

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

For our baseline, we define ZLB or extended monetary accommodation times to be 1932q2 - 1951q1 and 2008q4 - 2013q4 (the end of our sample). We do not include the early 1920s as a ZLB episode since the episode was so brief and theory tells us that the

32. The output gap for the earlier period is constructed similarly to Gordon and Krenn (2010). See the data appendix for details.

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 10 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.³³ Table 5 shows the F-tests for the periods split into ZLB periods and normal periods. The F-statistic is relatively low for the ZLB period, close to 5, so instrument relevance is a concern. One of our robustness checks excludes observations from WWII, so we also show the F-statistics for the ZLB periods excluding WWII. As the last line of the table shows, the F-statistic is below one, indicating a serious problem with instrument relevance. This should be kept in mind when we show our results for that restricted sample.

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 1889-2013.³⁴ Figure 13 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 cumulative multipliers in each state for the different horizons of two and four years, respectively. For the two-year integral, the multipliers in the ZLB state are slightly higher than normal times, and for the 4-year integral, the multiplier is higher in the normal state. However, the multiplier is less than one in both these cases and in no case do we find evidence of significantly higher multipliers during periods at the zero

33. An advantage of the Crafts and Mills (2012) analysis of UK data is that it has more defense news shocks during the 1930s.

34. 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.

lower bound or constant interest rates. Figure 14 shows the cumulative multiplier for the ZLB and normal state at various different horizons along with 95% confidence bands. The multiplier for both states is high on impact when the news shock hits the economy and is less than one after one year, but the multipliers across the two states are never significantly different.

We explore the robustness in several ways. First, we redefined the ZLB state as periods where the T-bill rate was less than or equal to 50 basis points. We have data for the 3-month T-bill rate starting in 1920, but we assume that prior to 1920 there was no monetary accommodation, and classify 1890-1919 as non-ZLB period. As shown in the first lines 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 higher multipliers in the ZLB state.

We also check the robustness of our results under Blanchard and Perotti (2002) identification, since the F-statistic for our news shock for the ZLB period is relatively low and might be a concern. The next set of results in Table 7 shows that under this alternative identification scheme, the linear multiplier is less than one for both two and four year horizons. The state dependent multipliers are also less than one: the 2 year multiplier under ZLB is 0.66 and 0.51 in normal times, whereas at four year horizon the multiplier is 0.79 under ZLB and slightly higher at 0.93 in normal times. Thus, again we do not find any evidence of significantly higher multipliers near the zero lower bound than normal times.

We consider the role of financing by controlling for taxes, by adding 2 lags of tax revenues as controls in our baseline specification. In this case, the multipliers are slightly lower in the ZLB state and higher in the normal state relative to the baseline case, for both 2 and 4 years. The difference between the multiplier in ZLB and normal states at 2 year is negligible and at 4 years, if anything, the multiplier is higher when the shock hits during a normal state than the ZLB state (see third panel of Table 7) .

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 exclude 1941q3 through 1945q4 from the data sample.³⁵ Table 7 shows that this change results in multipliers slightly higher in the linear case, somewhat above 1. Considering the state dependent multipliers,

35. We exclude this time period from the sample by ensuring that we do not consider responses of output and government spending from 1941q3 through 1945q4, when rationing might be a concern.

it is apparent that exclusion of World War II does not impact the multiplier in the normal state, but results in larger multipliers at both two and four year horizons in the ZLB state. The difference in multipliers is driven by the impulse response function of government spending across the two different states, where it rises more robustly and faster in normal times than the ZLB state, when we exclude WWII. The response of output is statistically not different across the two states. Notably, the multiplier at the two year horizon in the ZLB state is 1.59 versus 0.54 in the normal state, and the difference in the multipliers has a p-value of 0.07. The difference between the four year multipliers across the two states narrows to 1.11 in the ZLB state and 0.84 in the normal state.

We are not sure how much weight to put on these results because of the issues of instrument relevance discussed earlier.³⁶ Moreover, these results are not very robust to small generalizations. For example, using the sample that excludes WWII, if we also control for two lags of taxes, the multiplier in the ZLB state falls. These results are shown in the last panel of Table 7. In this specification, the two year horizon multiplier in the ZLB state is 1.26 and in the normal state is 0.7. By the fourth year, the multiplier in the ZLB state is only 0.99 and is lower than the multiplier in the normal state. The differences in multipliers across states for these horizons are not statistically significant at conventional levels.

Thus, we do not find robust results in support of the New Keynesian model prediction that multipliers are greater at the zero lower bound. 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, neither our results nor those of Crafts and Mills (2012) for the UK are consistent with the predictions of the New Keynesian model at the ZLB.³⁷

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

36. If we try to overcome the instrument relevance problem by identifying shocks using Blanchard-Perotti's method, we obtain implausible results. Because the response of government spending turns negative at quarter 4 in the recession state, the implied multipliers become increasingly large as the cumulative change in government spending shrinks and then becomes negative. For example, the multiplier is calculated to be 56 at quarter 12, but -108 at quarter 13.

37. 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.

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 in most specifications multipliers are not higher at the zero lower bound. The only exception is when we exclude World War II from the sample; however, these results are questionable because of issues of instrument relevance and a lack of robustness to additional controls. Thus, we did not find evidence to support 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 infrastructure spending. 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

- Aruoba, S. Boragan and Frank Schorfheide**, 2013. “Macroeconomic Dynamics Near the ZLB: A Tale of Two Equilibria.” NBER Working Paper 19248.
- 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.
- Braun, R. Anton, Lena Maren Korber, and Yuichiro Waki**, 2013. “Small and orthodox fiscal multipliers at the zero lower bound.” Working Paper 2013-13, Federal Reserve Bank of Atlanta.
- Bruckner, Marcus and Anita Tuladhar**, 2013. “Local Government Spending Multipliers and Financial Distress: Evidence from Japanese Prefectures.” *Economic Journal*, forthcoming.
- Caggiano, Giovanni, Efrem Castelnuovo, Valentina Colombo, and Gabriela Nodari**, 2014. “Estimating Fiscal Multipliers: News from a Nonlinear World.” Working Paper 179, Universita Degli Studi di Padova.

