Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data

Valerie A. Ramey

University of California, San Diego, and National Bureau of Economic Research

Sarah Zubairy

Texas A&M University

We investigate whether US government spending multipliers are higher during periods of economic slack or when interest rates are near the zero lower bound. Using new quarterly historical US data covering multiple large wars and deep recessions, we estimate multipliers that are below unity irrespective of the amount of slack in the economy. These results are robust to two leading identification schemes, two different estimation methodologies, and many alternative specifications. In contrast, the results are more mixed for the zero lower bound state, with a few specifications implying multipliers as high as 1.5.

I. Introduction

What is the multiplier on government spending? The policy debates that started during the Great Recession have led to an outpouring of research on this question. Most studies have found estimates of modest multipliers in aggregate data, often below unity. If multipliers are indeed this low, they

We are grateful to Harald Uhlig, Roy Allen, Alan Auerbach, Graham Elliott, Yuriy Gorodnichenko, Jim Hamilton, Òscar Jordà, Michael Owyang, Carolin Pflueger, Garey Ramey, Tatevik Sekhposyan, Mark Watson, Johannes Wieland, anonymous referees, and participants at many conferences and seminars for very helpful suggestions. We thank Michelle Ramey for excellent research assistance. Data are provided as supplementary material online.

Electronically published March 7, 2018
[*Journal of Political Economy*, 2018, vol. 126, no. 2]
 0022-3808/2018/12602-0005$10.00

suggest that increases in government purchases do not stimulate private activity and that fiscal consolidations based on reducing government purchases are unlikely to do much harm to the private sector.

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 multipliers. Unfortunately, the data do not present us with clean natural experiments that can answer these questions. While the recent US stimulus package was purely deficit financed and was 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 higher than normal during recessions (e.g., Barro and Redlick 2011; Auerbach and Gorodnichenko 2012, 2013; Fazzari, Morley, and Panovska 2015). Another strand of the literature considers how monetary policy affects government spending multipliers. New Keynesian dynamic stochastic general equilibrium (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, Eichenbaum, and Rebelo 2011; Coenen et al. 2012).

This paper contributes to the empirical literature by conducting a comprehensive investigation of whether government spending multipliers in the United States 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. We show that the post–World War II US data do not contain enough information to distinguish multipliers across either of these states at most horizons. Extending the initial analysis in Owyang, Ramey, and Zubairy (2013), we exploit the fact that the entire twentieth century contains potentially richer information than the post-WWII data that have been the focus of most of the recent research. We create a new quarterly data set for the United States 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.

This paper extends the small, but growing, literature on state dependence of government spending multipliers in two additional ways. First, our paper analyzes state dependence involving the important zero lower bound state. Only two previous papers specifically estimated multipliers over an episode of the zero lower bound—Ramey (2011) for the United States and Crafts and Mills (2013) for the United Kingdom—but neither tested for differences relative to normal times. Second, our paper contributes to the general state-dependent multiplier literature by highlighting some key methodological issues that arise. In particular, we show that some of the most widely cited findings of high multipliers during recessions are due to assumptions that may be at odds with the data-generating process. We show that the finding of high multipliers during low-growth periods disappears when data-consistent assumptions are used.

Using Jordà's (2005) local projection method, we find no evidence that government spending multipliers are high during high-unemployment states. Most estimates of the multiplier are between 0.3 and 0.8. We find a statistically significant difference in multipliers across states only when we identify spending shocks following Blanchard and Perotti's (2002) method; however, the difference is due not to high multipliers in the high-unemployment state but to very low multipliers in the low-unemployment state. We perform extensive robustness checks with respect to our measures of state, sample period, the behavior of taxes, and alternative estimation frameworks and find little change in the estimates.

We find mixed evidence on the size of the multiplier at the zero lower bound. For the full sample, there is no evidence of multipliers greater than one at the zero lower bound. When we exclude the rationing periods of WWII, however, we find multipliers as high as 1.5 in the zero lower bound state in some cases.

We also demonstrate that most of the differences in conclusions between our work and that of the leading alternative study on state-dependent multipliers of Auerbach and Gorodnichenko (2012) lie in subtle, yet crucial, assumptions underlying 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, constructing impulse response functions in nonlinear models is fraught with complications. Furthermore, when we apply their threshold vector autoregression (VAR) method to our longer sample, but in a way that is more consistent with the data-generating process, we find results that are very similar to those produced by the Jordà method.

The paper proceeds as follows. We begin by discussing the motivation for using a historical sample and then conduct some case studies of wars in Section II. In Section III we introduce the econometric methodology. In Section IV, 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 also conduct various robustness checks. Section V tests theories that predict that multipliers should be greater when interest rates are at the zero lower bound. Section VI explores alternative methodologies and explains why our results are different from the preexisting estimates in the literature, and Section VII presents conclusions.

II. Historical Sample and Case Studies

In this section, we begin by motivating why we construct a new historical data set to study multipliers. We then briefly describe the data construction, leaving most details to the data appendix. Finally, because there are three wars in our sample that potentially play an influential role for our estimates, we conduct brief case studies of those three periods.

A. Why Use Historical Data?

The ideal way to measure the effects of government purchases on an economy would be to ask the International Monetary Fund to conduct a randomized control trial across countries, randomly assigning changes in government spending (and how they are financed) across countries and then using simple statistical techniques to estimate the effects. Obviously, such an experiment is impossible. Thus, macroeconomists must resort to estimating multipliers by exploiting "natural experiments" or other identification methods using time series on national historical data.[1] To be informative, the identified changes in government spending must be exogenous and big enough for their effects to be extracted from the many other economic shocks hitting the economy. The challenge becomes even greater once one attempts to estimate state-dependent multipliers since informative estimates require that the states span a sufficient portion of the sample and that the exogenous changes in government spending be spread across the states.

Long samples of historical data for the United States meet this challenge well since US historical data include many more periods of slack, one more extended period near the zero lower bound, and much larger variations in government spending during world wars. Historical samples come with their own potential problems, though. For example, one may wonder whether the US economy has changed so much over time that estimates from historical samples are uninformative for modern policy. We would argue that, if anything, the changes over time would reduce multipliers in recent years. The models that produce some of the highest mul-

[1] The natural experiments ideally involve aggregate data. As Nakamura and Steinsson (2014) and Farhi and Werning (2016) show, the multipliers estimated from natural experiments that involve cross-state or cross-province differences are not the same as aggregate multipliers. A number of cross-state analyses find some evidence of higher multipliers when the state unemployment rate is higher, but translating those to aggregate multipliers is not straightforward.

tipliers are ones in which a higher fraction of consumers are rule-of-thumb or hand-to-mouth consumers. Increases over time in financial market access and consumer sophistication should reduce the fraction of rule-of-thumb consumers, thus reducing multipliers in recent years. Separately, monetary policy and fiscal policy have been conducted differently over various periods, but both the pre-WWII and post-WWII sample display periods of more or less monetary accommodation and more or less deficit financing of government spending. Thus, we believe that estimates from historical samples can be informative for modern policy debates.

Alternatively, one might argue that since wars are "abnormal," we should exclude them. Friedman (1952, 612) countered this argument years ago:

> The widespread tendency in empirical studies of economic behavior to discard war years as "abnormal," while doubtless often justified, is, on the whole, unfortunate. The major defect of the data on which economists must rely—data generated by experience rather than deliberately contrived experiment—is the small range of variation they encompass. Experience in general proceeds smoothly and continuously. In consequence, it is difficult to disentangle systematic effects from random variation since both are of much the same order of magnitude.
>
> From this point of view, data for wartime periods are peculiarly valuable. At such times, violent changes in major economic magnitudes occur over relatively brief periods, thereby providing precisely the kind of evidence that we would like [to] get by "critical" experiments if we could conduct them. Of course, the source of the changes means that the effects in which we are interested are necessarily intertwined with others that we would eliminate from a contrived experiment. But this difficulty applies to all our data, not to data for wartime periods alone.

We also believe that there is much to be learned from wartime periods but do recognize the potential effects of confounding factors. We will discuss those factors in the case studies below and in the sample exclusions in the econometric estimation.

A separate issue is whether the economy responds to military spending in the same way it would respond to other types of government purchases, such as nondefense consumption, infrastructure, and so forth. This is a valid concern and is related to the standard question of whether a local average treatment effect is equal to the average treatment effect. Our baseline instrument will be an updated version of Ramey's (2011) military news variable, so it captures only news about changes in military spending and most of actual spending arrives with delay. In order to broaden our range of treatments, we will also use the Blanchard and Perotti (2002) shock.

This identification scheme is based on the assumption that within-quarter government spending does not contemporaneously respond to macroeconomic variables. By their nature, the Blanchard-Perotti shocks lead to immediate rises in government spending with peaks close to impact, whereas the military news shocks lead to delayed rises in government spending. Using the Blanchard-Perotti shock involves a trade-off, however, since this type of shock is both more sensitive to potential measurement errors in the historical data and subject to the critique that it is likely to have been anticipated.

B. *Data Description*

In order to exploit the information in the historical sample, we construct quarterly data from 1889–2015 for the United States. 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. From 1939 to the present, 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 United States* (Carter et al. 2006) for 1889–1928 and the National Income and Product Accounts (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 Consumer Price Index (CPI) for 1939–46. 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

[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 World War I in our analysis.

[4] We also check the robustness of our results by using alternative series constructed by Christina Romer in the supplemental appendix, available online. See Romer (1999) for a discussion of her data.

1929. Following Gordon and Krenn (2010), we use monthly federal outlay series from the NBER Macrohistory database to interpolate annual government spending from 1889 to 1938. We use the 1954 quarterly NIPA data from 1939–46 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.

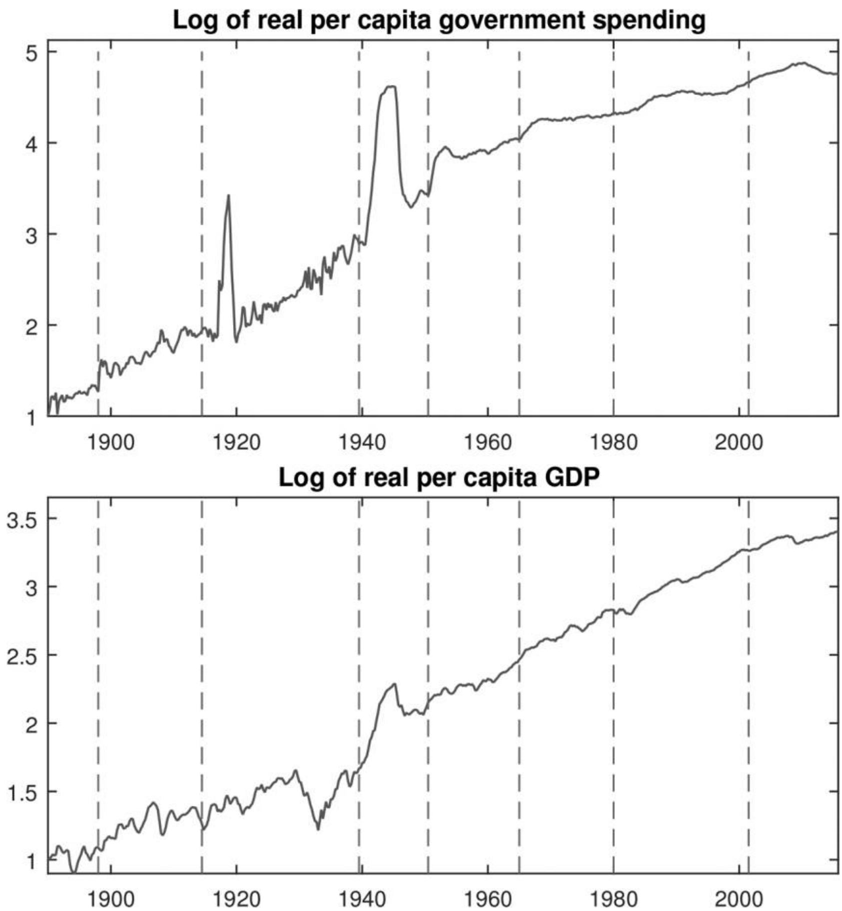

FIG. 1.—Government spending and GDP. The vertical lines indicate major military events: 1898q1 (Spanish-American War), 1914q3 (WWI), 1939q3 (WWII), 1950q3 (Korean War), 1965q1 (Vietnam War), 1980q1 (Soviet invasion of Afghanistan), and 2001q3 (9/11). Color version available as an online enhancement.

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 not important for our baseline multiplier estimates because we instrument for government spending using a narrative series that is uncorrelated with this measurement error.

The unemployment series is constructed by interpolating Weir's (1992) annual unemployment series, adjusted for emergency worker employment.[5] For 1929–48 we use the monthly unemployment series available from the NBER Macrohistory database back to April 1929 to interpolate. Before 1929, we interpolate Weir's 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 reveals that they are surprisingly close.

Because it is important to identify a shock that not only is exogenous to the state of the economy but is also unanticipated, we use narrative methods to extend the Ramey (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 arises through the wealth effect. Thus, the news series is constructed as changes in the expected present discounted value of government spending. The narrative underlying series is available in Ramey (2016). The particular form of the variable used as the shock is the nominal value divided by a one-quarter lag of the GDP deflator times trend real GDP. The real GDP time trend is estimated as a sixth-degree polynomial for the logarithm of GDP, from 1889q1 through 2015q4 excluding 1930q1–1946q4.[6] This method for estimating trend real GDP is similar to the method used by Gordon and Krenn (2010). We display the military news series in later sections when we construct the states so that one can see the juxtaposition.

For the local average treatment effect issues discussed in the last section, we will also explore results using the Blanchard and Perotti (2002) shock. This shock is identified simply from a Cholesky decomposition in a VAR with total government spending ordered first. Unfortunately, because this shock is constructed directly from the government spending series, any measurement error in that series will also be incorporated into the shock, which can lead to attenuation of the multiplier estimates. We will

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

[6] We also show the robustness of our results for an alternative potential GDP measure in the online supplemental appendix.

show that the relevance of each shock as an instrument varies by horizon and that using both as instruments together can have advantages.

