Are Government Spending Multipliers State Dependent? Evidence from U.S. and Canadian Historical Data

Michael T. Owyang
Research Division, Federal Reserve Bank of St. Louis

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

Sarah Zubairy
Texas A&M University

Preliminary version: 7 September 2013

The views expressed herein should not be taken to be the official opinions of the Federal Reserve Bank of St. Louis or the Federal Reserve System. The authors thank Robert Barro and Gordon Liao for use of their Canadian newspaper excerpts from WWI through the Korean War and Michelle Alexopoulos for providing us with some of the historical Canadian data. We are also grateful to Yuriy Gorodnichenko, Òscar Jordà and Garey Ramey for very helpful suggestions and to Kate Vermann and Michelle Ramey for research assistance.

Abstract

This paper investigates whether government spending multipliers differ according to two potentially important features of the economy: (1) the amount of slack in the economy and (2) whether interest rates are near the zero lower bound. We shed light on these questions by analyzing new quarterly historical data covering multiple large wars and deep recessions in the U.S. and Canada. We find no evidence that multipliers are greater during periods of high unemployment in the U.S, while we do find some evidence of higher multipliers during periods of slack in Canada. We also explore in detail the potential sources of difference in our results for the US from the pre-existing literature. Considering the impact of the zero lower bound, our preliminary analysis of multipliers in the U.S. based on monetary stance do not provide any evidence of larger multipliers at the zero lower bound for the U.S.

1 Introduction

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

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

A small empirical literature has begun to explore whether estimates of government spending multipliers vary depending on circumstances. One strand of this literature considers the possibility that multipliers are different during recessions. Auerbach and Gorodnichenko (2012) was one of the first papers to explore this possibility. Another strand estimates multipliers when interest rates are at the zero lower bound (e.g. Ramey (2011), Crafts and Mills (2012)). In addition, there are several estimated New Keynesian DSGE models that calculate multipliers that are conditional on the behavior monetary policy (e.g. Coenen and et al (2012)).

This paper contributes to the empirical literature by investigating whether government spending multipliers differ according to two potentially important features of the economy: (1) the amount of slack in the economy and (2) whether interest rates are near the zero lower bound. Extending the initial steps in our shorter paper (Owyang, Ramey, Zubairy (2013)), we exploit the fact that the entire 20th Century contains potentially richer information than the post-WWII data that has been the focus of most of the recent research. We create a new quarterly data set for the U.S. extending back to 1889 and for Canada extending back to 1912. These samples include episodes of huge variations in government spending, wide fluctuations in unemployment, prolonged periods near the zero lower bound of interest rates, and a variety of tax responses.

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

For the U.S., we find no evidence that the multiplier on government purchases is higher during high unemployment states or when interest rates are near the zero lower bound. Most estimates of the multiplier are between 0.8 and 1.1. In contrast, for Canada we find that multipliers are significant higher during high unemployment states, but in both states multipliers are modest, between 0 and 1.1.

The paper proceeds as follows. We begin by discussing the data construction and describing the indicators of the key states of the economy in Section 2. In Section 3 we introduce the econometric methodology and discuss issues with calculating impulse responses.

In Section 4, we then present estimates of a model in which multipliers are allowed to vary according to the amount of slack in the economy. We first present baseline results using our new data and methodology. We then conduct various robustness checks and then explain why our results are different from those in the literature. We also explore possible explanations for our results, such as the behavior of taxes. Section 5 tests theories that predict that multipliers should be greater when interest rates are at the zero lower bound and the final section concludes.

2 Data Description and Features

A key contribution of the paper is the construction of a new data set that spans historical periods that involve potentially informative movements of the key variables. In particular, we construct quarterly data from 1889 through 2010 for the U.S. and from 1912 through 2011 for Canada. 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.[1] The series for each country include real GDP, the GDP deflator, government purchases, tax revenue, deficit, population, the unemployment rate, interest rates, and defense news.

The separate data appendix contains full details, but we highlight some of the features of the data here. We use available quarterly series for the later sample, typically 1947 on for the U.S. and 1961 on for Canada. For the earlier periods, we use various higher frequency series to interpolate existing annual series, similar to the procedure used by Gordon and Krenn (2010). In most cases, we use the proportional Denton procedure which results in series that average up to the annual series. For example, for the earlier period in the U.S., we interpolate annual data from *Historical Statistics of the U.S.* (Carter et al. (2006)) using Gordon and Balke's (1986) quarterly real GNP series (which were themselves constructed using industrial production as an interpolator). For Canada, we interpolate annual data compiled by McInnis (2001) using monthly industrial production. We use monthly series on nominal federal revenues and outlays to interpolate annual series on nominal purchases, outlays and revenues.

The method for unemployment is somewhat difference. For the U.S., before 1948 we use the monthly unemployment series available from the NBER Macrohistory database back to April 1929 to interpolate Weir's (1992) annual unemployment series. Before 1929,

1. For example, the unemployment rate in the U.S. fell from over 10 percent to 5 percent between mid-1941 to mid-1942.

we interpolate Weir's (1992) annual unemployment series using business cycle dates and the additive version of Denton's method. Our comparison of the series produced using this method with the actual quarterly series in the post-WWII period reveal that they were surprisingly close. For Canada also, for pre-1954, we combine modern sources with historical sources for the annual rate and used business cycle dates to interpolate.

It should be noted that in both countries, the interpolator series for government spending and taxes are very noisy and we suspect the many large jumps we observe are due to vagaries of government budget accounting rather than actual jumps in spending or taxes. For this reason, our government spending series should ***not*** be used to identify shocks using standard Choleski decompositions (such as the method of Blanchard and Perotti (2002)). Fortunately, the measurement errors are less of an issue for us because we identify the shocks using narrative methods.[2]

Because it is important to identify a shock that is not only exogenous to the state of the economy but is also unanticipated, we use narrative methods to extend Ramey's (2011) news series in time for the U.S. and to construct an entire series for Canada.[3] This news series focuses on changes in government spending that are linked to political and military events, since these changes are most likely to be independent of the state of the economy. Moreover, changes in defense spending are anticipated long before they actually show up in the NIPA accounts. For a benchmark neoclassical model, the key effect of government spending is through the wealth effect. Thus, the news series is constructed as changes in the expected present discounted value of government spending. The particular form of the variable used as the shock is this nominal value divided by one-quarter lag of nominal GDP.