- 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. “What determines government spending multipliers?” *Economic Policy* 27(72): 521–565.
- 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.
- DeLong, J. Bradford and Lawrence H. Summers**, 2012. “Fiscal Policy in a Depressed Economy.” *Brookings Papers on Economic Activity* 44(1): 233–297.
- Drautzburg, Thorsten and Harald Uhlig**, 2011. “Fiscal Stimulus and Distortionary Taxation.” NBER Working Paper 17111, National Bureau of Economic Research.
- Driscoll, John C. and Aart C. Kraay**, 1998. “Consistent Covariance Matrix Estimation With Spatially Dependent Panel Data.” *The Review of Economics and Statistics* 80(4): 549–560.
- 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 Paper 18058.
- 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.” NBER Working paper 16380.
- 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.
- Kiley, Michael T.**, 2014. “Policy Paradoxes in the New Keynesian Model.” Finance and Economics Discussion Series 2014-29, Board of Governors of the Federal Reserve System.
- 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(4): 1212–1247.
- Mertens, Karel and Morten O. Ravn**, 2014. “Fiscal Policy in an Expectations Driven Liquidity Trap.” *Review of Economic Studies* forthcoming.
- Michaillat, Pascal**, 2014. “A Theory of Countercyclical Government Multiplier.” *American Economic Journal: Macroeconomics* 6(1): 190–217.
- Mittnik, Stefan and Willi Semmler**, 2012. “Regime Dependence of the Fiscal Multiplier.” *Journal of Economic Behavior 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**, 2013. “Are Government Spending Multipliers Greater During Times of Slack? Evidence from 20th Century Historical Data.” *American Economic Review* 103(2): 129–34.
- 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. and Sarah Zubairy**, 2014. “Biases in Computing Government Spending Multipliers.” Work-in-progress.
- Romer, Christina D.**, 1999. “Changes in Business Cycles: Evidence and Explanations.” *Journal of Economic Perspectives* 13(2): 23–44.
- 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 Estimates 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*, 55(1).

Data Appendix

GDP and GDP deflator:

1947 - 2013: Quarterly data on chain-weighted real GDP, nominal GDP, and GDP deflator from BEA NIPA (downloaded from FRED, March 27, 2014 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 2014 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. Our method of constructing real potential GDP is similar to the method advocated by Gordon and Krenn (2010). They illustrate the problems that arise

when one uses standard filters to estimate trends during samples that involve the Great Depression and World War II, and advocate instead using a piecewise exponential trend based on benchmark years. Our procedure is a smoothed version of theirs. 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 - 2013: Quarterly data on nominal "Government Consumption Expenditures and Gross Investment," from BEA NIPA (downloaded from FRED, March 27, 2014 revision).

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 from *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/macrophistory/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 build-downs of government spending, so we did not use it. On the other hand, the alternative series gave very similar results for the multiplier.

Population:

1890-2013: 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:

1947-2013: Quarterly data on nominal "Federal Government Current receipts," BEA Table 3.2, line 1, March 27, 2014 version. Note that all NIPA BEA data is on an accrual basis.

1879-1938: Monthly data on federal budget receipts. Source: NBER Macro-History Database
<http://www.nber.org/databases/macrophistory/contents/chapter15.html> . These data are on a cash basis.

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

1939-1946: Quarterly data on nominal federal receipts from *National Income, 1954 Edition, A Supplement to the Survey of Current Business* is used to interpolate the modern annual NIPA values. We construct the quarterly federal receipts interpolator from federal personal taxes + total corporate taxes + total indirect taxes.

1889-1928: Annual data on federal receipts. Source: Historical Statistics - fiscal year basis (e.g. fiscal year 1890 starts July 1, 1889).

1929-1946: Annual data on nominal "Federal Government Current receipts," BEA Table 3.2, line 1, March 27, 2014 version.

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:

1948-2013: 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/macrophistory/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:

1934-2013: 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/macroeconomy/contents/chapter13.html>

m13029a U.S. Yields On Short-Term United States Securities, 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
1890:2 - 2013:4 - All	17.88	495
1890:2 - 2013:4 - Slack	10.77	180
1890:2 - 2013:4 - No slack	13.69	315
1947:3 - 2013:4 - All	19.14	266
1947:3 - 2013:4 - Slack	2.03	82
1947:3 - 2013:4 - No slack	15.84	184

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 two lags, two lags of log real per capita GDP, current and two 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.				
2 year integral	0.75 (0.104)	0.69 (0.092)	0.79 (0.195)	0.661
4 year integral	0.84 (0.098)	0.76 (0.059)	0.96 (0.230)	0.342

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			
2 year integral	0.64	0.58	0.67
4 year integral	0.76	0.65	0.84
Omitting the quartic trend			
2 year integral	0.65	0.53	0.71
4 year integral	0.74	0.63	0.81
Considering 8% unemployment rate threshold			
2 year integral	0.75	0.78	0.77
4 year integral	0.84	0.76	0.96
HP filtered time-varying threshold (with $\lambda = 10^6$)			
2 year integral	0.75	0.82	0.75
4 year integral	0.84	0.82	0.88
Subsample: 1947-2013			
2 year integral	0.89	-5.52	1.05
4 year integral	0.47	24.97	0.62
Blanchard-Perotti identification			
2 year integral	0.56	0.73	0.51
4 year integral	0.80	0.89	0.77
Exclude World War II			
2 year integral	1.00	0.77	0.99
4 year integral	1.22	1.04	1.22
	Linear	Recession	Expansion
7 qtr. moving avg. output growth, $F(z)$			
2 year integral	0.76	0.56	0.76
4 year integral	0.84	0.65	0.85

Table 4. Comparison to Auerbach-Gorodnichenko (2012) Multipliers

Direct Comparisons	Extreme Recession ($F(z)=1$)	Extreme Expansion ($F(z)=0$)	Difference
AG's Estimates, Constant State			
5 year integral	2.24 (0.24)	-0.33 (0.20)	2.57
Jordà Method Applied to AG Specification			
5 year integral	0.84	-0.59	1.43
2 year integral	0.24	0.36	-0.12
Alternative Multipliers using AG's STVAR Estimates	Severe Recession ($F(z) \geq 0.95$)	Severe Expansion ($F(z) \leq 0.05$)	Difference
Constant State, No Feedback			
5 year integral	2.16	-0.31	2.47
2 year integral	1.56	0.10	1.46
Actual State Dynamics, No Feedback			
5 year integral	1.41	0.19	1.22
2 year integral	1.13	0.15	0.97
AG Partial Feedback			
5 year integral	1.36	-0.04	1.40
2 year integral	1.01	0.15	0.86
Actual State Dynamics, Partial Feedback			
5 year integral	1.07	0.14	0.93
2 year integral	1.07	0.12	0.95

Note: STVAR denotes the Smooth Transition Vector Autoregression used by AG-12. Impulse responses are calculated based on the VAR parameter estimates and auxiliary assumptions. The values in brackets under the multipliers give the standard errors. $F(z)$ is AG's indicator of the state of the economy. $F(z) = 1$ indicates the most severe recession possible and $F(z) = 0$ indicates the most extreme boom possible.

"Constant state" means that the impulse responses are calculated assuming that the economy remains in its current state for the duration of the multiplier. "Feedback" means that the estimates allow government spending to change the state of the economy going forward.