C. *Case Studies of Three Wars*

Our main results are based on time-series econometrics. Nevertheless, since the wartime periods contain influential observations for the estimates, it is useful to give a brief overview of the three most important wars in our sample: WWI, WWII, and the Korean War. As Ramey (2013) argues, if the within-quarter government spending multiplier is greater than unity, then the response of private spending (i.e., GDP minus government spending) must be positive. Thus, it is instructive to look at the comovement of private spending and government spending. Figure 2 shows real private spending (dashed line) and real government spending (solid line), both deflated with the same GDP deflator but not divided by trend, in the left column, the military news shock in the middle panel, and the civilian unemployment rate in the right column. Each row shows the data from one of the three wars. The shaded areas in the middle column indicate times when interest rates were near the zero lower bound.

Consider first WWI. The war started in Europe in August 1914, but the United States did not expect to get involved until subsequent events led the United States to break off official relations with Germany in February 1917 and to declare war in April 1917. Both the first large military news shock and the first small jump in government spending occurred in the second quarter of 1917. Government spending rose rapidly to a peak of 33 percent of GDP at the end of 1918, when the armistice was signed.

The graphs highlight several key aspects. First, private spending tended to move in the opposite direction of government spending during WWI. There was no mandatory rationing in the United States during WWI, only a campaign for victory gardens and voluntary rationing of food to show solidarity with the European allies. Thus, the behavior of private spending cannot be attributed to rationing. Second, the unemployment rate had already fallen below 6 percent when government spending began to increase. The civilian unemployment rate continued to decline as government spending increased, in large part because of the dramatic rise in the armed forces: the armed forces rose from 0.4 percent of the total labor force (civilian plus armed forces) in 1916 to 9.9 percent in 1918q4. Thus, WWI illustrates the case of big government spending shocks hitting the economy when there was not much slack and interest rates were well above the zero lower bound.[7] It appears that government spending partially

[7] The Federal Reserve had been established only in 1914. At the start of WWI, it lowered the discount rate from 5.75 percent to 3.75 percent but then raised it to 4.56 percent after the United States became involved.

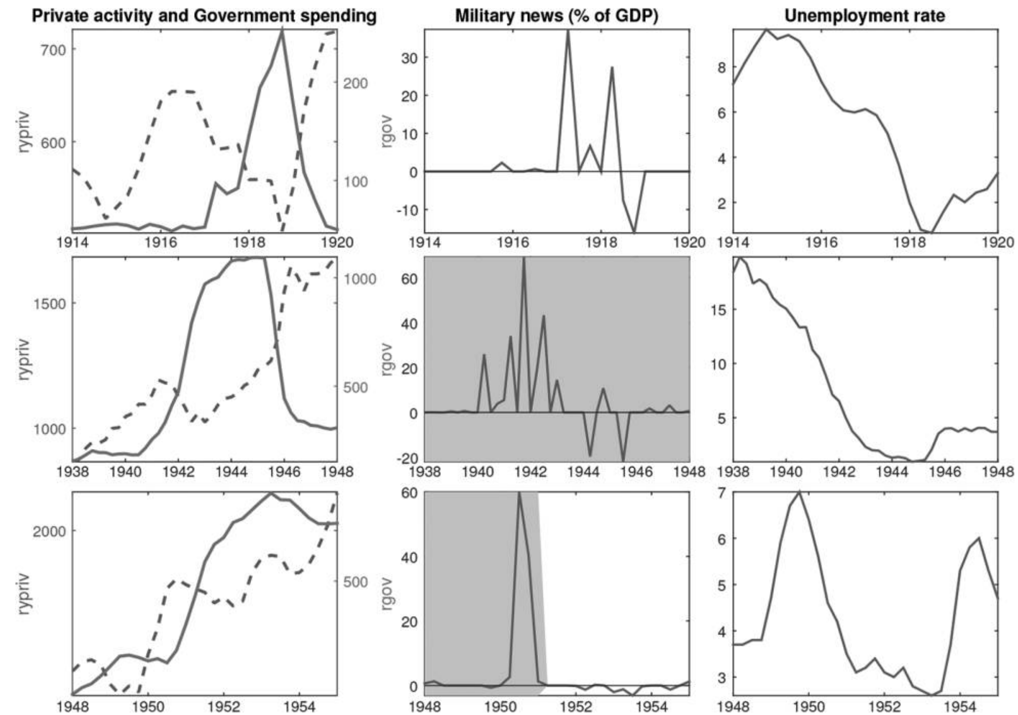

Fig. 2.—Evolution of variables during war episodes. The first column shows real private activity (left axis, dashed line) and real government spending (right axis, solid line). The second column shows military spending news, with shaded areas indicating periods we classify as the zero lower bound period for interest rate, and the third column shows the civilian unemployment rate. The first row corresponds to the period around World War I, the second row shows the time period around World War II, and the last row shows the period around the Korean War. Color version available as an online enhancement.

crowded out private spending. Overall, GDP rose and the unemployment rate fell, so the multiplier appears to be between zero and one during WWI.

In contrast, the buildup to WWII occurred when there was significant slack in the economy and the economy was at the zero lower bound. The war in Europe began in September 1939, and the ominous events of spring 1940 made it clear that the United States had to raise defense spending dramatically (see, e.g., the narratives of Gordon and Krenn [2010] and Ramey [2016]). As the second row of figure 2 shows, the civilian unemployment rate (defined here to include emergency workers in New Deal jobs) was falling steadily in 1938 and 1939 but was still 14 percent when the first big news shock hit in 1940. The US government imposed the draft in September 1940, and the unemployment rate continued to fall as the armed forces' percentage of the total labor force rose from 0.6 percent to over 18 percent.

Government spending rose from around 15 percent of GDP in early 1940 to almost 50 percent of GDP in 1944 and 1945 and then fell to 17 percent by the end of 1946. Meanwhile, private spending rose briskly from 1938 through the first half of 1941 and then stalled for the rest of the war. It soared when government spending fell at the end of the war.

There were two important complicating factors during WWII. The first was the dramatic rise in the labor force participation rate, due to both conscription and patriotism. The total labor force (civilian and military) rose 12 percent from 1939 to 1945. This rise allowed much more output to be produced than one would expect during non–war times. The second factor was the presence of price and credit controls and rationing, which began to be imposed on some goods in early 1942 and were lifted at the end of the war. The standard story is that private spending declined in WWII because of the rationing and rose after WWII when rationing was lifted. However, this story does not discuss the counterfactual: what would private spending have done if the government had not imposed price controls and rationing? It is not implausible to believe that changes in relative prices, interest rates, and other market forces would have led private spending to respond in a similar way.[8]

In sum, WWII contains potentially rich information because interest rates were at the zero lower bound before, during, and after the war, whereas the unemployment rate was elevated only before 1942. However, how rationing and conscription affected the path of private spending relative to what it would have done if prices, wages, and interest rates had been allowed to adjust remains an outstanding question.

Consider finally the Korean War, shown in the bottom row of figure 2. North Korea invaded South Korea in the last days of June 1950, and the

[8] McGrattan and Ohanian (2010) argue that the neoclassical model explains the behavior of quantities very well.

first big spending news shock hit in 1950q3. Government spending itself rose slightly in 1950q4 and then briskly in 1951 and 1952. As discussed later in the paper, we time the ending of the zero lower bound period as the Treasury Accord of March 1951. Unemployment was already low when the war started, and conscription contributed to further declines as the war progressed, with the armed forces' share of the total labor force rising from 2.3 percent in 1950 to 5.5 percent in 1952. Private spending was rising briskly before the war started and before government spending rose significantly, and then slowed down.

These case studies highlight several elements of the historical data we use. First, the wars give multiple, potentially informative, observations for big changes in government spending. Second, some of those changes come when the unemployment rate is high and some when it is low, and some when interest rates were near the zero lower bound. Third, confounding factors, such as the effects of military conscription, temporary increases in labor force participation, and controls on the economy, must be kept in mind.[9]

III. Econometric Methodology

In this section, we discuss a number of important details of the methodology. We first describe the Jordà local projection method that we use for our baseline estimates. We then discuss several pitfalls in calculating multipliers. We show that several widely used methods for translating estimates to multipliers can result in upward biases in multipliers. In addition, we introduce a new instrumental variables method for estimating cumulative multipliers in a one-step instrumental variables regression. This new method also allows us to use multiple candidates for government spending shocks at the same time.

A. Model Estimation Using Local Projection

We use Jordà's (2005) local projection method to estimate impulse responses and multipliers in our baseline. 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.[10] The Jordà method simply requires estimation of a series of regressions for each horizon h for each variable. The linear model looks as follows:

$$x_{t+h} = \alpha_h + \psi_h(L) z_{t-1} + \beta_h \text{shock}_t + \varepsilon_{t+h} \quad \text{for} \quad h = 0, 1, 2, \ldots, \qquad (1)$$

[9] The online supplemental appendix shows the behavior of taxes and deficits during the three wars. Both WWI and WWII were financed by a mix of deficit spending and taxes. As Ohanian (1997) shows, the Korean War was mostly financed with tax increases.

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

where x is the variable of interest, z is a vector of control variables, $\psi_h(L)$ is a polynomial in the lag operator, and shock is the identified shock. The baseline shock is the defense news variable scaled by trend GDP. Our vector of baseline control variables, z, contains real per capita GDP and government spending, each divided by trend GDP. In addition, z includes lags of the news variable to control for any serial correlation in the news variable. The term $\psi(L)$ is a polynomial of order 4. When we employ the Blanchard-Perotti identification, the shock is simply given by current government spending, since the set of controls, z, includes lagged measures of GDP and government spending. Thus, this is equivalent to the Blanchard-Perotti structural VAR (SVAR) identification.[11] 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:

$$\begin{aligned} x_{t+h} = {} & I_{t-1}[\alpha_{A,h} + \psi_{A,h}(L)z_{t-1} + \beta_{A,h}\text{shock}_t] \\ & + (1 - I_{t-1})[\alpha_{B,h} + \psi_{B,h}(L)z_{t-1} + \beta_{B,h}\text{shock}_t] + \varepsilon_{t+h}, \end{aligned} \tag{2}$$

where 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 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).

B. *Pitfalls in Calculating Multipliers*

We now highlight two potential problems that affect multipliers computed not only from nonlinear VARs but also from all of the standard linear SVARs used in the literature.

1. Logs versus Levels

The first problem concerns the conversion of elasticities to multipliers. The usual practice in the literature is to use the log of variables, such as real

[11] Blanchard-Perotti identification also includes taxes in the VAR. We show in the following sections that our results for both Blanchard-Perotti and news shocks are robust to the inclusion of taxes in the set of controls.

GDP, government spending, and taxes. However, the estimated impulse response functions do not directly reveal the government spending multiplier because the estimated elasticities must be converted to dollar equivalents. Virtually all analyses using VAR methods obtain the spending multiplier by using an ex post conversion factor based on the sample average of the ratio of GDP to government spending, Y/G.

We first noticed a potential problem with this method when we extended our sample back in time. In the post-WWII sample, Y/G varies between 4 and 7, with a mean of 5. In our full sample from 1889–2015, Y/G varies from 2 to 24 and with a mean close to 8. We realized that we could estimate the same elasticity of output with respect to government spending but derive much higher multipliers simply because the mean of Y/G was so much higher. In the online supplemental appendix, we show the results of experiments indicating that using an ex post conversion factor biases the multiplier estimates up in our sample.

In order to avoid this bias, we use Gordon and Krenn's (2010) transformation. Instead of taking logarithms of the variables, they divide all NIPA variables by an estimate of potential, or trend, GDP. This puts all NIPA variables in the same units, so that one can estimate the multiplier directly. We do this as well, using a polynomial to estimate trend real GDP (as discussed previously in the data description).

An alternative transformation is the one used by Hall (2009) and Barro and Redlick (2011). Owyang et al. (2013), as well as previous versions of this paper, used that transformation. The estimates are very similar. We chose the Gordon and Krenn (2010) transformation because that transformation can also be used in a VAR. Later, we will be comparing our baseline estimates to those from a threshold VAR.

2. Computing Multipliers in a Dynamic Environment

The second pitfall concerns the definition of a multiplier in a dynamic setting. The original Blanchard and Perotti (2002) paper defined the multiplier as the ratio of the peak of the output response to the initial government spending shock. Numerous papers have used this same definition, or variations, such as the average of the output response to the initial government shock (e.g., Auerbach and Gorodnichenko 2012, 2013). As argued by Mountford and Uhlig (2009), Fisher and Peters (2010), and Uhlig (2010), multipliers should instead be calculated as the integral of the output response divided by the integral government spending response.[12] The inte-

[12] Mountford and Uhlig (2009) and Uhlig (2010) calculate a present value multiplier, using the long-run average interest rate to discount. We used the simple cumulative multiplier because of its close relationship to the areas under the impulse response functions; however, our robustness tests indicate that the present value and simple cumulative multipliers are very similar and are shown in the online supplemental appendix.

gral multipliers address the relevant policy question because they measure the cumulative GDP gain relative to the cumulative government spending during a given period. As we will discuss later, the Blanchard-Perotti method of reporting multipliers tends to produce higher estimates of multipliers relative to the cumulative method.

In fact, the cumulative multiplier is very easy to estimate in one step as an instrumental variable (IV) estimation. In particular, one can estimate the following equation in the linear case:

$$\sum_{j=0}^{h} y_{t+j} = \gamma_h + \phi_h(L) z_{t-1} + m_h \sum_{j=0}^{h} g_{t+j} + \omega_{t+h} \quad \text{for } h = 0, 1, 2, \ldots, \tag{3}$$

using shock_t as an instrument for $\Sigma_{j=0}^{h} g_{t+j}$. Here, $\Sigma_{j=0}^{h} y_{t+j}$ is the sum of the GDP variable from t to $t + h$ and $\Sigma_{j=0}^{h} g_{t+j}$ is the sum of the government spending variable from t to $t + h$.[13] This one-step estimate of the cumulative multiplier at horizon h, m_h, is identical to the result from the following three-step method: (i) estimate equation (1) for GDP for each horizon j up to h and sum the β_j; (ii) estimate equation (1) for government spending for each horizon j up to h and sum those β_j; (iii) compute the multiplier as the answer to step 1 divided by the answer to step 2.[14] This one-step IV method has multiple advantages. First, the standard error of the multiplier is estimated directly. Second, both the shock and the government spending variable can have measurement error as long as their measurement errors are uncorrelated. Third, formulating the estimation as an IV problem highlights the importance of instrument relevance. Fourth, one can also use more than one instrument per endogenous variable if additional instruments are available. This can be useful since the leading government spending shocks tend to be relevant at different horizons. In subsequent sections, we show multipliers that are estimated using military news shocks and Blanchard-Perotti shocks separately, as well as in combination.