In the estimation, we consider two key ways in which the economy's state can vary. The first is the amount of slack in the economy. The traditional Keynesian idea of government spending multipliers, on which so much of modern intuition is based, assumes an economy with substantial underutilization of resources so that output is demand-determined rather than supply-determined. There are various potential measures of slack, such as output gaps, the unemployment rate, or capacity utilization. Based on data availability and the fact that it is is generally accepted as a key measure of underutilized resources, we use the unemployment rate as our indicator of slack. We define an economy to be in a slack state

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

3. The background narrative will be made available in the near future.

when the unemployment rate is above some threshold. For our baseline results, we use 6.5 as the threshold based on Chairman Bernanke's recent announcement about policy. We also conduct various robustness checks using time varying thresholds.

Note that our use of the unemployment rate to define the state is different from using NBER recessions or Auerbach and Gorodnichenko's (2012) moving average of GDP growth. The latter two measures, which are highly correlated, indicate periods in which the economy is moving from its peak to its trough. A typical recession encompasses periods in which unemployment is *rising* from its low point to its high point, and hence is not an indicator of a state of slack. We know of no theory that suggests that multipliers should differ according to the *change* in unemployment rates. In both Canada and the U.S., only half of the quarters that are official recessions are also periods of high unemployment.

Figure 1 shows the logarithm of per capita government purchases (deflated by the GDP deflator) for the U.S. and Canada. These include all federal, state (province), and local purchases; they exclude transfer payments. Note that the large events include WWI, WWII and the Korean War, where the entry of Canada in the World Wars is earlier than the U.S. in both cases. Particularly in the post-WWII period, it is difficult to see the military buildups in Canada's total government purchases series. Our separate analysis of defense purchases (not shown here) shows that with the exception of the Vietnam War, Canada's military build-ups and draw-downs are qualitatively similar to those of the U.S. While the percent change in defense spending in Canada during the post-WWII buildups was as large as in the U.S., military spending as a percent of GDP has been less than half as much in Canada as in the U.S.

Figure 2 and Figure 3 show the unemployment rate and the military spending news shocks for the two countries. As Figure 2 shows, the largest military spending news shocks are distributed across periods with a variety of unemployment rates. For example, the largest news shocks about WWI and the Korean War occurred when the unemployment rate was below the threshold. In contrast, the initial large news shocks about WWII occurred when the unemployment rate was still very high.

Because our method for estimation can be interpreted as an instrumental variables regression, it is important to gauge the relevance of the news variable as an instrument. Table 1 shows the F-tests for the exclusion of the news variables in regressions of log real per capita government spending on four lags of its own value, four lags of log real per capita GDP, and the current value and four lags of the news variable (scaled by the previous quarter's nominal GDP), as well as a quartic trend. The table shows these for the full sample as well as the post-WWII sample, and splits each of these according to whether the unemployment rate is above 6.5 percent. According to Staiger and Stock (1997), a first-stage F-statistic below 10 can indicate that the instrument may have low relevance. For the full historical and post-WWII samples, the statistics are all around 10 or above. The F-statistics fall when the samples are split between slack and non-slack quarters. While they are less than the safety threshold of 10, most are substantially higher than the typical macro F-statistic. The exception is the slack state in both countries in the post-WWII period, where both F-statistics are barely above 1. These low statistics indicate that not much can be learned about the difference in multipliers between slack and non-slack states in the post-WWII period. The F-statistics during slack states are much higher for the full historical samples. This difference supports our initial conjecture that the post-WWII sample was not sufficiently rich to be able to distinguish multipliers across states using the military news instrument.[4]

The second way in which we allow the economy's state to vary is by considering whether government interest rates are near the zero lower bound or are being held constant to accommodate fiscal policy. As we will discuss below, New Keynesian models suggest that government spending multipliers will be higher when the economy is at the zero lower bound. In fact, the prediction holds any time the monetary authority does not change the interest rate in response to changes in government spending. To test this hypothesis, we identify periods that satisfy these criterion. So far we have only done so for the U.S. Figure 4 shows short-term interest rates on 3-month Treasury bills (since the federal funds rate is not available until 1954). We classify as "near zero lower bound" or extremely accommodative the periods starting in 1932q2 and extending through 1951q1 after which the Treasury Accord freed up the Federal Reserve, and again starting in 2008q4 through 2011. The shaded areas show those sample periods.

Table 1 also shows the F-tests for the periods split into ZLB periods and normal periods. For the full sample, the F-test is only 2.25 for the ZLB periods, but is a surprisingly high 15 for the 26 ZLB quarters during the post-WWII period.

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

3 Econometric Methodology and Issues

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

3.1 Local Projections

Auerbach and Gorodnichenko (2013) were the first to recognize the advantages of applying Jordà's (2005) local projection technique to estimate state-dependent models and calculate impulse responses.[5] The Jordà method simply requires estimation of a series of regressions for each horizon h for each variable. The linear model looks as follows:

$$z_{t+h} = \alpha_h + \psi_h(L)y_{t-1} + \beta_h shock_t + \varepsilon_t + h, \text{for h = 0, 1, 2, ...} \tag{1}$$

z is the variable of interest (discussed in much detail below), y is a vector of control variables, $\psi_h(L)$ is a polynomial in the lag operator, and *shock* is the identified shock. The coefficient β_h gives the response of z 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 separate regressions for each horizon. Our vector of control variables, y, contains logs of real per capita GDP, government spending and federal revenues, and $\psi(L)$ is a polynomial of order 4.[6]

This 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:

$$z_{t+h} = I_{t-1}\left[\alpha_{A,h} + \psi_{A,h}(L)y_{t-1} + \beta_{A,h} shock_t\right] + (1 - I_{t-1})\left[\alpha_{B,h} + \psi_{B,h}(L)y_{t-1} + \beta_{B,h} shock_t\right] + \varepsilon_{t+h}. \tag{2}$$

I is a dummy variable that indicates the state of the economy before the shock hits. We allow all of the coefficients of the model (other than deterministic trends) to vary according to the state of the economy. As discussed in Section 2, the shock is identified as the news variable scaled by lagged nominal GDP. The only complication associated with the Jordà

5. They applied this method in their analysis of state-dependent multipliers in OECD panel data.

6. Note here in departure from Owyang et al. (2013), we additionally use log of tax revenues as a control variable.

method is the serial correlation in the error terms induced by the successive leading of the dependent variable. Thus, we use the Newey-West correction for our standard errors (Newey and West (1987)).