Table 5. Tests of Instrument Relevance Across Monetary Policy Regimes

	F-statistic	Number of observations
1890:2 - 2013:4 - All	17.88	495
1890:2 - 2013:4 - ZLB	4.59	97
1890:2 - 2013:4 - Normal	21.68	398
1890:2 - 2013:4 - ZLB, No WWII	0.73	73

Note: "ZLB" is when interest rates are near the zero lower bound or the Fed is being very accommodative of fiscal policy (1932q1-1951q1, 2008q4-2013q4). The F-tests are the joint significance of news variables in a regression of log real per capita government spending on its own two lags, two lags of log real per capita GDP, current and two 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
2 year integral	0.75 (0.104)	0.81 (0.139)	0.54 (0.196)	0.327
4 year integral	0.84 (0.098)	0.76 (0.074)	0.84 (0.525)	0.888

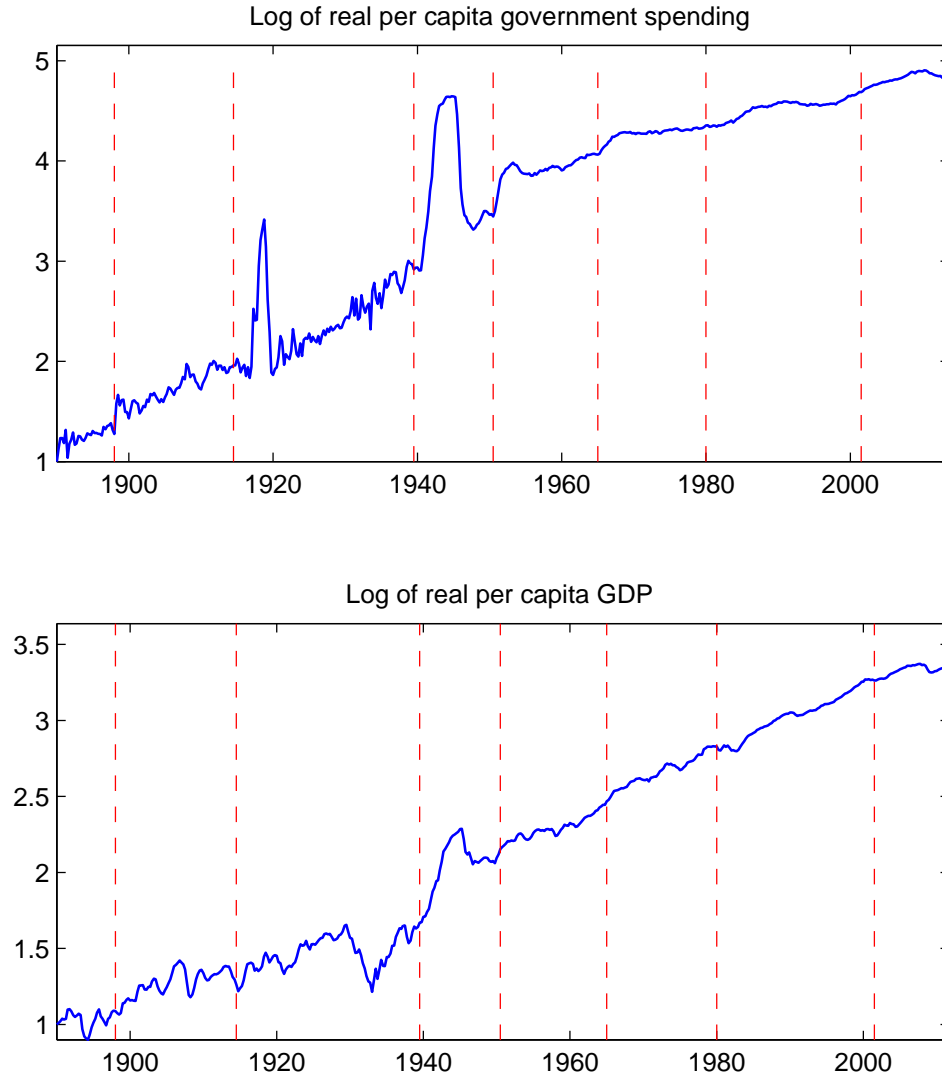
Note: The values in brackets under the multipliers give the standard errors.

Table 7. Robustness Checks of Multipliers Across Monetary Policy Regimes

	Linear Model	Near Zero Lower Bound	Normal	P-value for difference in multipliers across states
Defining ZLB as T-bill rate \leq 0.5				
2 year integral	0.75	0.70	1.01	
4 year integral	0.84	0.72	1.30	
Blanchard-Perotti identification				
2 year integral	0.56	0.66	0.51	
4 year integral	0.81	0.79	0.93	
Including taxes				
2 year integral	0.74	0.74	0.70	
4 year integral	0.83	0.72	1.41	
Excluding World War II				
2 year integral	1.00 (0.362)	1.59 (0.441)	0.54 (0.197)	0.071
4 year integral	1.22 (0.365)	1.11 (0.284)	0.84 (0.526)	0.701
Excluding World War II, including taxes				
2 year integral	1.01 (0.240)	1.26 (0.218)	0.70 (0.271)	0.120
4 year integral	1.27 (0.261)	0.99 (0.210)	1.42 (0.565)	0.467

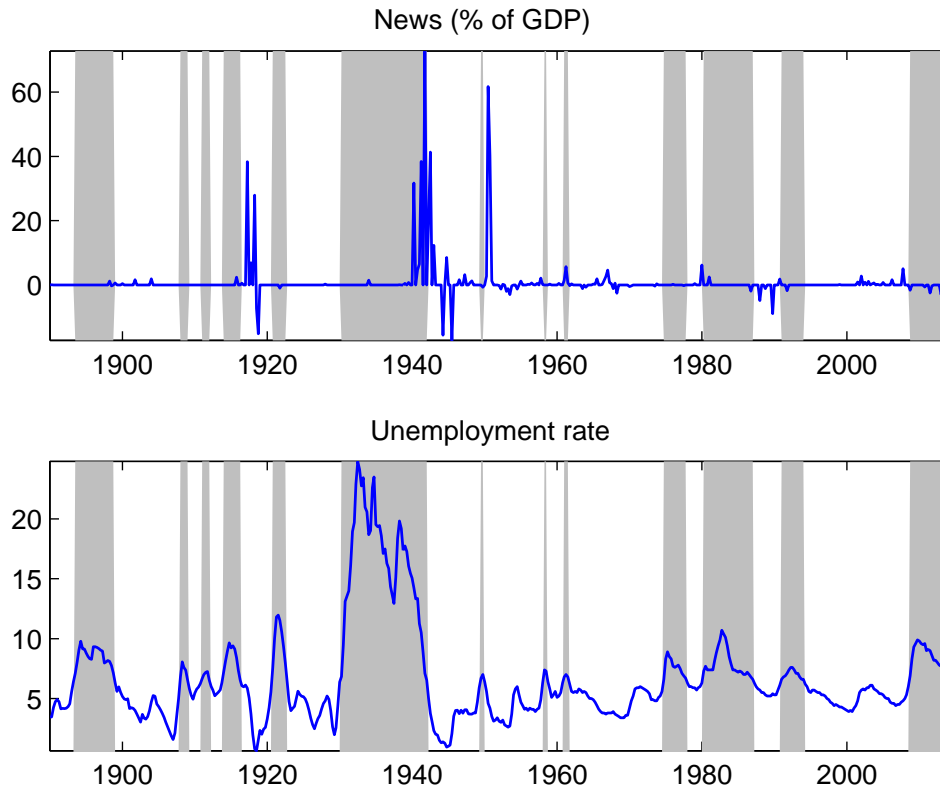
Note: The values in brackets under the multipliers give the standard errors.