The one-step equation for the state-dependent case is given by

$$\begin{aligned} \sum_{j=0}^{h} y_{t+j} = {} & I_{t-1}\left[\gamma_{A,h} + \phi_{A,h}(L) z_{t-1} + m_{A,h} \sum_{j=0}^{h} g_{t+j}\right] \\ & + (1 - I_{t-1})\left[\gamma_{B,h} + \phi_{B,h}(L) z_{t-1} + m_{B,h} \sum_{j=0}^{h} g_{t+j}\right] + \omega_{t+h}, \end{aligned} \tag{4}$$

[13] If one prefers to calculate present value cumulative multipliers, one can redefine the summation variables as discounted sums.

[14] The results are identical only if all of the regressions are estimated on the same sample; i.e., the regressions for horizons 0, 1, ... must also drop the h last observations.

using $I_{t-1} \times \text{shock}_t$ and $(1 - I_{t-1}) \times \text{shock}_t$ as the instruments for the respective interaction of cumulative government spending with the two state indicators. Again, this produces state-dependent multipliers, $m_{A,h}$ and $m_{B,h}$, that are identical to those estimated and calculated using the three-step method, as long as the sample is held constant. Moreover, one can use additional instruments if they are available.

IV. Multipliers during Times of Slack

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. Other than the zero lower bound papers, which make a distinct argument that we will discuss below, there is only a limited literature analyzing rigorous models that produce fiscal multipliers that are higher during times of high unemployment. Michaillat (2014) is one of the few examples, but his model applies only to government spending on public employment.[15] Thus, there is still a gap between Keynes's original notion and modern theories.

In this section, we analyze the issue empirically. Section IV.A discusses our measure of slack and shows graphs of the data and periods of slack. Section IV.B presents statistics showing the relevance of the military news shock, the Blanchard-Perotti shock, and their combination at various horizons. Section IV.C presents the main results. Section IV.D conducts robustness checks.

A. *Measurement of Slack States*

There are various potential measures of slack, such as output gaps, the unemployment rate, or capacity utilization. On the basis of data availability and the fact that it is generally accepted as a key measure of underutilized resources, we use the unemployment rate as our baseline 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 follow Owyang et al. (2013) and use 6.5 as the threshold.[16] We also conduct various robustness checks using different thresholds.

[15] Numerous papers explore theoretically the possibility of state-dependent multipliers that depend on alternative states, such as the debt-to-GDP ratio, the condition of the financial system, degree of openness, and exchange rate regimes. For example, see Corsetti, Meier, and Mueller (2012) for a brief survey of this literature, as well as Sims and Wolff (2013) and Canzoneri et al. (2016).

[16] They chose that threshold based on the US Federal Reserve's use of that threshold in its policy announcements at the time. Barro and Redlick (2011) used 5.57 as the threshold, based on the median unemployment rate from 1914 through 2006.

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

Figure 3 shows the unemployment rate, the military spending news shocks, and the estimated Blanchard-Perotti 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

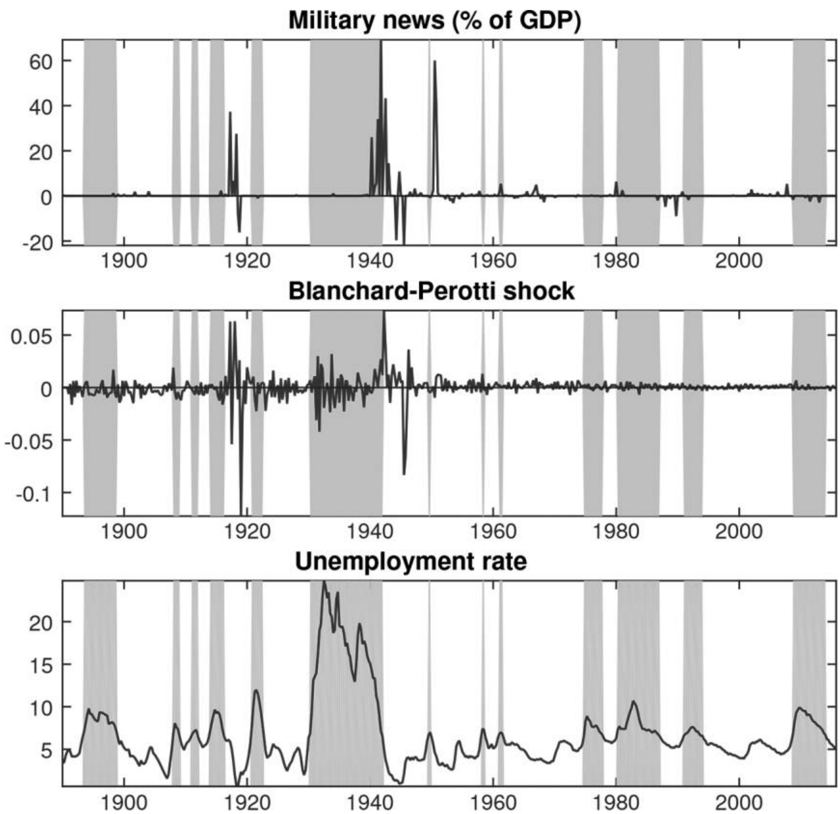

FIG. 3.—Military spending news, Blanchard-Perotti shock, and unemployment rate. Shaded areas indicate periods when the unemployment rate is above the threshold of 6.5 percent. Color version available as an online enhancement.

threshold. In contrast, the initial large news shocks about WWII occurred when the unemployment rate was still very high. The Blanchard-Perotti shocks tend to have large swings around wars. However, they also have substantial volatility at other times. Some of this volatility in the historical periods may be due to measurement error in the constructed government spending series, though.

B. *Instrument Relevance across States of Slack*

As discussed in the last section, multiplier estimates are the outcome of IV regressions. Because the military news variable is based on changes in defense spending due to political events, it should be exogenous to the economy. The question remains, however, whether it is a relevant instrument. The standard rule of thumb is that an *F*-statistic below 10 indicates a potential problem with instrument relevance (Staiger and Stock 1997). However, Olea and Pflueger (2013) show that the threshold can be different, and sometimes higher, when the errors are serially correlated. Since there is inherent serial correlation based on using the Jordà method, we use the Olea and Pflueger effective *F*-statistics and thresholds.[17]

Figure 4 shows the difference between the first-stage effective *F*-statistics and the Olea and Pflueger (2013) thresholds.[18] A value above zero means that the effective *F*-statistic exceeds the threshold. The *F*-statistics are from the regression of the sum of real government spending from t to $t + h$ on the shock(s) at t. The regression also includes all the other controls from the second stage, which include lagged GDP, government spending, and the news variable in the case of military news shock. For the Blanchard-Perotti shock specification, current and lagged military news are not included. The figure shows these for the full historical sample, the historical sample excluding WWII, and the post-WWII sample and splits each of these according to whether the unemployment rate is above 6.5 percent. When we exclude WWII, we exclude observations when either the dependent variable, the shock, or the lagged control variables occur in the period 1941q3–1945q4. Rationing did not start until 1942q1, but Gordon and Krenn (2010) have argued that various other capacity constraints occurred starting the second half of 1941. The results are shown

[17] Even at horizon 0, we detected some serial correlation. Thus, we used automatic bandwidth selection at all horizons.

[18] We use the threshold for the 5 percent critical value for testing the null hypothesis that the two-stage least squares bias exceeds 10 percent of the ordinary least squares bias. For one instrument, this threshold is always 23.1. The threshold is 19.7 percent for the 10 percent critical value. The effective *F*-statistics and thresholds were calculated using Pflueger and Wang (2015), Stata command "weakivtest."

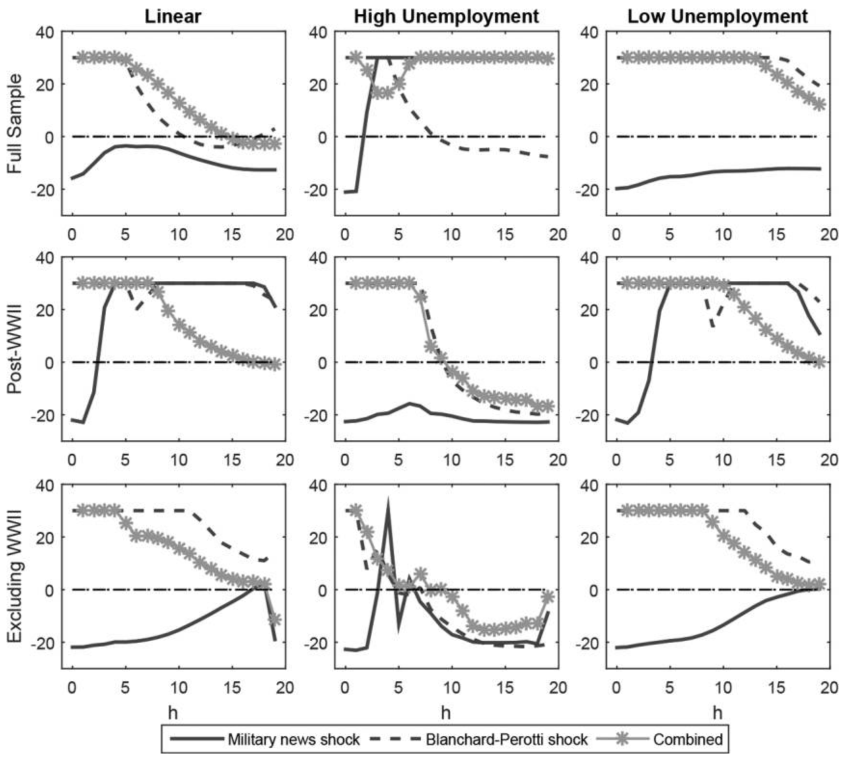

FIG. 4.—Tests of instrument relevance across states of slack. "Slack" is when the unemployment rate exceeds 6.5 percent. The lines show the difference between the effective F-statistic and the relevant threshold for the 5 percent level and are capped at 30. The effective F-statistics are from the regression of the sum of government spending through horizon h on the shock at t and all the other controls from the second stage, separately for the military news variable (solid line), the Blanchard-Perotti shock (dashed line), and both instruments (line with asterisks). The first column shows the linear case, the second column shows the high-unemployment state, and the last column shows the low-unemployment state. The full sample is 1890q1–2015q4, and the post-WWII sample spans 1947q3–2015q4. Color version available as an online enhancement.

for military news as the instrument (solid line), for the Blanchard-Perotti shock as the instrument (dashed line), and for both shocks as instruments.

Several features are evident from figure 4. First, military news has potential relevance problems at very short horizons whereas the Blanchard-Perotti shock has high relevance at very short horizons. These results should be expected because the entire point of Ramey (2011) is that the news about government spending occurs at least several quarters before the government spending actually rises. In contrast, the Blanchard-Perotti shock is identified as the part of current government spending not explained by the other lagged variables control variables. Second, moving beyond the first year or two, the military news shock effective F-statistic often rises above the threshold, whereas the Blanchard-Perotti shock often falls below the threshold.

Since the Blanchard-Perotti shock tends to do well at short horizons and the military news at longer horizons, it is natural to consider using both shocks as instruments. The line with stars in figure 4 shows that when both shocks are used as instruments, the effective F-statistics are above the threshold for more samples and horizons.

Note that none of the instrument alternatives has statistics above the threshold during slack states in the post-WWII period for horizons beyond the 2-year horizon. These results support our initial conjecture that the post-WWII sample is not sufficiently rich to be able to distinguish multipliers across states very precisely.

Using both shocks as instruments may come at a cost of exogeneity, though, since even conditioning on lagged military news, the Blanchard-Perotti shock may be anticipated. Furthermore, the likely measurement error in the historical government spending series will be highly correlated with the Blanchard-Perotti shock, since the shock is equal to the forecast error of government spending. As we shall see, the multiplier estimates that use the Blanchard-Perotti shock are noticeably lower than those estimated using the military news shock, consistent with attenuation bias from measurement error.

Because of possible problems with instrument relevance for some samples and some horizons, we will also conduct some key hypothesis tests using Anderson and Rubin (1949) statistics, which are robust to weak instruments. These tests have lower power, though.

C. Baseline Results for Slack States

We now present the main results of our analysis using the full historical sample and the local projections method. Figure 5 shows the impulse response functions. We first consider results from the linear model, which assumes that multipliers are invariant to the state of the economy. The second column of figure 5 shows the responses of government spending and output to a military news shock in the linear model. The bands are 95 percent confidence bands and are based on Newey-West standard errors that account for serial correlation. After a shock to news, output and government spending begin to rise and then peak at around 12 quarters.

We compute cumulative multipliers for a 2-year and 4-year horizon, using m_h from equation (3). As indicated in the first column of the top panel of table 1, the implied multipliers are around 0.7.