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

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

3.2 Pitfalls in Estimating Impulse Responses in Nonlinear Models

Most of the literature has estimated non-linear VAR models to address state-dependence of multipliers. For example, Auerbach and Gorodnichenko (2012) (henceforth AG-12) and Bachmann and Sims (2012) use a regime switching model in which the transition across states is smooth (Smooth Transition VAR, or STVAR). Fazzari et al. (2013), Baum et al. (2012), Mittnik and Semmler (2012), and Semmler and Semmler (2013) use threshold vector autoregressions (TARs) and employ different indicators of the state. As Koop et al. (1996) point out, estimating impulse responses in nonlinear models such as these is far from straightforward. We now discuss the issues involved in the context of the AG-12 model since so much other work builds on it. We compare what they do to the method advocated by Koop et al. (1996). Finally, we compare the computation of impulse responses using the

Jordà method to the Koop et al. (1996) method.

3.2.1 The Auerbach and Gorodnichenko (2012) STVAR Model

Let X_t be a vector of macroeconomic variables of interest—in their case, government spending, taxes, and output, in that order. The STVAR model is a combination of two linear VARs:

$$X_t = [1 - F(z_{t-1})] A_E(L) X_{t-1} + F(z_{t-1}) A_R(L) X_{t-1} + \Pi(L) z_{t-1} + \varepsilon_t, \tag{3}$$

where $A_E(L) X_{t-1}$ and $A_R(L) X_{t-1}$ reflect the dynamics when the economy is in the most extreme expansion and recession, respectively. The model implies that the economy is a convex combination of the two extreme regime dynamics, where $F(z_{t-1})$ is the transition function that determines how the two regimes are combined at any given t and demeaned z_t is a centered seven-quarter moving average of output growth. AG-12 specify a logistic transition function:

$$F(z_{t-1}) = \frac{\exp(-\gamma z_{t-1})}{1 + \exp(-\gamma z_{t-1})}, \tag{4}$$

which is calibrated to the match features of the U.S. business cycle. The errors are regime switching:

$$\varepsilon_t \sim N(0, \Omega_t),$$

where

$$\Omega_t = [1 - F(z_{t-1})] \Omega_E + F(z_{t-1}) \Omega_R.$$

3.2.2 Impulse Responses in Nonlinear VAR Models

Government spending multipliers are computed by comparing the response of output to the response of government spending after a shock. Thus, the impulse responses are the key building blocks for estimating multipliers. We can think of an impulse response as the difference between two conditional expectations. The impulse response to a shock $\varepsilon_t = \delta$

occurring at time t computed at horizon h is

$$IR(h) = E_t [X_{t+h}|\varepsilon_t = \delta] - E_t [X_{t+h}|\varepsilon_t = 0], \tag{5}$$

where we have implicitly conditioned on identical histories Θ_{t-1}. In a linear model, the impulse response is invariant to history and δ only acts as a scaling factor. In a nonlinear model, the response can depend on the magnitude (and sign) of the shock and the history up to time t. In the STVAR model, varying the history can alter, for example, z_t which determines the dynamics. Altering the size of the shock can scale the future response and, again, alter the future realizations of z_t. In the linear model, the sequence of subsequent innovations does not change the impulse responses since the difference nullifies their (linear) effect. In the nonlinear model, the sequence of subsequent innovations can affect the future values of z_{t-1} and, thus, alter the dynamics. This is especially important in the STVAR model, because the (future) values of z_t and, thus, future values of $F(z_t)$, alter both the variance-covariance matrix from which the (future) innovations are drawn and the dynamics of the model.

In order to simplify their calculations, AG-12 compute regime-dependent impulse responses, that is, they condition on a fixed value of $F(z_{t-1})$.[7] They compute two sets of responses:

$$IR_k(h) = E_t [X_{t+h}|\Theta_{t-1}, \varepsilon_t = \delta, F(z_{t-1}) = k] - E_t [X_{t+h}|\Theta_{t-1}, \varepsilon_t = 0, F(z_{t-1}) = k]$$

for $k = 0$ (expansion) and $k = 1$ (recession). Conditional on the regime, the model becomes linear, meaning the economy's history and the scale of the shock can be factored out. The sequence of future shocks are also irrelevant as they disappear in the difference. However, restricting $F(z_{t-1}) = k$ in AG-12's case is problematic. First, the economy rarely resides in either $F(z_{t-1}) = 0$ or $F(z_{t-1}) = 1$ for any length of time.[8] Second, even if the economy started in one of the discrete regimes, Y_{t+h-i} affects z_{t+h-1}—the response of output affects

7. The regime-dependent responses for STVAR models are similar to those used by Ehrmann et al. (2003) for Markov-switching models. The key difference between the STVAR models and Ehrmann et al. (2003) is that the latter's state variable is independent of the vector of variables of interest. Thus, the impulse responses do not feed back into the state variable.

8. If the state variables were independent of the macro variables in (3), the impulse responses could be computed as a convex combination of the two regime-dependent response. Because there is feedback, however, this is no longer true.

future regimes and, thus, the dynamics of future responses.[9]

The generalized impulse responses of Koop et al. (1996) attempt to solve the three problems (history dependence, future innovation dependence, and shock dependence) through Monte Carlo integration. The Koop et al. (1996) method, called the generalized impulse response function (GIRF) computes the difference in the conditional expectations, (5), by calculating

$$IR(h) = E_t [X_{t+h} | \varepsilon_t = \delta] - E_t [X_{t+h} | \varepsilon_t = 0] \tag{6}$$

and by calculating

$$IR(h) = E_t [X_{t+h} | \Theta_{t-1}, \varepsilon_t = \delta] - E_t [X_{t+h} | \Theta_{t-1}, \varepsilon_t = 0] \tag{7}$$

for each history. Koop et al. (1996) then integrate over the set of observed histories to obtain the GIRF.