Figure 1. Government Spending and GDP



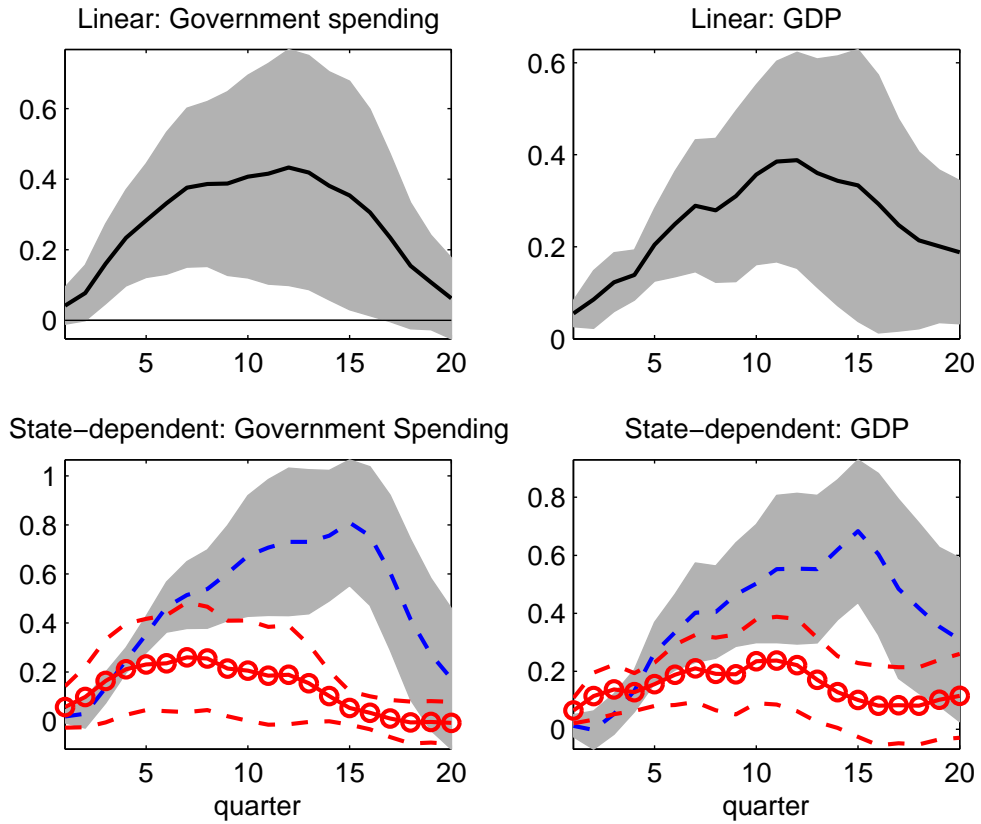
Note: The vertical lines indicate major military events: 1898q1(Spanish-American War), 1914q3 (WWI), 1939q3 (WWII), 1950q3 (Korean War), 1965q1 (Vietnam War), 1980q1 (Soviet invasion of Afghanistan), 2001q3 (9/11).

Figure 2. Military spending news and unemployment rate



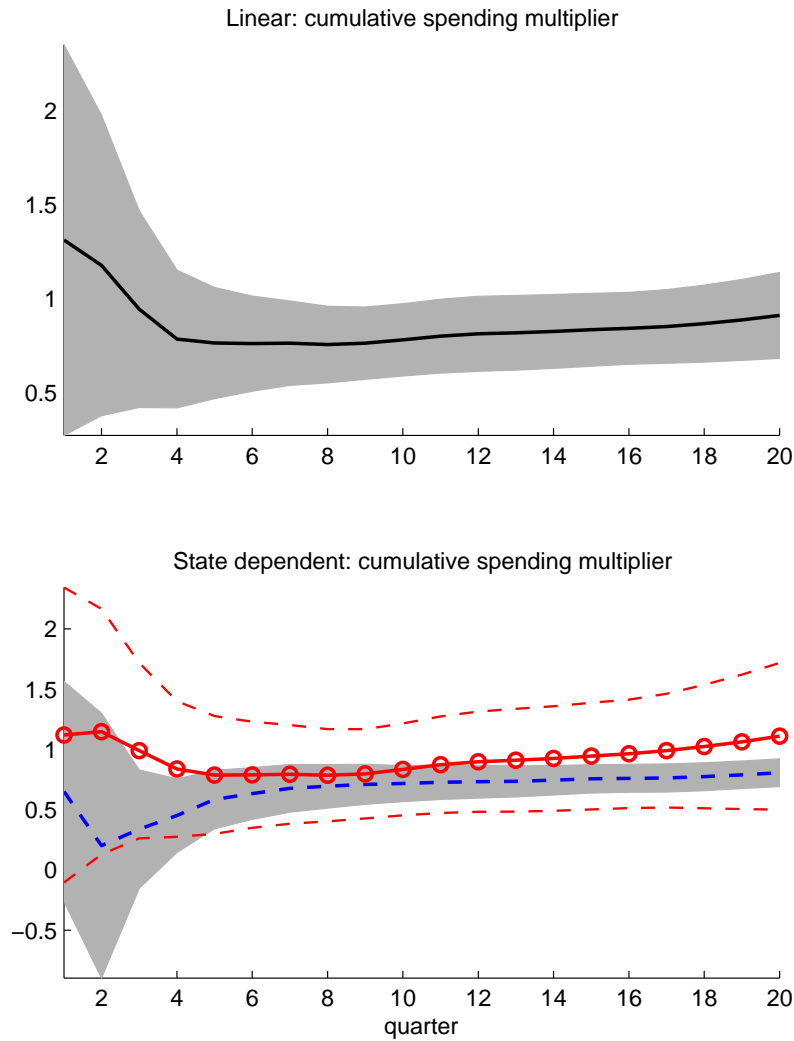
Note: Shaded areas indicate periods when the unemployment rate is above the threshold of 6.5%.