The main question addressed in this paper is whether the multipliers are state dependent and, in particular, whether they are high during

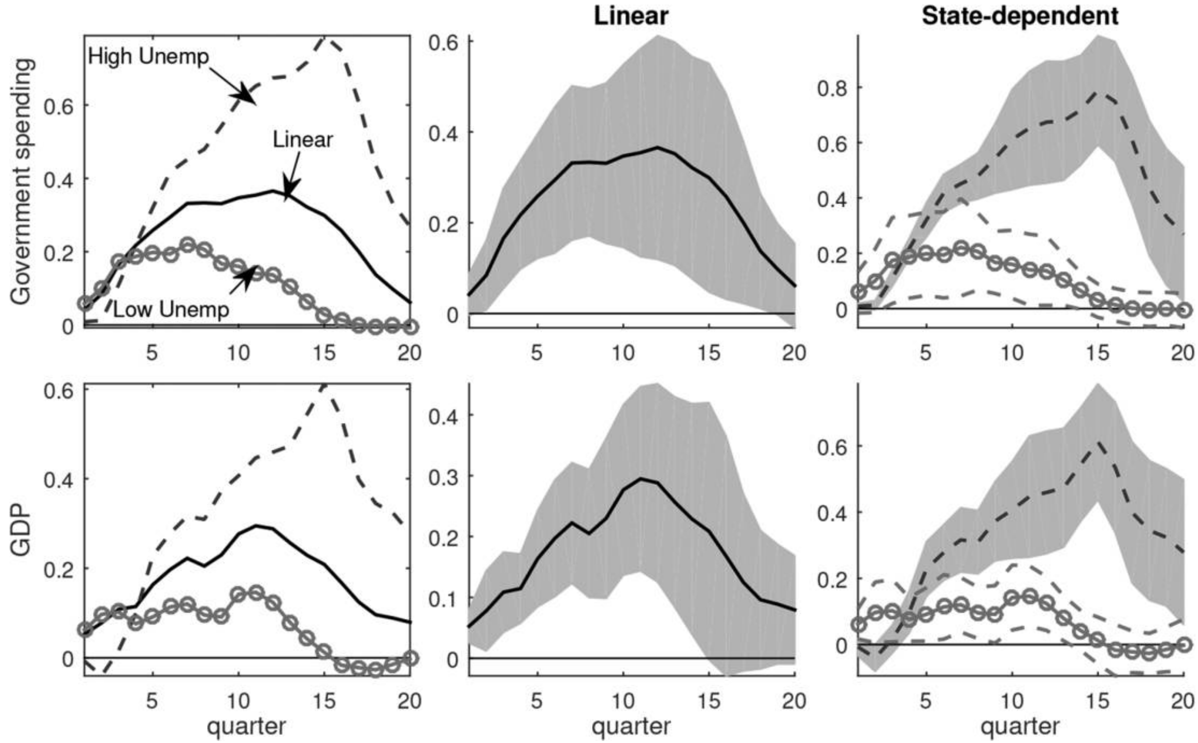

FIG. 5.—Government spending and GDP responses to a news shock: considering slack states. Response of government spending and GDP to a news shock equal to 1 percent of GDP. The top row shows the response of government spending and the second row shows the response of GDP. The first column shows the responses in the linear and state-dependent models. The second column shows the responses in the linear model. The last column 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 percent confidence intervals are shown in the second and third columns. Color version available as an online enhancement.

TABLE 1
Estimates of Multipliers across States of Slack

	Linear Model	High Unemployment	Low Unemployment	p-Value for Difference in Multipliers across States
Military news shock:				
2-year integral	.66	.60	.59	HAC = .954
	(.067)	(.095)	(.091)	Anderson-Rubin = .954
4-year integral	.71	.68	.67	HAC = .924
	(.044)	(.052)	(.121)	Anderson-Rubin = .924
Blanchard-Perotti shock:				
2-year integral	.38	.68	.30	HAC = .005
	(.111)	(.102)	(.111)	Anderson-Rubin = .070
4-year integral	.47	.77	.35	HAC = .001
	(.110)	(.075)	(.107)	Anderson-Rubin = .031
Combined:				
2-year integral	.42	.62	.33	HAC = .099
	(.098)	(.098)	(.110)	Anderson-Rubin = .228
4-year integral	.56	.68	.39	HAC = .021
	(.084)	(.052)	(.110)	Anderson-Rubin = .199

Note.—The values in parentheses under the multipliers give the standard errors. HAC indicates HAC-robust p-values and Anderson-Rubin indicates weak instrument robust Anderson-Rubin p-values.

periods of slack. The impulse response functions 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 last column of figure 5 shows the responses when we estimate the state-dependent model, where we distinguish between periods with and without slack in the economy. Similarly to many preexisting studies (e.g., Auerbach and Gorodnichenko 2012), we find that output responds more robustly during high-unemployment states. However, 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 1, the implied 2- and 4-year multipliers are very similar across the two states, both around 0.6 or 0.7. The final column shows the p-values for the test that the multiplier estimates differ across states. The first p-value reported is based on heteroscedastic- and autocorrelation-consistent (HAC) standard errors and is valid only for strong instruments; the second is based on the Anderson and Rubin (1949) test and is robust to weak instruments.[19] However, it has lower power, so we prefer the HAC-based test for the sample-horizon combinations when the instruments are strong. There is no evidence of differences in multipliers, either quantitatively or statistically.

Figure 6 shows the cumulative multipliers for each horizon from impact to 5 years out.[20] 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): output responds immediately to news about future government spending increases. Since output rises more quickly than government spending, the calculated multiplier looks large. 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.

The second panel of table 1 shows alternative results using the Blanchard-Perotti shock as the instrument.[21] Estimated multipliers are lower in this case, 0.4–0.5 in the linear case. Considering state dependence, multipliers are estimated to be higher in the high-unemployment state, and even the Anderson-Rubin test suggests some differences. However, the estimates imply that multipliers differ across the states not because they are so elevated in high-unemployment states but because they are so low in low-unemployment states. In all cases, they are below unity.

There are two reasons why the Blanchard-Perotti shocks would be expected to yield lower estimates of multipliers. First, as Ramey (2009) shows in DSGE Monte Carlo experiments, if the shocks are anticipated, then the impulse responses will not capture the anticipatory rise in GDP. This results in smaller multipliers. Second, as discussed in Section II.B, there is likely significant measurement error in the government spending series. Since the Blanchard-Perotti shock is defined as the part of government spending not explained by lagged GDP and government spending, it will inherit much of the measurement error. Thus, the measurement error in the instrument will be correlated with the measurement error

[19] We constructed the Anderson-Rubin test conditional on the assumption that there was no instrument relevance problem for the linear term in government spending and then tested the state-dependent term.

[20] We estimate multipliers out only 5 years because the Jordà method is less reliable at long horizons. Thus, we may be neglecting the negative effects due to the eventual increase in distortionary tax, as highlighted by Drautzburg and Uhlig (2015).

[21] In these regressions, lagged news variables are excluded from the controls. The impulse response functions (IRFs) are available in the online supplemental appendix. These IRFs also show both government spending and output responding more during high–unemployment rate states. In contrast to the military news IRFs, government spending rises as soon as the shock hits.

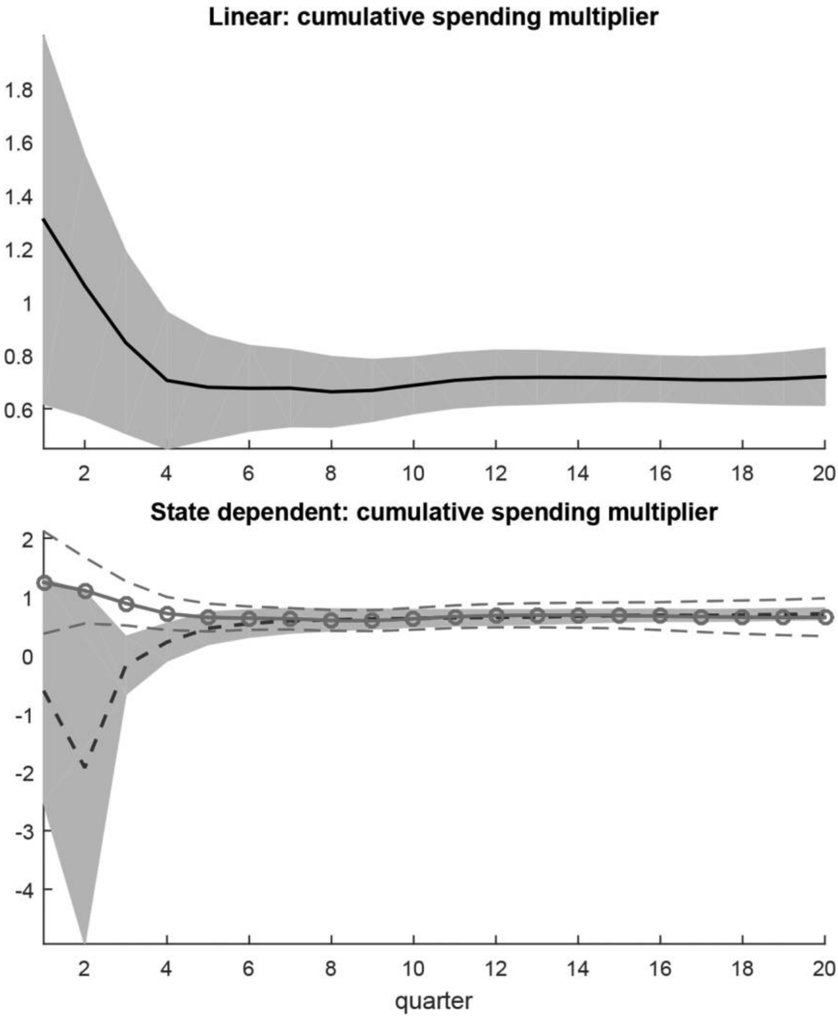

FIG. 6.—Cumulative multipliers for a news shock, considering slack states: cumulative spending multipliers across different horizons to a news shock. The top panel shows the cumulative multipliers in the linear model. The bottom panel shows the state-dependent multipliers, where the blue dashed lines are multipliers in the high-unemployment state and the lines with red circles are multipliers in the low-unemployment state; 95 percent confidence intervals are shown in all cases. Color version available as an online enhancement.

in government spending, so we should expect attenuation bias in the multiplier estimate.

The third panel of table 1 shows the estimated multipliers using both military news and the Blanchard-Perotti shock as instruments. Recall that the combination of instruments had effective F-statistics above the thresholds for all horizons when the full sample was used. The estimates here are closer to those obtained using the Blanchard-Perotti shock alone, with most multipliers lower than those estimated for military news.[22] There is a difference in multipliers using the HAC tests (which are the preferred ones for strong instruments) at the 4-year horizon. Again, though, all multipliers are well below unity.

To summarize, across all three instrument sets we find multipliers that are less than one in all cases (beyond the first couple of quarters). Considering state dependence, we find no evidence of sizable multipliers in the periods of slack; the differences across states for the Blanchard-Perotti shock stem from multipliers being so low during nonslack states.

D. *Robustness of Slack Estimates*

Our baseline 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 begin by conducting robustness checks by changing the definition of the slack state. We first allow for a time-varying threshold, where we consider deviations from trend for a Hodrick-Prescott (HP) filtered unemployment rate.[23] This definition of threshold results in about 50 percent of the observations being above the threshold. As shown in figure 7, 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: multipliers less than one for the state-dependent case, no significant difference between the multipliers when military news is used as the instrument, but some evidence of a difference when the Blanchard-Perotti shock is used (see the first panel of table 2).

Second, we analyze the effect of raising the unemployment rate cutoff for the threshold, to allow for the possibility of state dependence only for a higher degree of slack in the economy. The second panel in table 2 shows that when we choose the threshold for the unemployment rate to be higher than 8 percent, the slack state multiplier rises slightly to 0.8 for military news and to 0.7 for Blanchard-Perotti shocks. Otherwise, the results are similar to the baseline.

[22] A test of overidentifying restrictions using the Hansen *J*-statistic rejects the restrictions in the linear case at all horizons; the *p*-values (not shown in the table) range from .03 to .05. On the other hand, we cannot reject the overidentifying restrictions for the state-dependent model; the *p*-values range from .09 to .17 for nonslack periods and .3 to .9 for slack periods.

[23] We use a very high smoothing parameter of $\lambda = 1{,}000{,}000$; but even with this the Great Depression and WWII have a big influence. Thus, we fit the HP filter over a split sample, 1889–1929 and 1947–2015, and linearly interpolate the small gap in trend unemployment between 1929 and 1947.

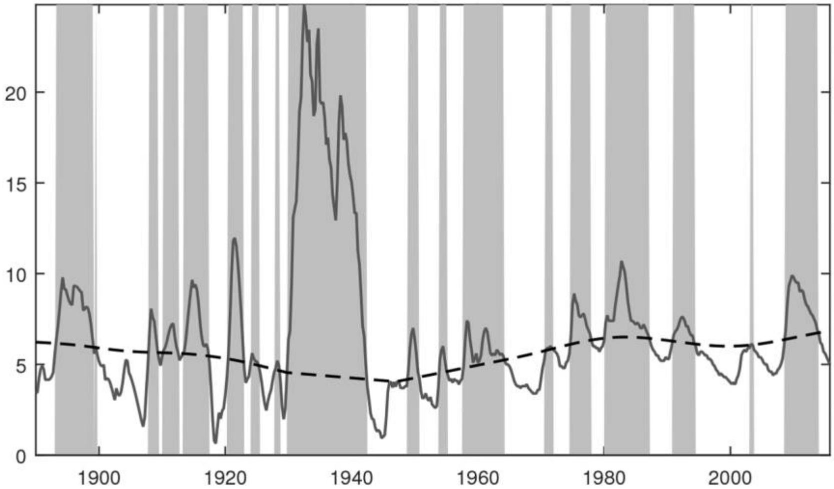

FIG. 7.—Alternative threshold of unemployment rate based on time-varying trend: unemployment rate with a time-varying trend. The solid line is the unemployment rate and the black dashed line shows the time-varying trend based on the HP filter with $\lambda = 10^6$, over a split sample, 1889–1929 and 1947–2015 and linearly interpolated for the small gap in trend unemployment between 1929 and 1947. Shaded areas indicate periods when the unemployment rate is above the time-varying trend. Color version available as an online enhancement.

We also consider NBER recession periods and a smooth transition threshold based on a 7-quarter moving average of output growth, as in Auerbach and Gorodnichenko (2012).[24] Results in the bottom panel of table 2 show that in both cases we still get multipliers less than one across both recession and expansion regimes and do not find any evidence of higher multipliers in recessions versus expansions. In fact, for Blanchard-Perotti shocks, the multipliers are statistically significantly higher in expansions than in NBER recessions.

In order to account for the role of financing, we control for taxes by adding lags of the average tax rate, given by tax revenues as a ratio of GDP, to our specification. The top panel of table 3 shows that our baseline results for both type of shocks are robust to the inclusion of taxes. We have also conducted further analysis considering the role of financing, which is detailed in the online supplemental appendix. This analysis shows that the behavior of deficit and taxes does not seem to explain why multipliers are not higher during times of slack.

We next consider different samples. As discussed in the earlier case study, rationing was a confounding influence during part of WWII. In

[24] We use the same definition as in Auerbach and Gorodnichenko (2012), and the online supplemental appendix shows this smooth transition function for our historical sample.