Later in the paper, we show that the Jordà method applied to AG-12 data and definition of state gives very different results from those obtained by AG-12. To understand the differences, we compute GIRFs for the estimated STVAR model. In particular, we compute the future values of X_{t+h}, conditional on $\varepsilon_t = \delta$ and conditional on $\varepsilon_t = 0$ for some set number of Monte Carlo innovations, $\varepsilon_{t+1}, ..., \varepsilon_{t+H}$.[10] Thus, we compute

$$E_t [X_{t+h} | \Theta_{t-1}, \varepsilon_t = \delta] = \frac{1}{M} \sum_{m=1}^{M} \left\{ \begin{array}{c} [1 - F(z_{t+h-1})] A_E(L) X_{t+h-1}^{shock} \\ +F(z_{t-1}) A_R(L) X_{t+h-1}^{shock} + \Pi(L) z_{t+h-1}^{shock} + \varepsilon_{t+h,m}^{shock} \end{array} \right\},$$

where the superscript *shock* represents the fact that the expectation was generated conditional on $\varepsilon_t = \delta$ and these are summed over M Monte Carlo draws. The same sequence innovations, taken from a draw of $N(0, \Omega_t)$, are used to compute each of the conditional

9. There is an additional problem associated with the use of the centered moving average as a transition variable. The centered moving average uses forward data to determine the state–i.e., Y_{t+1} determines the state used to compute Y_t. Thus, the response of Y_t to a shock is determined by Y_{t+1}, which is also affected by the Y_t, making the model internally inconsistent.

10. Again, in linear models, impulse responses are typically assumed to measure the evolution of variables in the absence of future shocks, because these future shocks affect each conditional expectation equally. The difference (6) removes their effect. Because of the nonlinearity and the feedback into the transition equation in AG's model, future shocks affect the future state and must be integrated out of the expectation.

expectations (i.e., each term in (7)). To be internally consistent, the GIRF includes the response from the last period in the history that generates the next response. Thus, the GIRF generates M future paths of the economy using Monte Carlo methods to draw from the distribution of the innovations. The final response is formed from computing the difference in the conditional expectations as the average of the M draws over the observed histories. In this case, the observed histories act as the starting values for each set of Monte Carlo experiments. The state evolves since z_{t+h-1} is recomputed at each h accounting for how values in X evolve. Note also that the model specification is important for correctly propagating the shock: If the model is misspecified, the bias in the response can be compounded for each additional horizon.

3.2.3 Impulse Responses in Local Projection Models

In this paper as well as our previous work in Owyang et al. (2013), we use regime-dependent local projections of Jordà (2005) as the method of computing impulse responses, which computes the difference between the conditional expectations directly versus iteratively as in the generalized impulse response function.[11]

To see how this method compares to the GIRFs, suppose we have a model that is constructed as:

$$X_t = F\left(\widetilde{X}_{t-1}, \widetilde{\varepsilon}_t\right),$$

where X_t is the period t value of the variable of interest, $\widetilde{X}_{t-1}$ is the history of realizations, and $\widetilde{\varepsilon}_t$ is the history of shocks. We have suppressed exogenous variables for notational convenience.

The GIRF computes the conditional expectation as

$$E_t\left[X_{t+h}|\varepsilon_t = \delta\right] = \frac{1}{N}\sum_{\{\widetilde{X}_{t-1}\}} E_t\left[X_{t+h}|\widetilde{X}_{t-1}, \varepsilon_t = \delta\right],$$

where the expectation is computed over the N observed histories. The GIRF uses Monte Carlo methods that integrate over possible outcomes to estimate the effect of future feedbacks. The GIRF, then, computes the expectation $E_t\left[X_{t+h}|\widetilde{X}_{t-1}, \varepsilon_t = \delta\right]$ using Monte Carlo methods by augmenting the history, $\widetilde{X}_{t-1}$, with the responses from the past horizons. Let

11. Marcellino et al. (2006) demonstrate some of the differences between direct and indirect methods for forecasting.

$\overline{X}_{t-1} = \left\{\widetilde{X}_{t-1}, \widehat{X}_t, ..., \widehat{X}_{t+h-1}\right\}$ be the augmented history, where $\widehat{X}_m$ is the period$-m$ conditional expectation, and $\overline{\varepsilon}_t = \{\widetilde{\varepsilon}_t, \widehat{\varepsilon}_t, ..., \widehat{\varepsilon}_{t+h-1}\}$, where $\widehat{\varepsilon}_m$ is the draw of the shock for period m. Then, we can compute

$$\widehat{X}_{t+h} = F\left(\overline{X}_{t+h-1}, \overline{\varepsilon}_{t+h}\right)$$

and the conditional expectation $E_t\left[X_{t+h}|\widetilde{X}_{t-1}, \varepsilon_t = \delta\right]$ as the average of M of draws of $\overline{\varepsilon}_{t+h}$.

The direct projection method also computes the impulse response as a difference in the conditional expectation. However, instead of using a single model relating X_t to $\widetilde{X}_{t-1}$, the direct projection method estimates the h-period-ahead response directly with a different model for each h:

$$X_{t+h} = F_h\left(\widetilde{X}_{t-1}\right).$$

The direct projection method computes the conditional expectation directly by generating a forecast of the h-period-ahead value of X regressing h-period-ahead on information available at time t. The direct projection method, then, embeds the historical observed feedback into the h-period-ahead forecast in computing the impulse response.