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



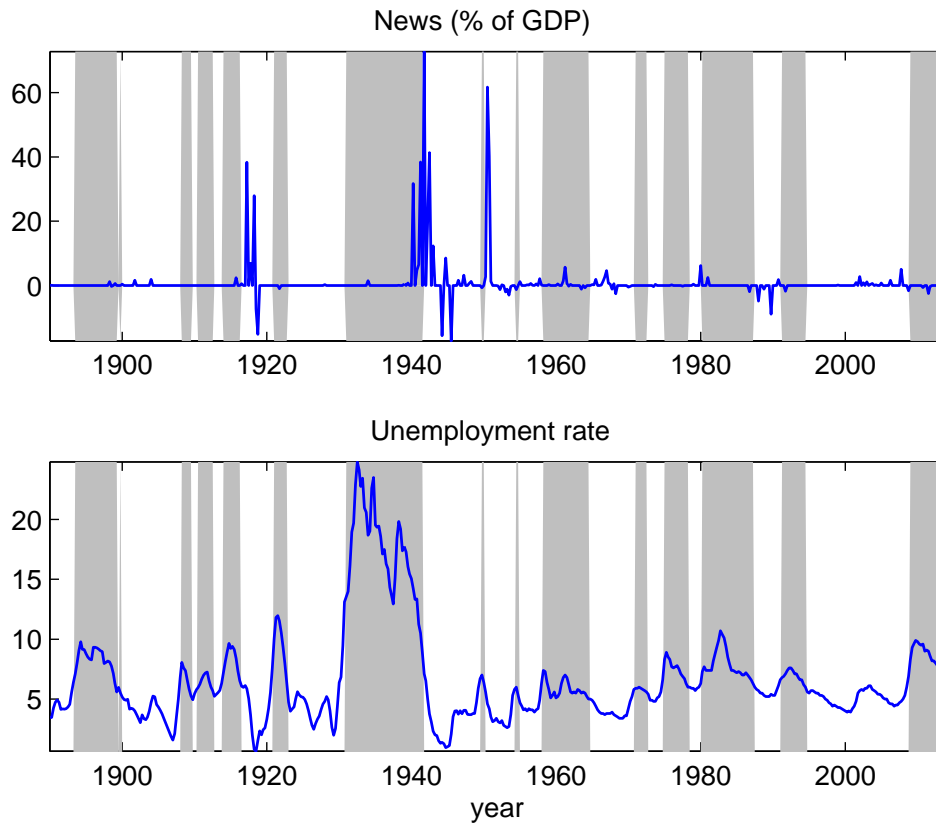
Note: Response of government spending and GDP to a news shock equal to 1% of GDP. The top panel shows the responses in the linear model. The bottom panel shows the state-dependent responses where the blue dashed lines are responses in the high unemployment state and the lines with red circles are responses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

Figure 4. Cumulative multipliers across slack states



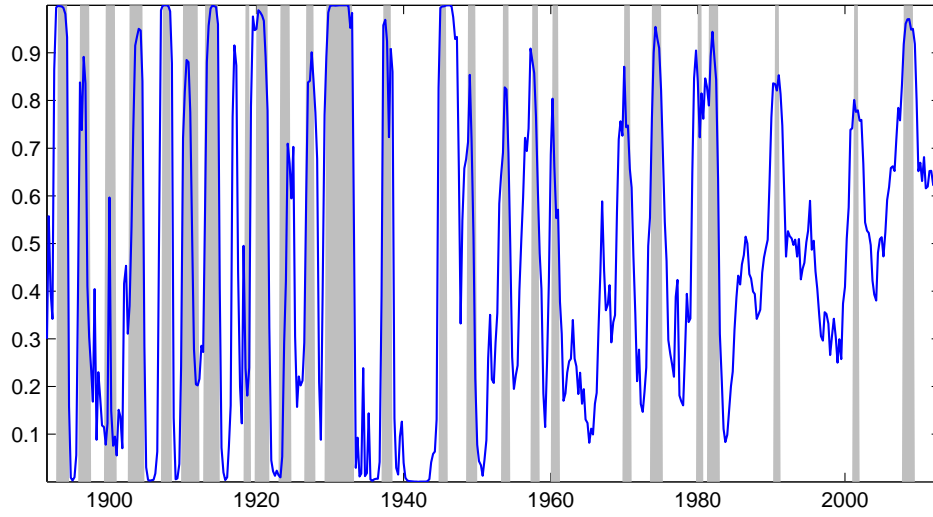
Note: Cumulative spending multipliers across different horizons. The top panel shows the cumulative multipliers in the linear model. The bottom panel shows the state-dependent multipliers where the blue dashed lines are multipliers in the high unemployment state and the lines with red circles are multipliers in the low unemployment state. 95% confidence intervals are shown in all cases.

Figure 5. Robustness check: New threshold of unemployment rate based on time-varying trend



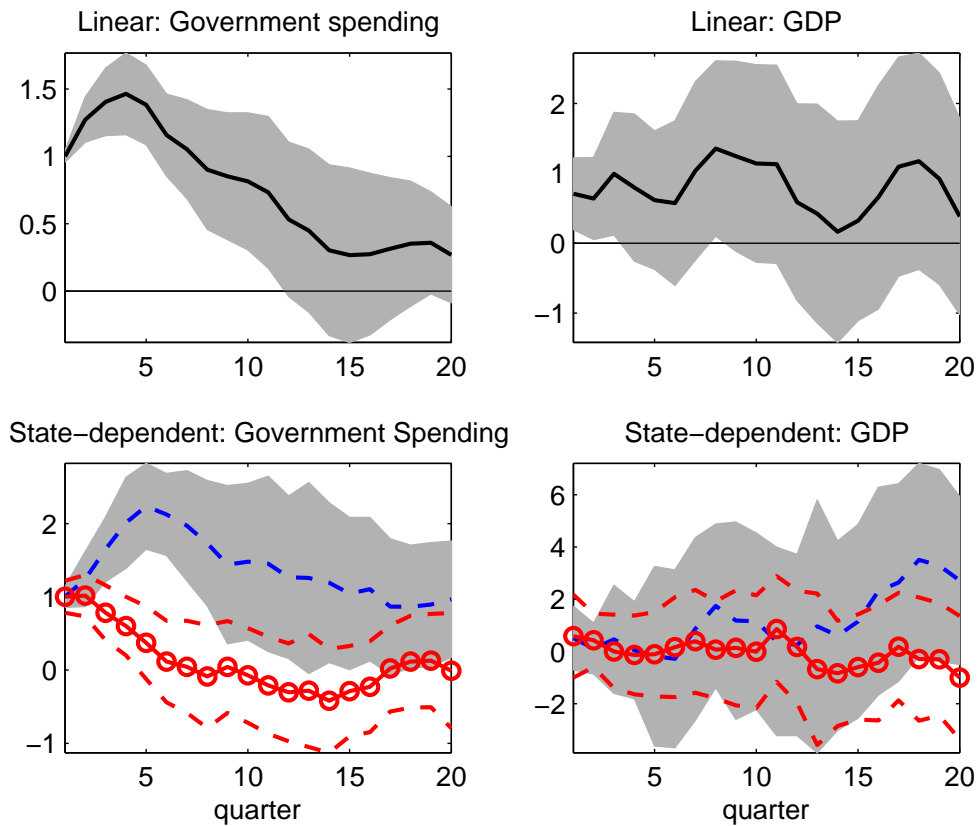
Note: Shaded areas indicate periods when the unemployment rate is above the time-varying trend based on HP filter with $\lambda = 10^6$.