TABLE 2
Robustness Check: Estimates of Multipliers across States of Slack

	Linear Model	High Unemployment	Low Unemployment
	HP-Filtered Time-Varying Threshold (with $\lambda = 10^6$)		
Military news shock:			
2-year integral	.66	.52	.66
4-year integral	.71	.56	.75
Blanchard-Perotti shock:			
2-year integral	.38	.58	.29†
4-year integral	.47	.69	.31†
	8% Unemployment Rate Threshold		
Military news shock:			
2-year integral	.66	.80	.60
4-year integral	.71	.76	.65
Blanchard-Perotti shock:			
2-year integral	.38	.64	.36†
4-year integral	.47	.69	.43†
	Linear	Recession	Expansion
	NBER Recession Dates		
Military news shock:			
2-year integral	.66	.63	.55
4-year integral	.71	.67	.64
Blanchard-Perotti shock:			
2-year integral	.38	.15	.50†
4-year integral	.47	.25	.58†
	Moving Average of Output Growth Weighting Function		
Military news shock:			
2-year integral	.66	.57	.62
4-year integral	.71	.65	.68
Blanchard-Perotti shock:			
2-year integral	.38	.51	.39
4-year integral	.47	.57	.52

† HAC-robust p-value for the difference in multipliers across states: $p_{\mathrm{HAC}} < .1$.

order to determine whether our results are sensitive to the constraints or the rationing, we exclude WWII from our sample.[25] Recall from Section IV.B, though, that all instrument sets appear to be weak for the high–unemployment rate state for horizons beyond 2 years if this period is excluded. The third panel of table 3 shows that multipliers rise to around 1 and are even 1.6 in the case of Blanchard-Perotti shocks at the 4-year horizon.[26] However, the confidence bands are so large that there is no

[25] See Sec. IV.B for details on how we exclude WWII from our sample.

[26] Exclusion of WWII and the use of military news shocks is the one instance in which the slight changes in sample make a difference in the multipliers calculated by summing IRFs vs. estimating things using the one-step method. The results shown for the sample excluding WWII are based on summing the IRFs.

TABLE 3
Robustness Check: Estimates of Multipliers across States of Slack

	Linear Model	High Unemployment	Low Unemployment
	Additional Control for Taxes		
Military news shock:			
2-year integral	.66	.67	.54
4-year integral	.72	.69	.60
Blanchard-Perotti shock:			
2-year integral	.37	.71	.35†
4-year integral	.45	.80	.39†*
	Excluding WWII		
Military news shock:			
2-year integral	.75	.72	.56
4-year integral	.73	.89	.53
Blanchard-Perotti shock:			
2-year integral	.14	.98	.13
4-year integral	.17	1.62	.18
	Subsample: 1947–2015		
Military news shock:			
2-year integral	.75	−1.63	.80†
4-year integral	.51	−2.77	.49
Blanchard-Perotti shock:			
2-year integral	.31	−.47	.39†
4-year integral	.32	−.44	.34

* Weak instrument robust p-value for the difference in multipliers across states: $p_{\mathrm{AR}} < .1$.
† HAC-robust p-value for the difference in multipliers across states: $p_{\mathrm{HAC}} < .1$.

statistically significant difference (even at the 10 percent level) between multipliers across the two states.[27]

The preexisting literature on state dependence of multipliers typically employs a shorter data sample that spans the post-WWII period (see, e.g., Auerbach and Gorodnichenko 2012; Bachmann and Sims 2012; Caggiano et al. 2015; Riera-Crichton, Vegh, and Vuletin 2015). As a robustness check we limit our sample to this period, 1947–2015, and the results are shown in the bottom panel of table 3. The multipliers in the linear case are similar to those in the full historical sample. In the state-dependent case the multipliers are estimated to be negative in the high–unemployment rate states, for both the military news shock and the Blanchard-Perotti shock. In both cases, the impulse response of GDP is negative at most horizons, but even the HAC-based standard error bands are very wide (not shown). However, recall that neither instrument was strong for the

[27] The confidence bands are not shown here. The Blanchard-Perotti multiplier estimate at the 4-year horizon of 1.6 during recessions has a HAC standard error above 1.9, so the estimate is not even significantly different from zero.

high–unemployment rate state in the post-WWII sample. Thus, the state-dependent estimates are not reliable.

We also conducted a number of other robustness checks, such as using data based on linear interpolation and including additional controls. The results are available in the online supplemental appendix.

V. 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. Some New Keynesian models suggest that government spending multipliers will be substantially higher (e.g., above 2) when the economy is at the zero lower bound.[28] This view has been challenged by a series of new papers, some of which construct models in which multipliers are lower at the zero lower bound (see, e.g., Aruoba and Schorfheide 2013; Braun, Korber, and Waki 2013; Kiley 2014; Mertens and Ravn 2014). 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.

Very few papers have attempted to test the predictions of the theory empirically in aggregate data. Ramey (2011) estimates her model for the United States over the subsample from 1939 through 1951 and shows that the multiplier is no higher during that sample. Crafts and Mills (2013) construct defense news shocks for the United Kingdom and estimate multipliers on quarterly data from 1922 through 1938. They find multipliers below unity even when interest rates were near zero.[29]

A. *Defining States by Monetary Policy*

The bottom panel of figure 8 shows the behavior of 3-month Treasury bill rates from 1920 through the present, where the shortened sample is based on data availability, as well as the discount rate for the period starting in 1914 until 1919 (dotted line) at the founding of the Fed. The Treasury bill 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 (by either design

[28] See, e.g., Christiano et al. (2011) and Eggertsson (2011). The relationship between government spending multipliers and the degree of monetary accommodation, even outside the zero lower bound, has been explored by many others, including Davig and Leeper (2011) and Zubairy (2014).

[29] Bruckner and Tuladhar (2014) focus on local, not aggregate, multipliers for Japan and find that the effects of local spending are larger in the zero lower bound (ZLB) period, but only modestly. A recent paper by Miyamoto, Nguyen, and Sergeyev (forthcoming) extends our analysis to Japan and finds some evidence of higher multipliers at the ZLB.

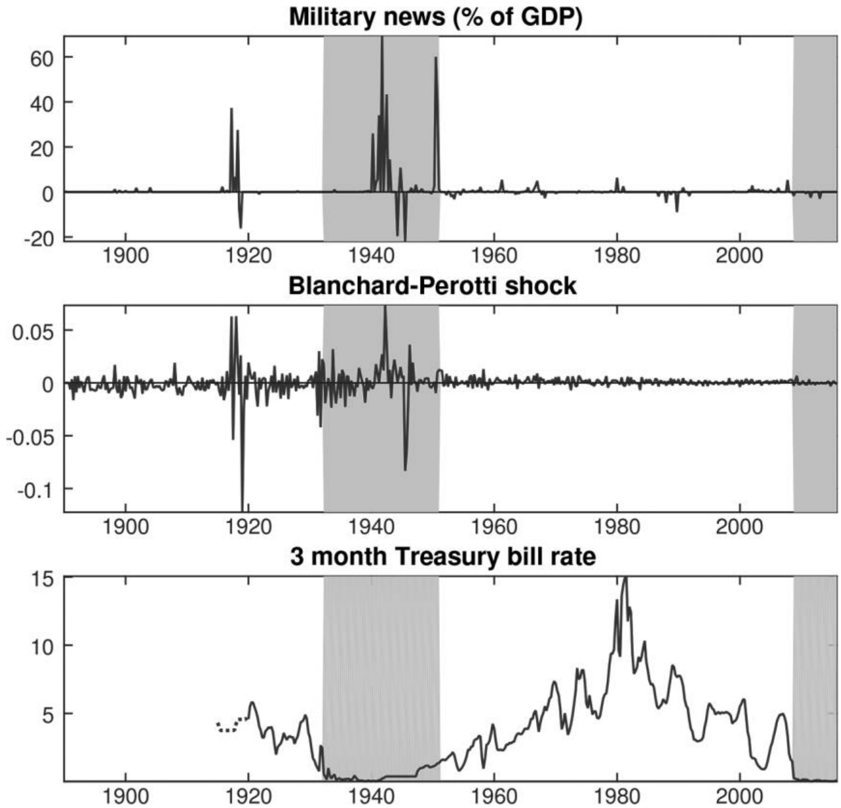

FIG. 8.—Military spending news, Blanchard-Perotti shock, and interest rate. Shaded areas indicate periods that we classify as the zero lower bound period for interest rate. Color version available as an online enhancement.

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:

$$\begin{aligned}\text{nominal interest rate} = 1 &+ 1.5 \text{ year-over-year inflation rate} \\ &+ 0.5 \text{ output gap}.\end{aligned} \tag{5}$$

Figure 9 shows the behavior of inflation and the output gap, which were quite volatile during the early period.[30] The last panel of figure 9 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 at the start of the sample in 1914

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

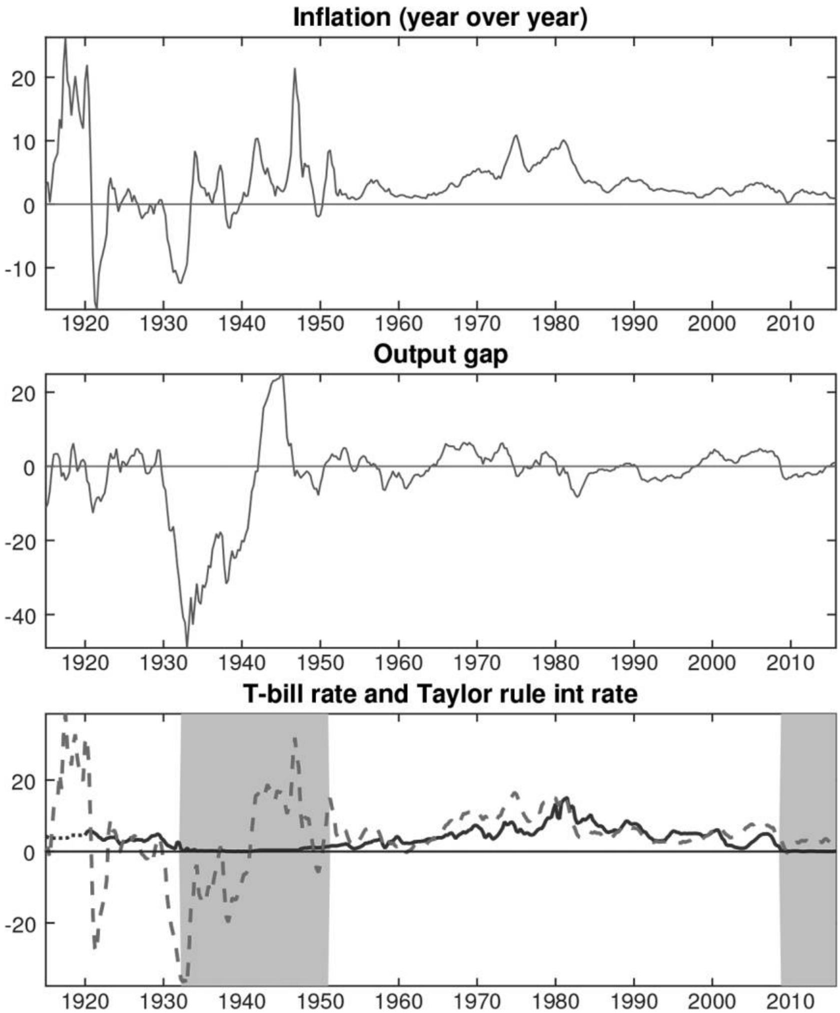

FIG. 9.—Inflation, output gap, and Taylor rule implied interest rate. The top panel shows the year-over-year GDP deflator inflation rate and the second panel shows the output gap, which is constructed as the percentage deviation between real GDP and potential GDP. In the last panel, the dashed line shows the Taylor rule implied nominal interest rate, and the solid line shows the data for the 3-month T-bill rate, with a dotted line for 1914–19 showing the discount rate. In the last panel, the shaded areas indicate periods that we classify as the zero lower bound period for interest rate. Color version available as an online enhancement.

into 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 2008q4–2015q4 (the end of our sample). We do not classify the early part of the sample as a ZLB episode, since the United States was under the gold standard then with the purpose of ensuring price stability. The United States maintained the gold standard only in a limited sense starting in 1914, at the onset of WWI, but did not completely suspend it (see Crabbe 1989). Thus, any actual inflation would have to be offset by future deflation and we would not expect high multipliers based on the expectations channel, as long as people in the economy expected to go back on the gold standard with the end of the war.

Also, while the deviation from the Taylor rule widens starting in 1930, we do not include the early 1930s in our ZLB state. The reason is that 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 onward. 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 8 shows the behavior of the military news series and the Blanchard-Perotti shock over the states defined this way. The main shocks to military spending news 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 military news during the 1930s.[31] There are sizable Blanchard-Perotti shocks during that period, though.

B. *Instrument Relevance for ZLB Periods*

Figure 10 shows the difference between the effective *F*-statistics and the thresholds for the periods split into ZLB periods and normal periods and for the defense news shock, Blanchard-Perotti shock, and the combined instruments.[32] For the ZLB periods for the full sample, military news just reaches the threshold from horizon 4 through around 8, whereas the Blanchard-Perotti shock instrument and the combined instruments have strong relevance through horizon 15. If WWII rationing

[31] An advantage of the Crafts and Mills (2013) analysis of UK data is that it has more military news shocks during the 1930s.

[32] See the earlier discussion on instrument relevance in Sec. IV.B for details about the tests and thresholds. That section also discusses the potential problems with using the Blanchard-Perotti shock.