In the end, the two methods differ (for our purposes) by how they account for the feedback. The GIRF takes the feedback from the model and computes it (iteratively) at each horizon, using the response at the $h-1$ horizon to estimate the response at the h horizon. Direct projections does not use information at the $h-1$ horizon to compute the horizon h response. Instead, it computes the average h-period-ahead value forecast given only the information at the time the forecast is constructed. Whatever feedback occurs in the history of the h-period-ahead realizations will be accounted for in these forecasts. While direct projections may still produce biases if the forecasting model is misspecified, the errors may not be compounded at longer horizons because each horizon estimates a new model without including the past responses in the history.

3.3 Pitfalls in Converting Elasticities to Multipliers

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

We first noticed a potential problem with this method when we extended our sample back in time. In the post-WWII sample in the U.S., Y/G varies between 4 and 7, with a mean of 5. In our full sample for the U.S. from 1889-2011, Y/G varies from 2 to 24 and with a mean of 8. We realized that we could estimate the same elasticity of output with respect to government spending, but derive much higher multipliers simply because the mean of Y/G was so much higher. To determine whether using *ex post* conversion factors can lead to inflated multipliers, we ran the following experiment. To be consistent with the standard linear SVAR literature, we first estimated a trivariate SVAR with military news, log real per capita government spending, and log real GDP on the U.S. data from 1889 - 2011. The estimated elasticity was around 0.23 (based on the ratio of the peak of response of $ln(Y)$ to the peak of the response of $ln(G)$).[12] We then multiplied the estimated elasticity by the average of Y/G for the full sample, and obtained an implied multiplier of 1.84. We then re-estimated the model after substituting the log of private spending, defined as $Y - G$, for GDP, and computed the impulse response functions. These implied that private spending fell when government spending rose. A fall in private spending implies that government spending is crowding out private spending, so the multiplier *must* be less than unity. Thus, the practice of backing out multipliers using *ex post* conversion factors can lead to upward biased multiplier estimates.[13]

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

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

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

12. The SVAR contains four lags of each variable and a quartic trend.

13. This insight also explains the discrepancy in estimates between two other papers. Ramey (2011) uses the standard method of estimating elasticities and converting them and calculates multipliers around 1.1 to 1.2. In contrast, Ramey (2013) estimates the same system with private spending substituted for GDP and finds statistically significant *negative* elasticities of real private spending to government spending.

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

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

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

4 Multipliers During Times of Slack

4.1 Theoretical Literature

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

Numerous papers explore theoretically the possibility of state-dependent multipliers that depend on the debt-to-GDP ratio, the condition of the financial system, and exchange rate regimes.[14] Other than the zero lower bound papers, which make a distinct argument that we will discuss below, there is only one paper of which we are aware that analyzes a rigorous model that produces fiscal multipliers that are higher during times of high unemployment. Michaillat (2014) develops a search and matching model and shows that the multiplier on one particular type of government spending doubles as the unemployment rate rises from 5 percent to 8 percent. In particular, he analyzes government spending on public employment. However, Michaillat (2014) does not model the original Keynesian notion that arbitrary government spending can stimulate private employment. Thus, there is still a gap between Keynes' original notion and modern theories.

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

4.2 Baseline Results for Slack States

We now present the main results of our analysis using the full historical samples and the local projections method. We first consider results from the linear model, which assumes that multipliers are invariant to the state of the economy. The top panel of Figure 5 shows the responses of government spending and output to a military news shock in the linear model using the U.S. data. The bands are 95 percent confidence bands and are based on Newey-West standard errors that account for the serial correlation induced in regressions when the horizon $h > 0$. After a shock to news, output and government spending begin to rise and peak at around 12 quarters.

In the linear model, the multipliers are derived from the estimated β_h from the Y and G equations. We compute multipliers over three horizons: the ratio of the peak of the GDP response to the peak of the government spending response, the ratio of the cumulative responses through two years, and the cumulative responses through four years.[15] As indicated in the first column of the top panel of Table 2, the implied multipliers are below one and range from 0.8 to 0.9. The estimates are not statistically different from one at the five percent significance level.[16]

The main question addressed in this paper is whether the multipliers are state-dependent. The impulse response functions and multipliers in the state-dependent case are derived from the estimated $\beta_{A,h}$ and $\beta_{B,h}$ for Y and G in equation 2. The bottom panel of Figure 5 shows the responses when we estimate the state-dependent model where we distinguish between periods with and without slack in the economy. Similar to many pre-existing studies (e.g. AG-12), we find that output responds more robustly during high unemployment states. Also, during the high unemployment state both output and government spending peak almost a year later than the linear case, at around 12 to 14 quarters after the shock to news. Note that government spending also has a stronger response during those high slack periods. Consequently, the larger output response during the high unemployment state does not imply a larger government spending multiplier. In fact, as shown in the second and third column of Table 2, the implied multipliers are slightly lower during the high unemployment state in the U.S. data. In no case do we find a statistically different multiplier across states.

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

16. The variance-covariance matrix is computed by estimating all of the regressions as one panel regression and using the Newey-West procedure to adjust the standard errors. We thank Yuriy Gorodnichenko for suggesting this method.

These results should not be interpreted as showing a systematic difference in the way that governments change their spending during recessions. Rather, the difference in paths of government are likely due to the particular types of military shocks that hit irrespective of the state. The initial WWII news shocks hit when unemployment was very high. This war was also the biggest war fought during the sample. This combination is likely the source of the differences across states.

The top panel of Figure 6 shows the results for the linear model using the Canadian data. Both government spending and output rise in a sustained manner, though the estimated government spending responses are rather erratic. As the first column of Table 2 shows, the implied multipliers are below unity in the linear model. The fiscal multipliers for Canada are consistently smaller than the ones for U.S. regardless of the definition used, and range between 0.4 and 0.5 and are usually statistically different from unity.