Figure 6. Robustness check: New smooth transition threshold based on moving average of output growth



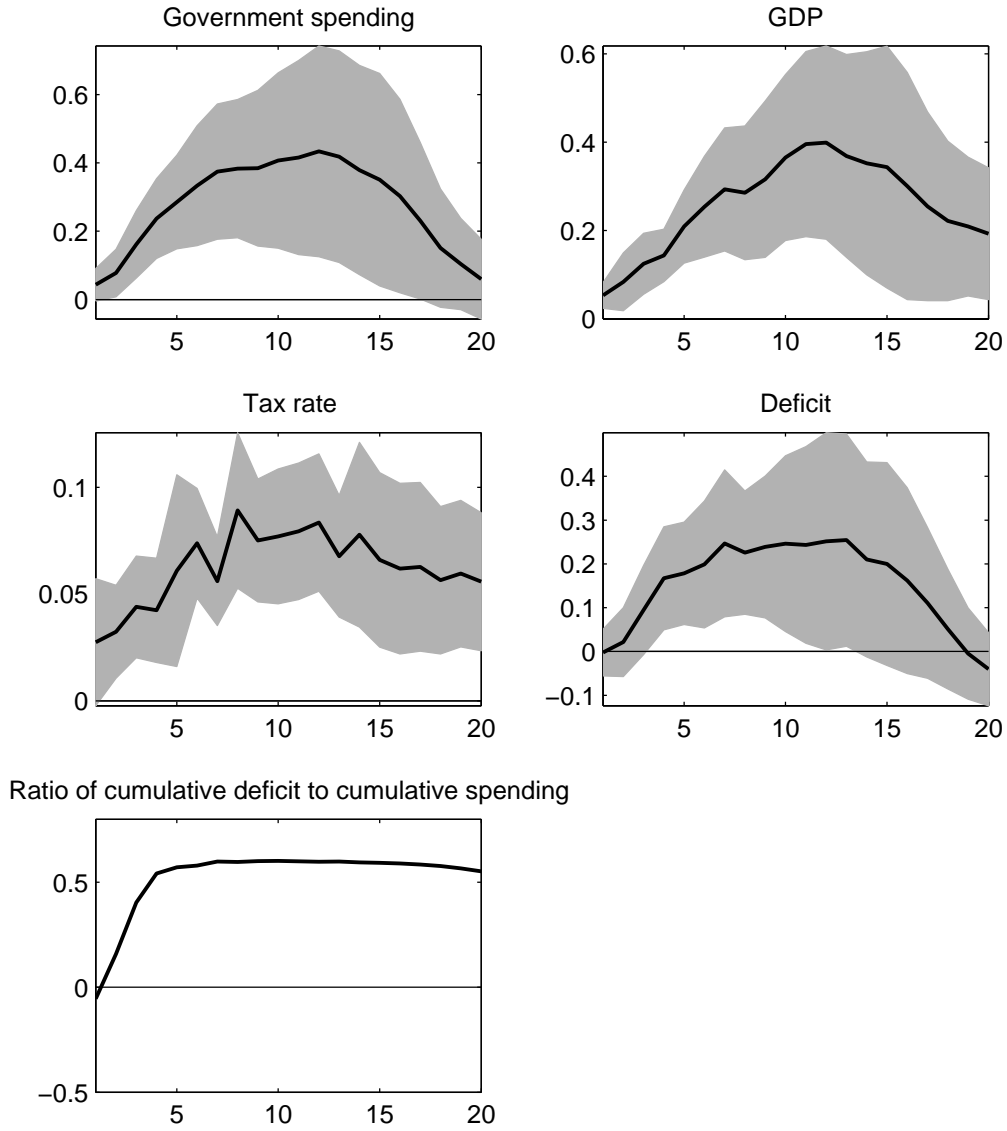
Note: The figures shows the weight on a recession regime, $F(z)$ and the shaded areas indicate recessions as defined by NBER.

Figure 7. Estimating Auerbach and Gorodnichenko (2012) with the Jorda method



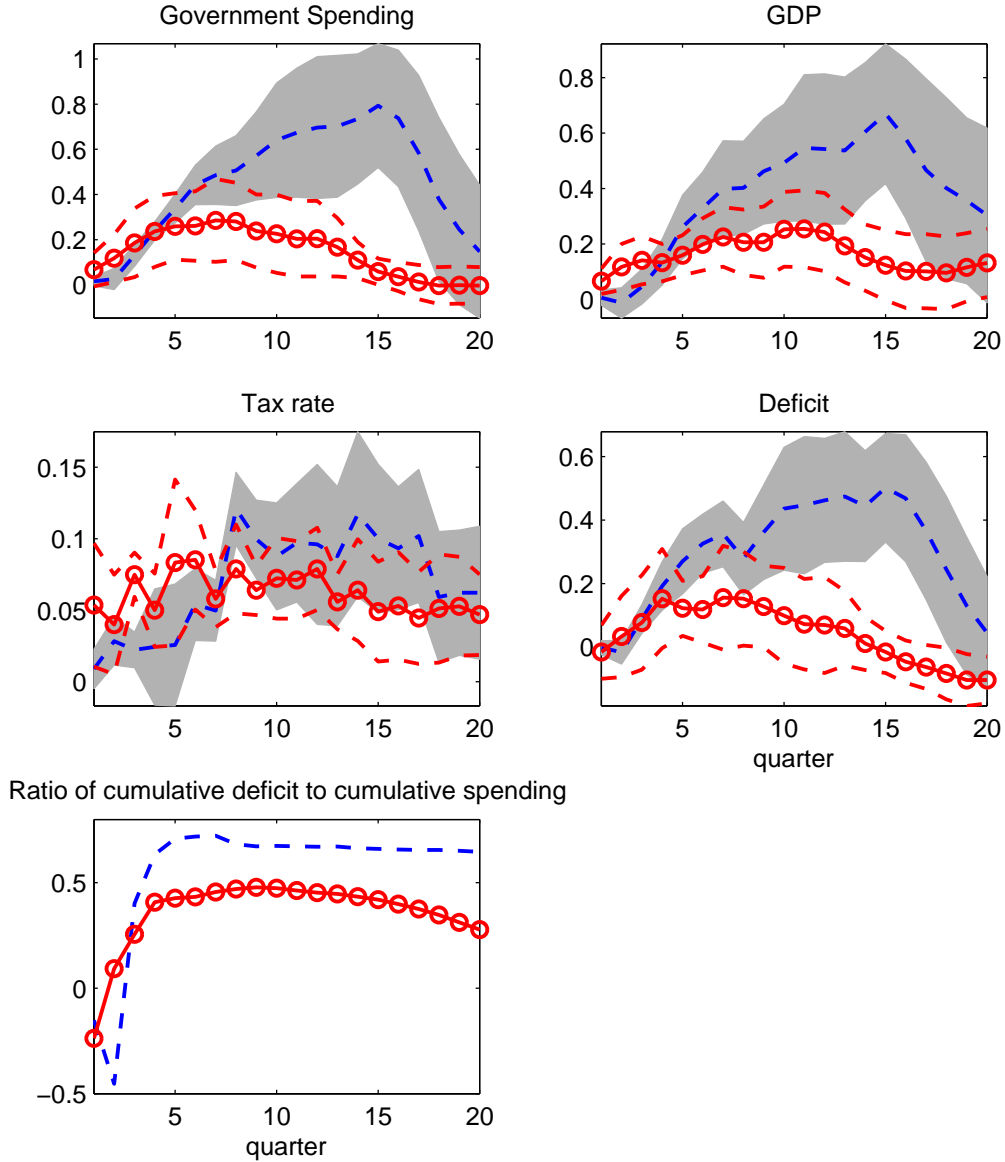
Note: Response of government spending and GDP to a government spending shock equal to 1% of GDP, with the same data, identification scheme and threshold definition as Auerbach and Gorodnichenko (2012). The top panel shows the responses in the linear model. The bottom panel shows the state-dependent responses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