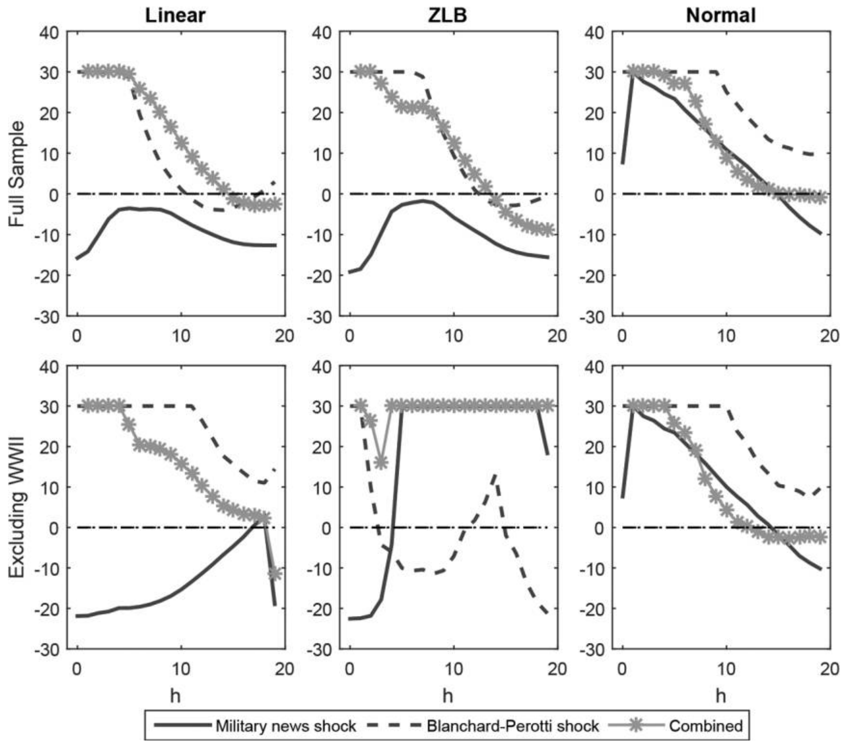

FIG. 10.—Tests of instrument relevance across monetary policy regimes. "ZLB" is when interest rates are near the zero lower bound or the Fed is being very accommodative of fiscal policy (1932q1–1951q1, 2008q4–2015q4). The lines show the difference between the effective F-statistic and the relevant threshold for the 5 percent level and are capped at 30. The effective F-statistics are from the regression of the sum of government spending through horizon h on the shock at t and all the other controls from the second stage, separately for the military news variable (solid line), the Blanchard-Perotti shock (dashed line), and both instruments (line with asterisks). The first column shows the linear case, the second column shows the high-unemployment state, and the last column shows the low-unemployment state. The full sample is 1890q1–2015q4. Color version available as an online enhancement.

periods are excluded, the Blanchard-Perotti shock instrument loses relevance after just a few horizons, but the military news and the combined instruments have higher effective F-statistics for most horizons for both states. Thus, unlike the case for slack, military news, as well as both instruments combined, appear to be strong instruments in the ZLB even when WWII rationing is omitted. We suspect that the reason that the F-statistics actually rise relative to the full sample is that the observations omitted may have represented cases in which the military news did not predict the actual path of government spending well.[33]

[33] For example, the D-Day invasion in June 1944 led the public to believe the war in Europe would be over in just a couple of months, which turned out to be wrong.

It is important to note that these effective F-statistic results depend heavily on our standard procedure of allowing the sample to change as the horizon advances. To be specific, for the full sample with the WWII rationing periods excluded, as we go from horizon h to $h + 1$, we drop two observations, one in the late 1930s or early 1940s and another near the end of the sample in the 2010s. Dropping the extra observation in the 2010s makes no difference, but sometimes dropping an observation in the late 1930s or early 1940s does make a difference because it means dropping a large military news shock. We considered fixing the sample at the maximum horizon of 20 quarters, but that involves throwing away all observations for the 10-year period from 1936q3 through 1946q4. Not surprisingly, the F-statistics for military news and the combined instruments are far below the threshold at virtually all horizons if we discard the information during the entire 10-year period (not shown).

C. Results for ZLB States

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 rather than slack. We consider our full sample spanning 1889–2015. Figure 11 shows the impulse responses. The results suggest that government spending responds more slowly but more persistently during ZLB states than in normal states.[34] The difference in GDP responses follows this pattern, but in a muted way.

Table 4 shows the cumulative multipliers in each state for the different horizons of 2 and 4 years, respectively. Using military news, we see little difference in multipliers in the ZLB state. Figure 12 shows the cumulative multiplier for the ZLB and normal states at various different horizons along with 95 percent confidence bands. The multiplier for both states is high on impact when the news shock hits the economy (since the shock is news about future government spending) and is less than one after 1 year, but the multipliers across the two states are never significantly different. For the Blanchard-Perotti shock and the combined shocks, the multipliers are estimated to be 0.64–0.76 in the ZLB state but only 0.1–0.26 in the normal (non-ZLB) state (see the middle and third panels of table 4). There is also statistical evidence of differences in multipliers, as evidenced by the p-values; we reference the HAC-based tests since the instruments appear to be strong. However, this difference is due not to elevated multipliers in the ZLB but to multipliers estimated to be near zero in the normal states.

[34] This result stems from the particular historical sample and is not necessarily a general result. In particular, the two large wars that resulted in persistent increases in government spending—WWII and the Korean War—occurred during the ZLB period. World War I, which involved less persistent increases in government spending, occurred in the non-ZLB, or normal, period.

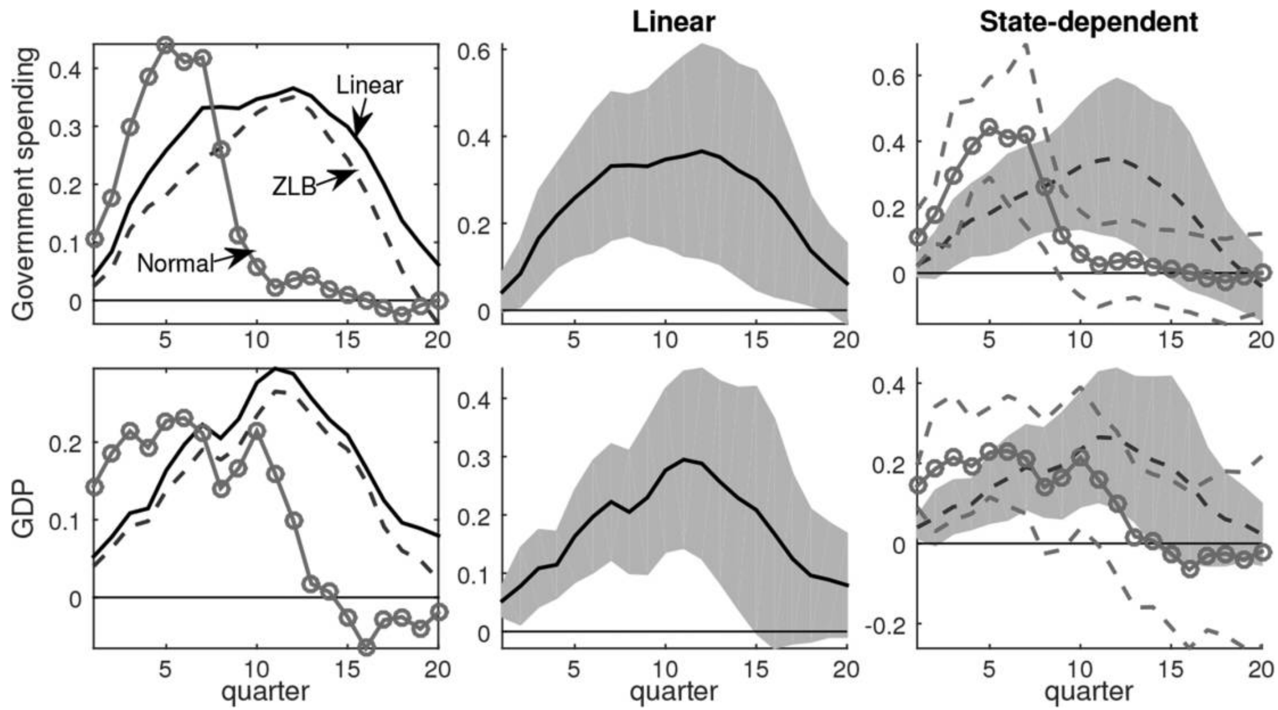

FIG. 11.—Government spending and GDP responses to a news shock: considering zero lower bound. Response of government spending and GDP to a news shock equal to 1 percent of GDP. The top row shows the response of government spending and the second row shows the response of GDP. The first column shows the responses in the linear and state-dependent models. The second column shows the responses in the linear model. The last column 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 percent confidence intervals are shown in the second and third columns. Color version available as an online enhancement.

TABLE 4
Estimates of Multipliers across Monetary Policy Regimes

Baseline	Linear Model	Near Zero Lower Bond	Normal	p-Value for Difference in Multipliers across States
Military news shock:				
2-year integral	.66	.77	.63	HAC = .429
	(.067)	(.106)	(.149)	Anderson-Rubin = .504
4-year integral	.71	.77	.77	HAC = .992
	(.044)	(.058)	(.376)	Anderson-Rubin = .992
Blanchard-Perotti shock:				
2-year integral	.38	.64	.10	HAC = .000
	(.111)	(.033)	(.112)	Anderson-Rubin = .066
4-year integral	.47	.71	.12	HAC = .000
	(.110)	(.033)	(.115)	Anderson-Rubin = .062
Combined:				
2-year integral	.42	.67	.26	HAC = .000
	(.098)	(.027)	(.103)	Anderson-Rubin = .184
4-year integral	.56	.76	.21	HAC = .000
	(.084)	(.040)	(.136)	Anderson-Rubin = .208

Note.—The values in parentheses under the multipliers give the standard errors. HAC indicates HAC-robust p-values and Anderson-Rubin indicates weak instrument robust Anderson-Rubin p-values.

Table 5 shows various robustness checks. These robustness checks include redefining the ZLB state periods to be those in which the T-bill rate was less than or equal to 50 basis points and including taxes and inflation as additional controls. The results show that our baseline estimates are robust to these modifications. (See the online supplemental appendix for some additional robustness checks.)

We then explore the effect of excluding the capacity constraint and rationing periods of WWII, excluding observations from the estimation if either the shock, the dependent variable, or the lagged controls occurred in any quarter from 1941q3 through 1945q4. Table 6 shows the estimates.[35] For the first time, we see evidence of multipliers above unity in a "bad" state, in this case the ZLB state. Using military news as an instrument, the multiplier is estimated to be 1.4 at 2 years and close to 1 at 4 years in the ZLB state. The Blanchard-Perotti instrument also produces higher multiplier estimates in the ZLB state, though they are still below unity and they have very large standard errors. The multiplier estimates based on using both instruments, shown in the lower panel of table 6, imply a multiplier of 1.6 at the 2-year horizon and 1.1 at the 4-year hori-

[35] As in the case of slack, the excluded WWII sample along with the news shock leads to some differences across the three-step and one-step methods because of some influential observations in the changing sample. The differences are smaller for the ZLB analysis, with the greatest differences being 0.2. We report the one-step estimates because they allow us also to use the combined instruments.

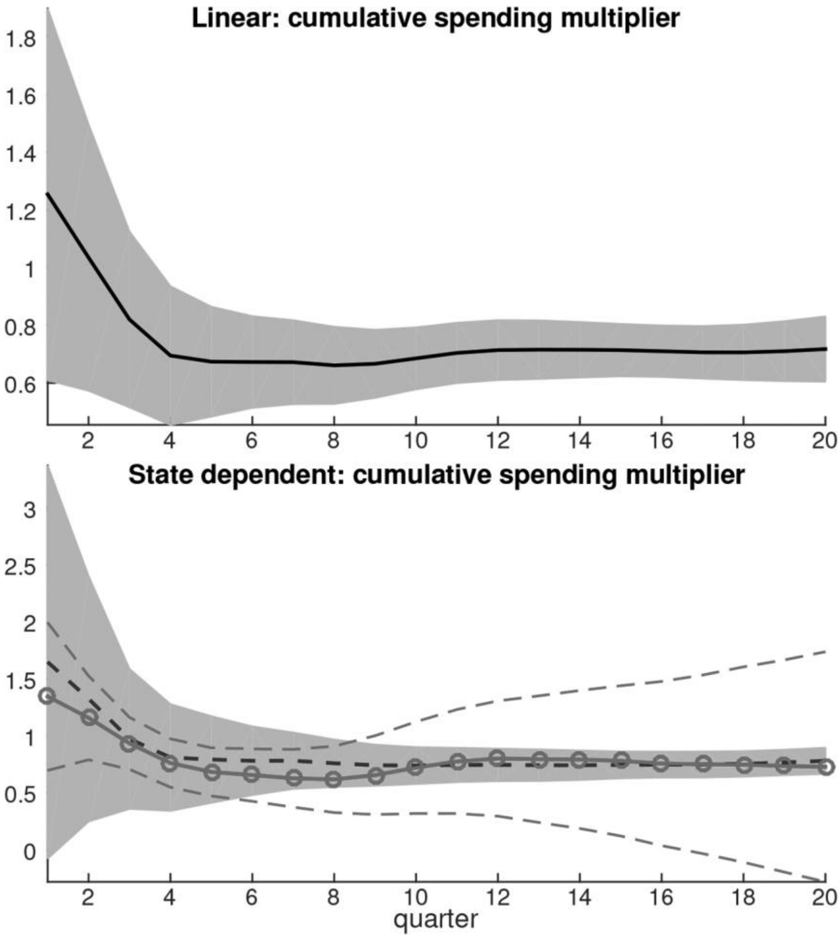

FIG. 12.—Cumulative multipliers for a news shock: considering zero lower bound; cumulative spending multipliers across different horizons for a news shock. The top panel shows the cumulative multipliers in the linear model. The bottom panel shows the state-dependent multipliers, where the blue dashed lines are multipliers in the near zero lower bound state and the lines with red circles are multipliers in the normal state; 95 percent confidence intervals are shown in all cases. Color version available as an online enhancement.

zon. We can reject equality of the multipliers across states using the HAC-based test, but not the Anderson-Rubin test. Since the *F*-statistics are above the threshold, we prefer the HAC-based test since it has better power.[36]

[36] We also tested the overidentifying restrictions when we use the two instruments together. As discussed in a previous section, the overidentifying restrictions are rejected for the linear case. We cannot, however, reject them for the ZLB and normal states; the *p*-values for the ZLB states range from .2 to .8, depending on the horizon, and for the normal states range from .14 to .27.

TABLE 5
Robustness Checks: Estimates of Multipliers across Monetary Policy Regimes

	Linear Model	Near Zero Lower Bound	Normal
	Defining ZLB as T-Bill Rate ≤ .5		
Military news shock:			
2-year integral	.66	.66	.78
4-year integral	.71	.74	.76
Blanchard-Perotti shock:			
2-year integral	.38	.63	.16†
4-year integral	.47	.70	.20†
	Additional Controls for Taxes and Inflation		
Military news shock:			
2-year integral	.67	.94	.55
4-year integral	.71	.86	.52
Blanchard-Perotti shock:			
2-year integral	.38	.67	.08†
4-year integral	.44	.74	−.02†

† HAC-robust p-value for the difference in multipliers across states: $p_{\text{HAC}} < .1$.