The bottom panel of Figure 6 shows the results from the state dependent model. The responses of government spending and GDP are not very different for the first two years across states, but then diverge starting in the third year when both government spending and GDP climb significantly in the high unemployment state. Table 2 shows that the implied multipliers are greater during periods of slack in Canada. For example, using the multipliers based on the integral through two years, the value is 1.08 when the initial shock hits during the high unemployment state in contrast to only 0.09 when it hits in the low unemployment state. Thus, the Canadian estimates suggest that multipliers are substantially greater in the high unemployment state. The exact values depend on the horizon. While the multipliers in high unemployment state can exceed 1, the low unemployment multiplier can also take a negative value.[17]

To summarize we find over our full sample that the multipliers for both U.S. and Canada tend to be 1 or less, where they are much smaller for Canada relative to the U.S. Considering state dependence, we find no evidence of larger multipliers in the periods of slack for the U.S, and multipliers vary between 0.8 and 1. In contrast, there is evidence in the Canadian data that suggests higher multiplier during periods of high unemployment in the economy.

17. These multipliers are rather different than the ones showed for Canada in Owyang et al. (2013), since we have now extended the data set to start in 1912 instead of 1921, and have further refined many of the time series that were preliminary estimates. In addition, we are now also accounting for taxes as a control variable.

4.3 Robustness

The next issue we address is the robustness of our findings to the choice of our specific threshold and our sample period.

We first consider a time-varying threshold, where we consider deviations from trend for Hodrick-Prescott filtered unemployment rate with a very high smoothing parameter of $\lambda = 1,000,000$. This definition of threshold results in about 40 percent of the observations being above the threshold for both U.S. and Canada. As shown in Figures 7 and 8, for U.S. and Canada respectively, 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 both the U.S. and Canada 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, that the two year integral multipliers are less than one for the state-dependent case for U.S. data, and there is no significant difference between the multipliers in the low unemployment state and the high unemployment state (see Table 3). The results for Canada are also very similar to the baseline case (see the top part of Table 4) and the multiplier is higher in the high unemployment state compared to the low unemployment state.

Next, we consider a threshold based on the moving average of output growth, instead of the unemployment rate. This is to help compare our findings with previous literature that considers state-dependence based on recessions and expansions. We construct a smooth transition threshold similarly to AG-12, where we replace the dummy variable I_{t-1} in equation (2) with the function $F(z_t)$, where z is the 7-quarter moving average of output growth. We also use the same definition of $F(z)$ as AG-12, which is given in (4).[18] Figure 9 shows the function $F(z)$ along with the NBER recessions for the US and Figure 10 shows the analogous function $F(z)$ for Canada, for our full samples.[19] Results in the bottom part of Table 3 show that when we use this weighting function for recessionary regimes in our specification to construct state-dependent multipliers, we still get multipliers typically less than one for U.S. across both recession and expansion regimes, and do not find any con-

18. The choice of $\gamma = 3$ ensures that $F(z)$ is greater than 0.8 close to 30 percent of the time for the U.S., which lines up with the total duration of recessions, as classified by the NBER business cycle dating committee, for the sample under consideration, starting in 1889. The same value of $\gamma = 3$ ensures that $F(z)$ is greater than 0.8 close to 20 percent of the time for Canada, which lines up with the total duration of recession for Canada in the full sample considered.

19. For the overlapping period, starting in 1949, the U.S. figure is essentially a replication of Figure 1 in AG-12.

vincing evidence of higher multipliers in expansions versus recessions. The only exception is the case in which the two year multiplier is found to be much larger in expansions than in recessions (where it is negative). For Canada, our baseline results are preserved and the multipliers are much larger in expansion, both compared to the baseline case and also recessions, exceeding 1 in the recessionary state under one definition.

Another point of departure with the pre-existing literature is the fact that most of the papers employ a shorter data sample that spans the post World War II period. As a robustness check we limit our sample to this period, 1948-2011, and employ the Jorda local projection method on this data set. In this shorter sub-sample too, about 30 percent of the observations are above our baseline threshold for unemployment rate, signifying state of slack.[20] As shown in the middle part of Table 3, in the linear case, the multipliers for U.S. are still smaller than 1. Looking at state-dependent multipliers, we find that the peak multiplier is comparable across the high and low unemployment states but the two year integral multiplier is large and negative at -2 and the four year integral multiplier is large and positive taking a value of 18! Since the military news variable has very low instrument relevance during slack periods in the post-WWII period, the impulse responses in this state are very imprecisely estimated. Also, rather counter-intuitively in this sub-sample, output has a negative response to the news shock in the high unemployment state, and the government spending response also becomes negative after 2-3 years. Thus, it is hard to take these state-dependent multipliers for the sub-sample seriously. For Canada, in the same post World War II sub-sample, the multiplier is larger than 1 in the linear case, in contrast to the full sample where it was close to 0.5. Across the two states, the multiplier under high unemployment for Canada also fluctuates between very large positive and negative values across definitions. Here too, the caveat remains for Canada as well that in this post World-War II sample, in the high unemployment state, the F-statistics for the instrumental relevance of the news variable in the slack state are barely above 1.

4.4 Comparison to Existing Results

While we find differences in multipliers across states for Canada, we do not find differences for the U.S. This result is contrary to a number of results from the literature that find higher multipliers when the economy is in a low growth state.

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

We explored some potential sources of differences in the previous section, such as choice of threshold and the sample period. To further understand the differences, we begin by applying the Jorda method to AG-12 post-WWII data for output, taxes and government spending, using their exact definition of states, and their identification of shocks.[21] Figure 11 shows the linear responses in the top panel which look fairly similar to the linear case in AG-12. However, the state-dependent responses shown in the lower panel look very different from them. Notice that the state-dependent responses for both government spending and output essentially lie on top of one another and are not significantly different across expansions and recessions, and neither are the resulting multipliers.

Thus, the difference between our results and those of AG-12 is not due to the sample period or the definition of slack, but rather due to the econometric method for estimating the model and calculating the impulse responses. As we discussed in an earlier section, AG-12 calculate their impulse responses under the assumption that the economy stays in its current state for the 20 quarters over which they compute their multiplier. As Figure 9 shows, during the post-WWII period the episodes of low growth states are much shorter than 20 quarters in duration, so the assumption is not consistent with the data. Also, the assumption implies that a positive shock to government spending during a low-growth state does not help the economy escape that state. Yet this very assumption produces impulse responses that imply multipliers greater than two in a recession state. Thus, the crucial assumption used to calculate the multipliers - that government spending does not help the economy escape the low state - is at odds with the conclusion that government spending has powerful stimulus effects on output.