Figure 8. Responses of taxes and deficits



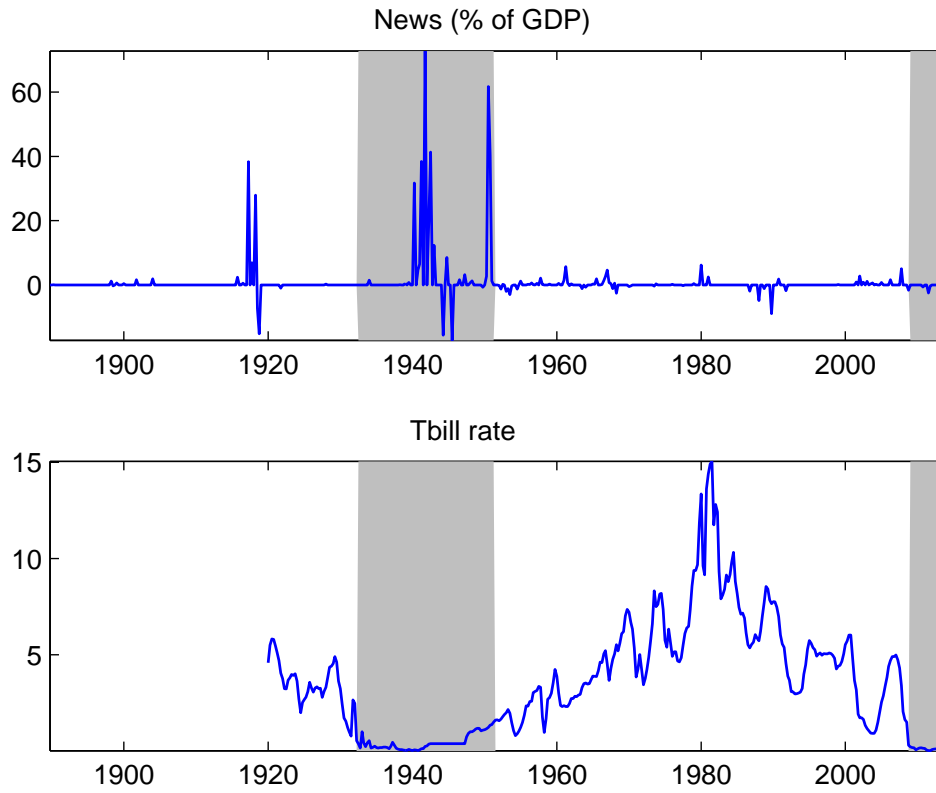
Note: These are responses for taxes and deficits in the linear model. The shaded areas indicate 95% confidence bands.

Figure 9. State-dependent responses of taxes and deficits: Considering slack state



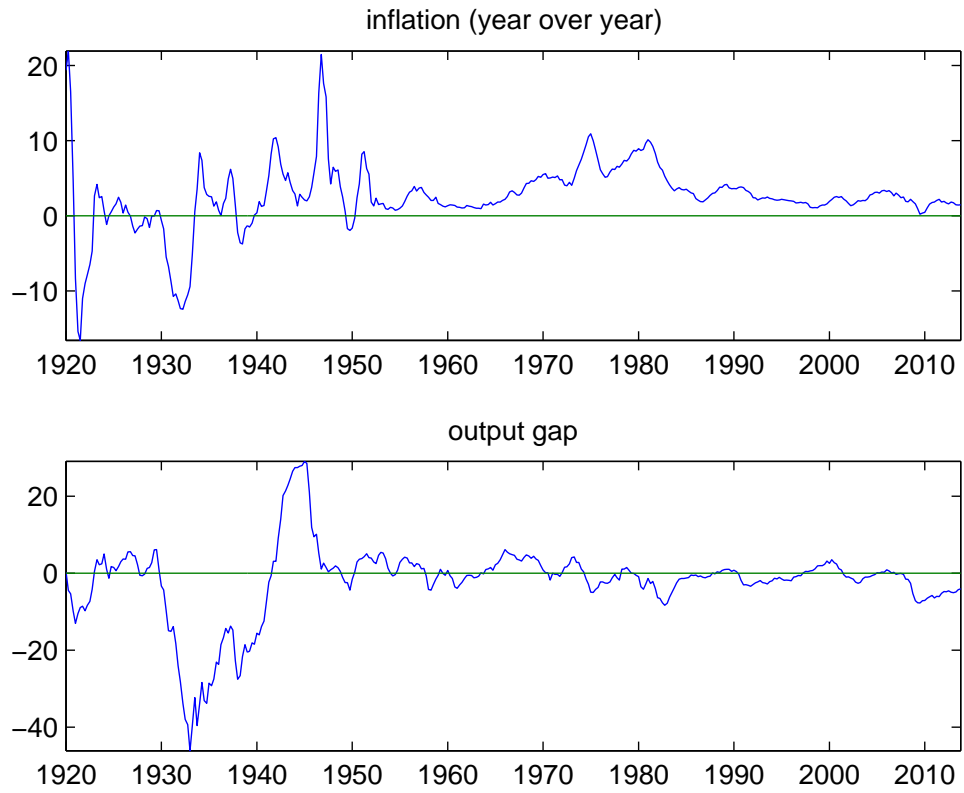
Note: These are state-dependent responses for taxes and deficits, where the black solid lines are responses in the high unemployment state and the lines with red circles are responses in the low unemployment state. 95% confidence intervals are also shown.

Figure 10. Military spending news and interest rate



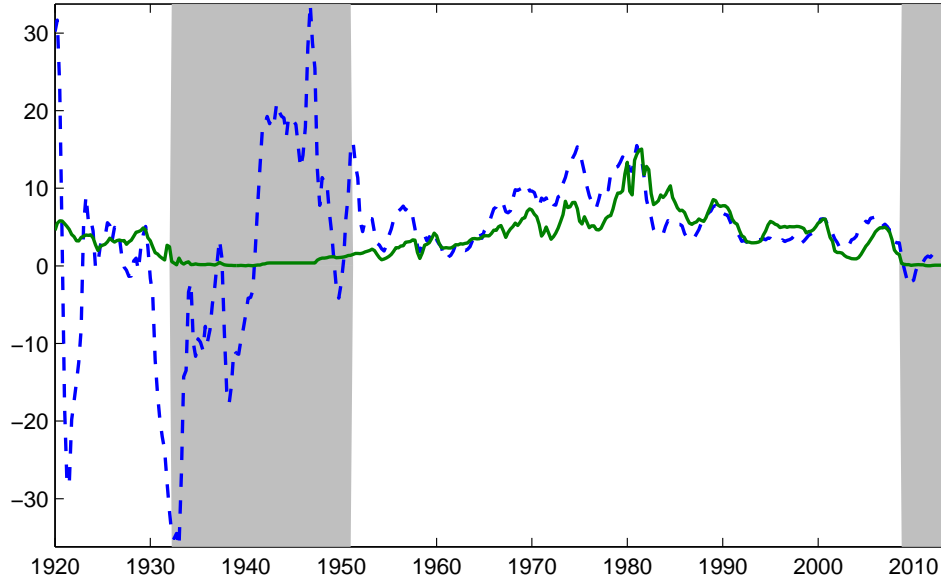
Note: Shaded areas indicate periods which we classify as the zero lower bound period for interest rate.