In sum, when we consider the sample that excludes the rationing in WWII, we find multipliers above unity at some horizons when we use the military news shock. That shock used alone produces a multiplier estimate of 1.4, with a HAC standard error of 0.15, which is statistically different from the one for normal times, estimated to be 0.6. Thus, in this restricted sample we find both a multiplier above unity during ZLB periods and a difference with normal periods. The online supplemental appendix shows that the multiplier estimates are above unity at the 2-year horizon for the military news shock when we control for taxes and inflation. The estimate during ZLB periods is 1.7, but it is estimated less precisely.

VI. A Comparison of Methodologies and Estimates

In Section IV.C we found no evidence of elevated multipliers during slack states. This result is consistent with results of Barro and Redlick (2011), who also find no differences in contemporaneous multipliers across states of slack. This finding stands in contrast, however, to the leading study of state dependence for the United States by Auerbach and Gorodnichenko (2012), who report multipliers as high as 2.5 for their definition of the recession state.

In this section, we show that the difference in results with those in Auerbach and Gorodnichenko (2012) is largely driven by the simplifying

TABLE 6
Estimates of Multipliers across Monetary Policy Regimes:
Excluding World War II

	Linear Model	Near Zero Lower Bound	Normal	p-Value for Difference in Multipliers across States
Military news shock:				
2-year integral	.77	1.40	.63	HAC = .000
	(.201)	(.153)	(.152)	Anderson-Rubin = .263
4-year integral	.74	.98	.77	HAC = .585
	(.158)	(.100)	(.375)	Anderson-Rubin = .637
Blanchard-Perotti shock:				
2-year integral	.13	1.08	.10	HAC = .197
	(.080)	(.749)	(.101)	Anderson-Rubin = .301
4-year integral	.15	.84	.12	HAC = .228
	(.093)	(.574)	(.115)	Anderson-Rubin = .416
Combined:				
2-year integral	.21	1.60	.26	HAC = .010
	(.087)	(.507)	(.103)	Anderson-Rubin = .216
4-year integral	.26	1.10	.21	HAC = .001
	(.105)	(.233)	(.136)	Anderson-Rubin = .354

Note.—The values in parentheses under the multipliers give the standard errors. HAC indicates HAC-robust p-values and Anderson-Rubin indicates weak instrument robust Anderson-Rubin p-values.

assumptions about state transitions that they use to convert their smooth transition VAR (STVAR) estimates into impulse responses. We first explain the implicit assumptions embedded in the Jordà method and compare them to the assumptions used by Auerbach and Gorodnichenko. We next show that making their assumptions more consistent with their data-generating process significantly reduces their recession-state multiplier estimates. Finally, we apply a threshold VAR (TVAR) method to our historical data in a way that is more consistent with the data-generating process and show that the estimates are very similar to those we obtained using the Jordà method.

A. *Methodological Differences with Auerbach and Gorodnichenko (2012)*

The key ingredients for estimating multipliers in a dynamic environment are the impulse responses of output and government spending. Constructing impulse responses in nonlinear VAR models is far from straightforward since many complexities arise when one moves from linear to nonlinear systems (e.g., Koop, Pesaran, and Potter 1996). 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 how the shocks affect the state, are key components of the constructed responses.

As discussed in Section III.A, the Jordà method is similar to a direct forecasting method. The impulse response estimate for GDP at $t + h$ is a forecast of how GDP will differ at $t + h$ if $\text{shock}_t = 1$ rather than $\text{shock}_t = 0$. This means that if the average shock is likely to change the state, it will be reflected in the impulse response estimate. On the other hand, natural transitions between states that are independent of the shock should be captured by the state-dependent control variables; that is, the coefficients on the state-dependent (and horizon-specific) constant terms and lagged variables will embed information on the average behavior of the economy to transition to the other state at future horizons.

In contrast, Auerbach and Gorodnichenko (2012) estimate a regime-switching VAR model, which switches between one set of reduced-form VAR parameters for recessions and another set for expansions.[37] The difficulty comes in generating impulse responses from those parameters because one must make assumptions about when the parameter sets should switch from one state to the other. Auerbach and Gorodnichenko calculate their baseline impulse responses under the assumption that the economy stays in its current state for at least the 20 quarters over which they compute their multiplier. This may be a reasonable approximation for expansions, which last for several years, but it is not a good approximation for recession states, which have a mean duration of only 3.3 quarters, according to their moving average of growth rates definition.[38] In fact, Hamilton (1989) has argued that GDP is well described by a regime-switching model with a short-duration low-growth regime ("recession") and a longer-duration high-growth regime ("expansion"). Auerbach and Gorodnichenko estimate the 5-year multipliers in recessions to be 2.24, but this high recession multiplier is not due to differences in impact effects on output, for those are estimated to be equal across states (around 0.5). Rather, their high multiplier stems from their constructed impulse response for the subsequent path of GDP after a shock hits in a recession state. As the bottom-left graph of figure 2 of their paper (also reproduced in our online supplemental appendix) shows, their constructed impulse response for GDP keeps rising indefinitely after a spending shock hits

[37] They assume that the transitions across states are smooth and the indicator function of the state of the economy varies between a maximum of one (extreme recession) and zero (extreme expansion).

[38] Even the Great Recession, which was not part of their estimation sample, lasted only 9 quarters by their definition. And the 9-quarter duration is an overestimate, since Auerbach and Gorodnichenko (2012) use only extreme recessions in their calculations.

during a recession, even though government spending does not continue to rise. A regime-switching model for GDP provides a ready explanation for their unusual response of output: on average, recessions do not last long, so during a recession one should forecast output growth in the next few years to be higher than current output growth. Because their method assumes that the economy continues in recession indefinitely, it looks like future growth will always be higher than current growth. With their method, the multiplier grows as the horizon grows since output keeps rising but government spending does not.

To show how the methodology changes the multiplier estimate, we first use the Jordà method on Auerbach and Gorodnichenko's exact data, sample, and identification scheme. When we do so, we estimate a 5-year cumulative multiplier of 0.84.[39] Thus, the Jordà method does not produce elevated multipliers during recessions on the Auerbach-Gorodnichenko sample and specification. In the online supplemental appendix, we show that the difference between the two methods is largely due to Auerbach and Gorodnichenko's baseline assumption that the recession state VAR parameters should apply for a 20-quarter period. We use their STVAR parameter estimates to construct alternative impulse responses that allow the state to change endogenously, with respect to the history of both the nongovernment spending shocks and the government spending shocks. We find multipliers in severe recessions that are around unity. Thus, their high estimated multiplier in recessions disappears when we allow more data-consistent transitions from state to state.[40]

B. *TVAR Estimation on the Historical Data*

The TVAR method is not intrinsically problematic because one can vary the assumptions when translating the reduced-form TVAR estimates to IRFs.[41] Also, alternative definitions of states or alternative samples may be consistent with the assumption of nonchanging states through rea-

[39] We find, however, that these results are not robust since almost any deviation from their exact specification results in negative multipliers during recessions. For example, if we omit the four lags of the moving average of GDP growth, use a backward-moving average of growth for the state, or use four instead of three lags of the endogenous variables, the results change significantly. Alloza (2014) conducts a much more systematic analysis of the importance of the two-sided moving average filter and shows that the results are not robust in the STVAR either.

[40] In later work, Auerbach and Gorodnichenko (2013) use the Jordà method on OECD data. Our online supplemental appendix explains that it is likely that the manner in which they converted IRFs to multipliers raised their multiplier estimates substantially.

[41] VARs with various methodologies to construct state-dependent IRFs and fiscal multipliers have been employed by Batini, Callegari, and Melina (2012), Baum, Poplawski-Ribeiro, and Weber (2012), and Fazzari et al. (2015). We explore the source of difference between our results and those of Fazzari et al. in the online supplemental appendix.

sonable horizons. For example, ZLB states tend to last many years, so holding the state constant is not contrary to the data.

In this section, we apply the TVAR methodology along with Auerbach and Gorodnichenko's baseline assumptions about the duration of states to the ZLB case. We find estimates that are exceedingly close to those we obtained using the Jordà method. We also explore the effect of applying their method to recession states and also find results similar to those we obtain using the Jordà method.

In particular, we consider the following reduced-form TVAR,

$$Y_t = I_{t-1}\Psi_A(L)Y_{t-1} + (1 - I_{t-1})\Psi_B(L)Y_{t-1} + u_t, \tag{6}$$

where, as before, I is a dummy variable that indicates the state of the economy when the shock hits and $u_t \sim N(0, \Omega)$. We also assume that $\Omega = I_{t-1}\Omega_A + (1 - I_{t-1})\Omega_B$, and $\Psi(L)$ is a polynomial of order 4. In order to identify a military news shock we set $Y_t = [\text{news}_t, g_t, y_t]$, and in order to identify a Blanchard-Perotti shock, we set $Y_t = [g_t, y_t]$ before doing a Cholesky decomposition. Here y_t and g_t are Gordon-Krenn transformations of output and government spending, respectively.

First, we define the state on the basis of whether the interest rates are subject to ZLB. In our full sample, we have classified two episodes as ZLB or extended monetary accommodation times, and both have a long duration. In fact, average duration of a ZLB period is about 52 quarters. Thus, in this case the assumption that the state lasts several years is data consistent even if we compute 5-year multipliers, since an average news or spending shock is not likely to cause the economy to leave its current state.

Table 7 shows the state-dependent multipliers for the ZLB state for both the Jordà method and the threshold VAR, assuming that the economy does not exit from its current state. The two methodologies give surprisingly similar results, for both the military news and Blanchard-Perotti identification, with multipliers between 0.6 and 0.8 in the ZLB state for the full historical sample.[42] Thus, when the TVAR method and Auerbach and Gorodnichenko's assumptions about constant states are applied to samples and horizons over which those assumptions are more consistent, the results look very much like those from the Jordà method. Examination of the impulse responses constructed from the TVAR (shown in the online supplemental appendix) reveals that the response of output during the ZLB state has the more usual hump shape, in contrast to Auerbach and Gorodnichenko's ever-increasing path of output.

Although slack periods do not last as long as ZLB periods, even in the historical sample, it is still interesting to compare the estimates from the

[42] Recall that the multipliers were higher when we excluded WWII rationing. We tried to estimate the TVAR on that restricted sample, but the roots were explosive.

TABLE 7
Estimates of Multipliers from Threshold VAR and Jordà Method on the Historical Data

	ZLB Monetary Policy Regime		
	Linear	ZLB	Normal
	Threshold VAR		
Military news shock:			
2-year integral	.61	.75	.43
4-year integral	.64	.83	.31
Blanchard-Perotti shock:			
2-year integral	.36	.62	.03
4-year integral	.40	.70	.07
	Jordà Method		
Military news shock:			
2-year integral	.66	.77	.63
4-year integral	.71	.77	.77
Blanchard-Perotti shock:			
2-year integral	.38	.64	.10
4-year integral	.47	.71	.12
	NBER Recession Dates		
	Linear	Recession	Expansion
	Threshold VAR		
Military news shock:			
2-year integral	.61	.59	.52
4-year integral	.64	.57	.52
Blanchard-Perotti shock:			
2-year integral	.36	−.20	.37
4-year integral	.40	−.25	.39
	Jordà Method		
Military news shock:			
2-year integral	.66	.63	.55
4-year integral	.71	.67	.64
Blanchard-Perotti shock:			
2-year integral	.38	.15	.50
4-year integral	.47	.25	.58

TVAR to those from the Jordà method in that sample. We initially estimated the TVAR for our baseline definition of slack states, but the roots were explosive. Thus, we explored the alternative of defining the state to be official NBER recessions. The longest recession in the post-WWII sample lasted 6 quarters, but in the historical sample there were four recessions that lasted 8 quarters or more. Table 7 shows both the Jordà estimates and the estimates from the TVAR, assuming the economy does not switch states. For the case of the military news shock, the multipliers are remarkably similar across the two approaches. For the Blanchard-Perotti shock, the TVAR multiplier estimates imply negative multipliers in recession and multipliers close to 0.4 in expansions. Thus, the TVAR approach also does not reveal any multiplier larger than one or provide any evidence of larger multipliers in recessions than in expansion.

Overall, the TVAR approach using the constant-state assumptions yields results very similar to our baseline estimates using the Jordà approach. We find much more similarity in the two methods in our application because all ZLB states last many quarters, and in our historical sample even the recession states last longer. Moreover, our small multiplier estimates are more consistent with Auerbach and Gorodnichenko's baseline assumption that the government spending shock cannot make the economy switch states. Thus, their simplifying assumption is a better approximation to the data in our application than in theirs.

VII. 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 United States. We considered two possible indicators of the state of the economy: the amount of slack, as measured by the unemployment rate, and whether interest rates were being held constant close to the zero lower bound. Using a more data-consistent 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 for slack states can be summarized as follows. We find no evidence of large multipliers when the US economy is experiencing substantial slack as measured by the unemployment rate. All estimates indicate multipliers below unity. When we use the Blanchard-Perotti shock identification, we find differences in multipliers across states of slack, but only because the multipliers are very low in nonslack states. Our numerous robustness checks suggest that our results are not sensitive to variations in our specification.

How do we reconcile these results with the common belief that government spending during WWII lifted the economy out of the Great Depression? Our results do not dispute this notion, but instead reinterpret it. World War II government spending did help lift the economy out of the Great Depression, not because multipliers were so large, but because the amount of government spending was so great. Although multipliers may be modest in magnitude, they are positive.

In our analysis of multipliers in zero lower bound interest rate states, we also find no evidence that multipliers are greater than one at the zero lower bound in the full sample. The results are mixed, however, when we exclude WWII from the sample. Our preferred shock, the military news shock, indicates multipliers around 1.4 at the 2-year horizon, and the estimates are reasonably precise. On the other hand, the Blanchard-Perotti shock suggests multipliers just below one, but they are not precisely estimated.

We also conduct a comparison of the Jordà method to the TVAR method used by Auerbach and Gorodnichenko (2012). We show that their results depend on a simplifying assumption that is not a good approximation for their sample. We demonstrate that their recession state multiplier estimates are much lower once we relax that assumption. We then implement a TVAR on our sample and states, for which the simplifying assumption is more consistent with the data, and find results very close to those we estimated using the Jordà method.

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 the military news shock measures only changes in defense spending and because the Blanchard-Perotti shock mixes all types of shocks to government purchases, our results do not inform us about the size of multipliers on specific classes of government outlays, such as 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.