To determine how this assumption affects their results, we compute alternative impulse responses using their data, their measure of slack, and their method for estimating the model. The left panel of Figure 12 computes the impulse responses using the Koop et al. (1996) method discussed in an earlier section, but restricting the feedback from the responses to the transition function, using $\gamma = 3$ in equation 4 and neglecting the exogenous variables term in the computation of the responses. For this experiment, we partition the histories into times that correspond to $F(z_{t-1}) \geq 0.8$ (recession) and $F(z_{t-1}) < 0.2$ (expansion).[22] For each of these sets of histories, we compute (7) setting $F(z_{t-1}) = 1$ and

21. As discussed earlier, AG-12 use a smooth transition threshold based on a 7 quarter moving average of output growth. We use exactly the same data and the threshold variables in their codes provided to us kindly by Yuriy Gorodnichenko. Also, they identify the government spending shock using Cholesky decomposition with government spending ordered first. We introduce this identification scheme in the Jorda framework by considering the coefficients of log of government spending at time t.

22. These cutoffs for the histories are consistent with the definition of the recessions in AG-12 and are more

$F(z_{t-1}) = 0$, respectively. Because of the restrictions on the feedback from the response to the state of the economy, the model is linear, conditional on the value of $F(z_{t-1})$, and becomes history independent and the scale of the shock. Future innovations are drawn appropriately from the either Ω_R or Ω_E, respectively. The shock to spending is a Blanchard and Perotti (2002)-type shock where the Wold causal chain has government spending ordered first in the VAR. In the expansion regime, output rises on impact but then declines, eventually becoming negative. Apart from a scaling of the output response, these figures are identical to those that are obtained using the regime-dependent responses in AG-12.[23] This is expected since shutting off the feedback makes the responses history independent.

The right panel of Figure 12 show the impulse responses allowing the feedback into the transition variable. In this case, the period-t value of the variance-covariance matrix is used to draw future innovations. Because there is feedback, the period-t values of the variance-covariance matrix with and without the shock (that is, Ω_t will not be the same for both terms in (7)) will differ. We approximate the smooth transition heteroskedasticity by drawing

$$\begin{aligned} e_{1,t+h} &\sim N(0, \Omega_1), \\ e_{2,t+h} &\sim N(0, \Omega_2), \end{aligned}$$

and computing the expectations using

$$\begin{aligned} \varepsilon_{t+h}^{shock} &= \left(1 - F\left(z_{t+h-1}^{shock}\right)\right) e_{1,t+h} + F\left(z_{t+h-1}^{shock}\right) e_{2,t+h}, \\ \varepsilon_{t+h}^{noshock} &= \left(1 - F\left(z_{t+h-1}^{noshock}\right)\right) e_{1,t+h} + F\left(z_{t+h-1}^{noshock}\right) e_{2,t+h}. \end{aligned}$$

Allowing for the feedback from the output response into the regime dramatically alters the path of output and the path of government spending. Starting from either regime, the response of government spending to its own shock is persistent. Output increases over the duration of the response horizon. Unfortunately, these generalized impulse responses cannot be easily interpreted as a convex combination of the regime dependent responses.

likely to provide evidence of differences across regimes. Results partitioning the histories at $F(z_{t-1}) \geq 0.5$ and $F(z_{t-1}) < 0.5$ were similar.

23. AG-12 scale this shock but we leave the shock unscaled. We found that the scaling factor makes the shock to government spending large enough that the economy switches states instantaneously, regardless of history. In this case, the responses across regimes are essentially identical.

However, the GIRFs are similar across regimes, suggesting that the feedback is important to determining how the economy responds to shocks.[24]

The left panel of Figure 13 show the responses of the three variables to an output shock using the regime-dependent method; the right panel of Figures 13 show the same responses using the GIRFs. The main noticeable difference between the two methods is the government spending response to an output shock. If computed holding the regime constant, government spending declines during a recession but eventually rises. If output feeds back into the state of the economy, spending does not increase over the length of the response horizon.

There are a few other papers using GIRFs to estimate the response of output to government spending shocks across business cycle regimes. Fazzari et al. (2013) and Baum et al. (2012) compute GIRFs from threshold autoregressions (TARs) with different variables to determine whether there are differences in the multiplier across regimes.[25] The TARs are different from the STAR model of AG-12 in that the economy moves discretely rather than smoothly from regime to regime but still allow feedback to the regime.[26] The model still produces feedback, requiring the use of the GIRFs rather than, say, regime-dependent impulse responses. Both papers find differences in the point values of the multipliers; however, error bands for the GIRFs overlap making the differnces across regimes likely statistically negligible. Fazzari et al. (2013) also test for and find breaks in the VAR coefficients. However, because the impulse responses are highly nonlinear functions of the coefficients, this does not necessarily mean that said breaks will results in statistically relevant differences in the multiplier.

As discussed earlier, Auerbach and Gorodnichenko (2013) (henceforth AG-13) also apply the Jordà method in their analysis of a panel of OECD countries, using semiannual data from 1985 to 2008. They continue to find much higher multipliers during recessions in contrast to our findings. In addition to the obvious difference in time period and country sample, the most likely reason for the difference is in two details of how they calculate multipliers. First, following the standard practice, they estimate everything in logarithms and then use the *ex post* conversion factor based on average Y/G during their sample to

24. AG-12 attempt to account for the feedback in one experiment by computing a response for each history and allowing the state to change over time.

25. Mittnik and Semmler (2012) and Semmler and Semmler (2013) estimate bivariate TARs, using GIRFs to determine the responses. However, their models do not include government spending explicitly. The regime dynamics in the multiplier is inferred.

26. Bognanni (2013) uses a Markov-switching VAR with regime-dependent responses and finds no significant differences across regimes.

convert elasticities to multipliers. Second, they depart from the rest of the literature by comparing the path or peak of output to the *impact* of government spending rather than to the peak or integral of the path of government spending.[27] This is a big difference because the effects of a shock to government usually build up for several quarters. Tist is also not the type of multiplier policy makers are interested in because it does not count the average cumulative cost of government spending associated with the path.