Figure 11. Inflation and output gap



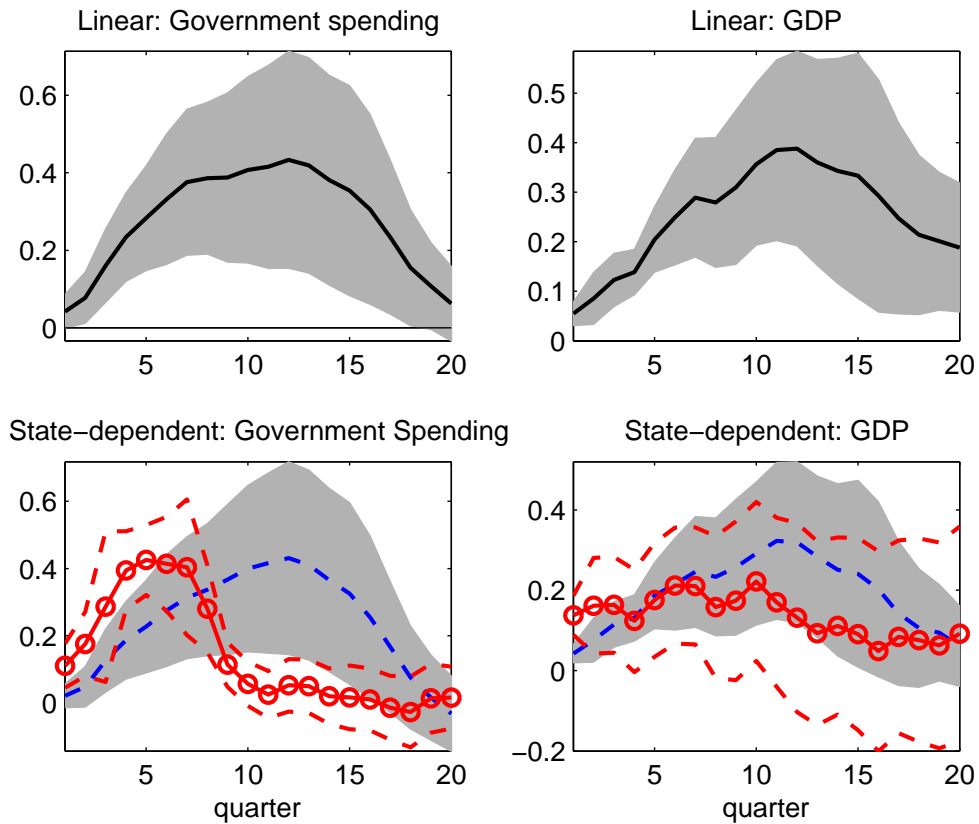
Note: The top panel shows the year-over-year GDP deflator inflation rate and the bottom panel shows the output gap, which is constructed as the percentage deviation between real GDP and potential GDP.

Figure 12. Taylor rule implied interest rate and the T-bill rate



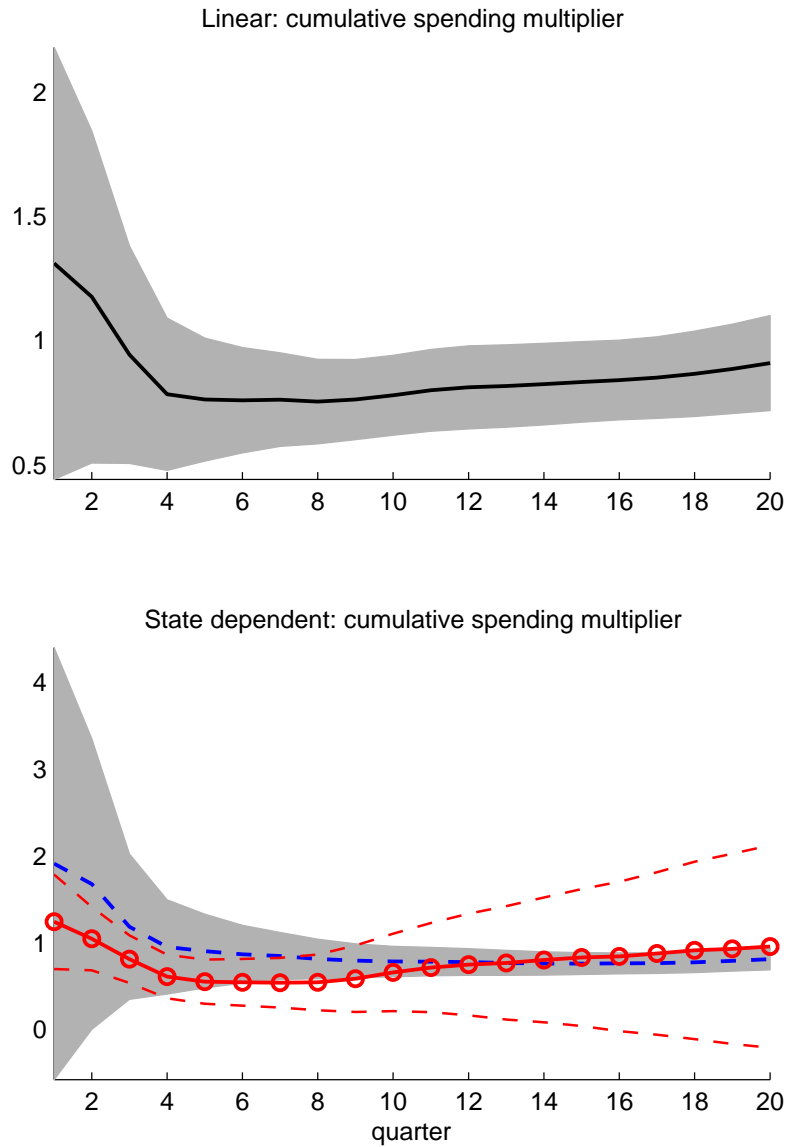
Note: The solid line shows the data for the 3-month T-bill rate, and the dashed line shows the Taylor-rule implied nominal interest rate. Shaded areas indicate periods which we classify as the zero lower bound period for interest rate.

Figure 13. 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 blue dashed 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.

Figure 14. Cumulative multipliers: Considering zero lower bound



Note: Cumulative spending multipliers across different horizons. The top panel shows the cumulative multipliers in the linear model. The bottom panel shows the state-dependent multipliers where the blue dashed lines are multipliers in the near zero-lower bound state and the lines with red circles are multipliers in the normal state. 95% confidence intervals are shown in all cases.