Data Appendix

GDP and GDP Deflator

1947–2015: Quarterly data on chain-weighted real GDP, nominal GDP, and GDP deflator from Bureau of Economic Analysis (BEA) NIPA (downloaded from the Federal Reserve Bank of St. Louis Economic Data [FRED], March 25, 2016 revision).

1889–1946: Annual data from 1929–46 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* (Carter et al. 2006). These series are based on the work of Kuznets, Kendrick, Gallman, and Balke-Gordon.

1939–46: 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 1939–46, we used a simplified version of the procedure used in Ramey (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 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 ensures 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 1939–46. Finally, we linked the earlier series to the modern quarterly NIPA series from 1947 to the present.

Potential GDP

The real GDP time trend is estimated as a sixth-degree polynomial for the logarithm of GDP, from 1889q1 through 2015q4 excluding 1930q1 through 1946q4. Somewhat lower-degree and somewhat higher-degree polynomials gave similar results for multipliers. 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 they 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

1947–2015: Quarterly data on nominal Government Consumption Expenditures and Gross Investment, from BEA NIPA (downloaded from FRED, March 25, 2016, revision).

1889–1946: NIPA annual nominal data from 1929–46 (BEA table 1.1.5, line 21) is spliced to annual data from 1889–1928. Source: Kendrick (1961, table A-II).

1939–46: Quarterly data on nominal government spending from *National Income, 1954 Edition: A Supplement to the Survey of Current Business*, are used to interpolate the modern annual NIPA values.

1889–1938: Monthly data on federal budget expenditures. Source: NBER Macrohistory Database (http://www.nber.org/databases/macrohistory/contents/chapter15.html): m15005a US Federal Budget Expenditures, Total 01/1879–09/1915; m15005b US Federal Budget Expenditures, Total 11/1914–06/1933; m15005c US 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, 1934:1, 1936:6, and 1937:6. 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 that of the previous quarter. Note that our use of the proportional Denton method creates a bumpier series than an alternative that uses the additive Denton method. However, the additive Denton method leads to series that behave very strangely around large buildups and builddowns of government spending, so we did not use it. On the other hand, the alternative series gave very similar results for the multiplier.

We seasonally adjust the monthly interpolator series, but the quarterly interpolation still had some residual seasonality. So we applied X-12 again to the quarterly interpolated series for 1879q3–1938q4.

Military News

The narrative underlying the series is available in Ramey (2016).

Population

1890–2015: Annual population data, based on July of each year, were taken from 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 FRED database, "POP."

Data adjustment: For 1890–1951, 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.

Federal Tax Revenues and Federal Expenditure

We create federal deficit series by subtracting our federal tax revenue series from our federal expenditures series (note these are total expenditures, not just government purchases).

1947–2015: Quarterly data on nominal Federal Government Current Receipts, BEA table 3.2, line 1, March 25, 2016, version. Note that all NIPA BEA data are on an accrual basis.

Quarterly data for 1959q3–2015q4 are from table 3.2, line 42, Total Expenditures. The period 1947q1–1959q2 did not show the total because one of the elements was missing. Since the missing element (net purchases of nonproduced assets, line 46) is so small, we assume it was 0 and added up the other elements (lines 43 + 44 + 45 − 47) to get the total.

1879–1938: Monthly data on federal budget receipts. Source: NBER Macrohistory Database (http://www.nber.org/databases/macrohistory/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.

Monthly data on federal expenditures. Source: NBER Macrohistory Database: m15005a U.S. Federal Expenditures, Total 01/1879–10/1914; m15005b U.S. Federal Expenditures, Total 11/1914–12/1931; m15005c U.S. Federal Expenditures, Total 01/1932–06/1937; m15005d U.S. Federal Expenditures, Total 07/1937–06/1939; m15005e U.S. Federal Expenditures, Total 07/1939–06/1945; m15005f U.S. Federal Expenditures, Total 07/1945–12/1946.

1939–46: 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. Expenditures are same source as receipts with expenditures = federal purchases + total transfers.

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

1929–46: Annual data on nominal Federal Government Current Receipts, BEA table 3.2, line 1, March 27, 2014, version. Annual data on nominal Federal Expenditures, BEA table 3.2, adding up lines 43 + 44 + 45 + 46 − 47, treating missing components as zeros since they were small once they became available, March 25, 2016, 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. Same adjustment for expenditures as for receipts. Note, we seasonally adjust the monthly interpolator series, but the quarterly interpolation still had some residual seasonality. So, we applied X-12 again to the quarterly interpolated series for 1879q3–1938q4.

Unemployment Rate

1948–2015: Monthly civilian unemployment rate. Source: 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 1933–43 to include emergency workers from Conference Board (1945).

1890–1929: NBER-based monthly recession indicators. Source: FRED database, USREC (http://research.stlouisfed.org/fred2/series/USREC).

1930–46: Monthly civilian unemployment rate (including emergency workers). Source: NBER Macrohistory Database (http://www.nber.org/databases

/macrohistory/contents/chapter08.html): m08292a U.S. Unemployment Rate, Seasonally Adjusted 04/1929–06/1942; m08292a U.S. Unemployment Rate, Seasonally Adjusted 01/1940, 03/1940–12/1946.

1947: Monthly civilian unemployment rate (including emergency workers, seasonally adjusted). Source: Geoffrey Moore, *Business Cycle Indicators*, vol. 2, NBER, p. 122.

Data adjustment: Monthly NBER recession data are used to interpolate annual data using the Denton interpolation from 1890–1929. For 1930–47 onward we use the monthly unemployment rate series to interpolate annual data using the Denton proportional interpolation.

Interest Rate

1934–2015: Monthly 3-month Treasury bill. Source: FRED database, TB3MS (http://research.stlouisfed.org/fred2/series/TB3MS).

1920–33: Monthly 3-month Treasury bill. Source: NBER Macrohistory Database (http://www.nber.org/databases/macrohistory/contents/chapter13.html): m13029a U.S. Yields On Short-Term United States 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.

References

Alloza, Mario. 2014. "Is Fiscal Policy More Effective in Uncertain Times or during Recessions?" Working paper, Univ. Coll. London.

Anderson, T. W., and H. Rubin. 1949. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *Ann. Math. Statis.* 20:46–63.

Aruoba, S. Boragan, and Frank Schorfheide. 2013. "Macroeconomic Dynamics Near the ZLB: A Tale of Two Equilibria." Working Paper no. 19248, NBER, Cambridge, MA.

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

———. 2013. "Fiscal Multipliers in Recession and Expansion." In *Fiscal Policy after the Financial Crisis*, edited by Alberto Alesina and Francesco Giavazzi, 63–98. Chicago: Univ. Chicago Press.

Bachmann, Rudiger, and Eric R. Sims. 2012. "Confidence and the Transmission of Government Spending Shocks." *J. Monetary Econ.* 59:235–49.

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

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

Batini, Nicoletta, Giovanni Callegari, and Giovanni Melina. 2012. "Successful Austerity in the United States, Europe and Japan." Working Paper no. 12/190, Internat. Monetary Fund, Washington, DC.

Baum, Anja, Marcos Poplawski-Ribeiro, and Anke Weber. 2012. "Fiscal Multipliers and the State of the Economy." Manuscript, Internat. Monetary Fund, Washington, DC.

Blanchard, Olivier, and Roberto Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *Q.J.E.* 117 (4): 1329–68.

Braun, R. Anton, Lena Mareen Korber, and Yuichiro Waki. 2013. "Small and Orthodox Fiscal Multipliers at the Zero Lower Bound." Working Paper no. 2013-13, Fed. Reserve Bank Atlanta.

Bruckner, Marcus, and Anita Tuladhar. 2014. "Local Government Spending Multipliers and Financial Distress: Evidence from Japanese Prefectures." *Econ. J.* 124 (581): 1279–1316.

Caggiano, Giovanni, Efrem Castelnuovo, Valentina Colombo, and Gabriela Nodari. 2015. "Estimating Fiscal Multipliers: News from a Non-linear World." *Econ. J.* 125 (584): 746–76.

Canzoneri, Matthew, Fabrice Collard, Harris Dellas, and Behzad Diba. 2016. "Fiscal Multipliers in Recessions." *Econ. J.* 126 (590): 75–108.

Carter, Susan B., Scott Sigmund Gartner, Michael R. Haines, et al. 2006. "National Income and Product." In *Historical Statistics of the United States*, Millennial Ed. Online, chap. Ca. Cambridge: Cambridge Univ. Press.

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

Coenen, Günter, et al. 2012. "Effects of Fiscal Stimulus in Structural Models." *American Econ. J.: 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." *J. Econ. Dynamics and Control* 34:281–95.

Conference Board. 1945. *The Economic Almanac, 1945–46*. New York: Nat. Indus. Conf. Board.

Corsetti, Giancarlo, Andre Meier, and Gernot Mueller. 2012. "What Determines Government Spending Multipliers?" *Econ. Policy* 27 (72): 521–65.

Crabbe, Leland. 1989. "The International Gold Standard and U.S. Monetary Policy from World War I to the New Deal." *Fed. Reserve Bull.* 75 (June): 423–40.

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

Davig, Troy, and Eric M. Leeper. 2011. "Monetary-Fiscal Policy Interactions and Fiscal Stimulus." *European Econ. Rev.* 55 (2): 211–27.

Drautzburg, Thorsten, and Harald Uhlig. 2015. "Fiscal Stimulus and Distortionary Taxation." *Rev. Econ. Dynamics* 18 (4): 894–920.

Eggertsson, Gauti B. 2011. "What Fiscal Policy Is Effective at Zero Interest Rates?" In *NBER Macroeconomics Annual*, vol. 25, edited by Daron Acemoglu and Michael Woodford, 59–112. Chicago: Univ. Chicago Press.

Farhi, Emmanuel, and Ivan Werning. 2016. "Fiscal Multipliers: Liquidity Traps and Currency Unions." In *Handbook of Macroeconomics*, vol. 2, edited by John B. Taylor and Harald Uhlig. Amsterdam: Elsevier.

Fazzari, Steven M., James Morley, and Irina Panovska. 2015. "State-Dependent Effects of Fiscal Policy." *Studies Nonlinear Dynamics and Econometrics* 19 (3): 285–315.

Fisher, Jonas D. M., and Ryan Peters. 2010. "Using Stock Returns to Identify Government Spending Shocks." *Econ. J.* 120:414–36.

Friedman, Milton. 1952. "Price, Income, and Monetary Changes in Three Wartime Periods." *A.E.R.* 42 (2): 612–25.

Gordon, Robert J., and Robert Krenn. 2010. "The End of the Great Depression: VAR Insight on the Roles of Monetary and Fiscal Policy." Working Paper no. 16380, NBER, Cambridge, MA.

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

Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57 (2): 357–84.

Jordà, Óscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *A.E.R.* 95 (1): 161–82.

Kendrick, John D. 1961. *Productivity Trends in the United States*. Princeton, NJ: Princeton Univ. Press (for NBER).

Kiley, Michael T. 2014. "Policy Paradoxes in the New Keynesian Model." Finance and Economics Discussion Series, no. 2014-29, Board Governors, Fed. Reserve System, Washington, DC.

Koop, Gary, Hashem M. Pesaran, and Simon M. Potter. 1996. "Impulse Response Analysis in Nonlinear Multivariate Models." *J. Econometrics* 74 (1): 119–47.

McGrattan, Ellen R., and Lee E. Ohanian. 2010. "Does Neoclassical Theory Account for the Effects of Big Fiscal Shocks? Evidence from World War II." *Internat. Econ. Rev.* 51 (2): 509–32.

Mertens, Karel, and Morten O. Ravn. 2014. "Fiscal Policy in an Expectations Driven Liquidity Trap." *Rev. Econ. Studies* 81 (4): 1637–67.

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

Miyamoto, Wataru, Thuy Lan Nguyen, and Dmitriy Sergeyev. Forthcoming. "Government Spending Multipliers under the Zero Lower Bound: Evidence from Japan." *American Econ. J.: Macroeconomics*.

Mountford, Andrew, and Harald Uhlig. 2009. "What Are the Effects of Fiscal Policy Shocks?" *J. Appl. Econometrics* 24:960–92.

Nakamura, Emi, and Jòn Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *A.E.R.* 104 (3): 753–92.

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3): 703–8.

Ohanian, Lee E. 1997. "The Macroeconomic Effects of War Finance in the United States: World War II and the Korean War." *A.E.R.* 87 (1): 23–40.

Olea, Jose Luis Montiel, and Carolin Pflueger. 2013. "A Robust Test for Weak Instruments." *J. Bus. and Econ. Statis.* 31 (3): 358–69.

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." *A.E.R.* 103 (2): 129–34.

Pflueger, Carolin E., and Su Wang. 2015. "A Robust Test for Weak Instruments in Stata." *Stata J.* 15 (1): 216–25.

Ramey, Valerie A. 2009. "Identifying Government Spending Shocks: It's All in the Timing." Working Paper no. 15464, NBER, Cambridge, MA.

———. 2011. "Identifying Government Spending Shocks: It's All in the Timing." *Q.J.E.* 126 (1): 51–102.

———. 2013. "Government Spending and Private Activity." In *Fiscal Policy after the Financial Crisis*, edited by Alberto Alesina and Francesco Giavazzi, 19–55. Chicago: Univ. Chicago Press.

———. 2016. "Defense News Shocks, 1889–2015: Estimates Based on News Sources." Manuscript, Univ. California San Diego.

Riera-Crichton, Daniel, Carlos A. Vegh, and Guillermo Vuletin. 2015. "Procyclical and Countercyclical Fiscal Multipliers: Evidence from OECD Countries." *J. Internat. Money and Finance* 52:15–31.

Romer, Christina D. 1999. "Changes in Business Cycles: Evidence and Explanations." *J. Econ. Perspectives* 13 (2): 23–44.

Sims, Eric, and Jonathan Wolff. 2013. "The Output and Welfare Effects of Government Spending Shocks over the Business Cycle." Working Paper no. 19749, NBER, Cambridge, MA.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–86.

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

Uhlig, Harald. 2010. "Some Fiscal Calculus." *A.E.R.* 100 (2): 30–34.

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, 301–46. Greenwich, CT: JAI.

Zubairy, Sarah. 2014. "On Fiscal Multipliers: Estimates from a Medium Scale DSGE Model." *Internat. Econ. Rev.* 55 (1): 169–95.