We now demonstrate the difference these two details make by applying their method for calculating multipliers to our U.S. historical sample. First, we estimate our baseline model in logs and use the sample average of Y/G of 8 to convert the elasticities. This results in 2-year integral multipliers of 1.9 in the high unemployment state and 1.8 in the low unemployment state. Thus, this change in method *doubles* the constructed multiplier relative to our baseline method, but does not lead to a difference across states. Second, still using the log specification and the conversion factor, we instead calculate multipliers by dividing the average response of output over the 2-year horizon by the the *initial* shock to government spending. In this case, we calculate 2-year multipliers of 9.9 for the high unemployment state and 4.7 for the low unemployment state. Thus, this method not only produces multipliers that are huge in both states, but also induces a large difference between states. As our figures make clear, this difference shows up because this calculation does not take into account the fact that government spending rises more robustly after an initial shock during a recession. Thus, it is clear that even using the same estimation method and same method for computing impulse responses, details of the calculations of multipliers can make a big difference.

4.5 The Behavior of Taxes

Our analysis so far has ignored the responses of taxes. Romer and Romer's (2010) estimates of tax effects indicate very significant negative multipliers on taxes, on the order of -2 to -3. Thus, it is important for us to consider how tax policy differs across states in order to understand our multiplier results.

We estimate our basic model with tax rates and deficits. Tax rates are computed as the ratio of federal receipts to nominal GDP, and represent average tax rates rather than marginal tax rates. We substitute tax rates for the real tax revenues in our baseline specification so that we can distinguish increases in revenues caused by rising output versus

27. This is also true for the first column of Table 1 in AG-12

rising rates. The deficit is the real total deficit. We include four lags of the logs of these two new variables along with GDP and government spending as controls all of the regressions. Figure 14 shows the results from the linear case for the U.S. The responses of government spending and GDP are almost identical to the baseline case. The bottom panels show that both average tax rates and the deficit increase, with deficits increasing more rapidly at the beginning and tax rates rising more slowly and more persistently. Taking the ratio of the path of deficits to government spending, we estimate that most of the increase in government spending during the first year is financed by deficits. The deficit fraction of government spending hits a peak at quarter 4. It stays high for about a year and then begins to decline.

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

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

$$1 = \beta E_t \left[\frac{u_{n,t+1}}{u_{n,t}} \frac{(1-\tau_t) w_t}{(1-\tau_{t+1}) w_{t+1}} (1+r_t) \right] \tag{8}$$

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

It is easy to show in a standard neoclassical model that the path of taxes we observe in the impulse responses results in a multiplier that is higher in the short-run but lower in the long-run relative to the lump-sum tax case. We have also conducted this experiment in the

Gali et al. (2007) model where 50 percent of the households are rule-of-thumb consumers. We found the same effect in that model as well. Drautzburg and Uhlig (2011) analyze an extension of the Smets-Wouters model and find that the timing of distortionary taxes is very important for the size of the multiplier. Given the impulse response of tax rates, and with these theoretical results in mind, it is very possible that our estimated multipliers are greater than we would expect if taxation were lump-sum.

Does the financing of government spending differ across states? Figure 15 shows the state-dependent results for the U.S. As we showed before, both government spending and GDP rise more if a news shock hits during a slack state, even after adjusting the initial size of the shock. The bottom panels show that tax rates and deficits also rise more during recessions, but there are other interesting differences in the patterns. When we study the ratio of the deficit to government spending at each point in time along the path, we find that more of government spending is financed with deficits when a shock hits during a slack state. For example, at quarter five the ratio of the deficit to government spending is 86 percent if a shock hits during a slack state but only 29 percent if the shock hits during a non-slack state. Comparing the peak of the deficit to the peak of the government spending response, the ratio is 0.6 for the slack state and 0.4 for the no slack state. Thus, on average the short-run government spending is financed more with deficits if the shock hits during a slack state.

Figures 16 and 17 show the results for Canada. In Canada, the responses are more erratic, but we can still distinguish a different pattern. It appears that a higher portion of government spending is financed by taxes in the short-run, but then the deficit rises more after the first two years. This differential behavior of taxes could explain why we find lower multipliers for Canada than for the U.S. Comparing across states, the ratio of the peak deficit to the peak government spending is 0.5 if the shock hits during a slack state and 0.4 if it hits during a non-slack state. Further looking at the deficit fraction of government spending for the first two years, it is much higher in the high unemployment state than the low unemployment state.

5 Multipliers at the Zero Lower Bound

5.1 Literature

Several recent papers have analyzed the effects of fiscal policy in New Keynesian models when the zero lower bound on interest rates prevents nominal interest rates from responding according to the Taylor rule. For example, Eggertsson and Woodford (2003), Eggertsson (2011), and Christiano et al. (2011) show that the government spending multiplier can be much larger if interest rates are at the zero lower bound. The intuition for why the zero lower bound can raise the multiplier is as follows. A deficit-financed increase in government spending leads expectations of inflation to rise. When nominal interest rates are held constant, this increase in expected inflation drives the real interest rate down, spurring the economy. Christiano et al. (2011) show that if interest rates are held constant for 12 quarters and government spending goes up during this time, the multiplier peaks at 2.3. Fernández-Villaverde et al. (2012) take into account the inherent nonlinearities at the ZLB in their analysis, but still find that the government spending multiplier can be three times greater at the zero lower bound.

Very few papers have attempted to test the predictions of the theory empirically. As far we know, only two examples exist. Ramey (2011) estimates her model for the U.S. over the subsample from 1939 through 1951 and shows that the multiplier is no higher during that sample. Crafts and Mills (2012) construct defense news shocks for the U.K. and estimate multipliers on quarterly data from 1922 through 1938. They find multipliers below unity even when interest rates were near zero.

In all of the theoretical models, it is not the zero lower bound per se, but rather the fact that nominal interest rates stay constant rather than following the Taylor rule. 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. As discussed above, we include the classic ZLB periods of the 1930s and the last several years, but we also include WWII through the Treasury Accord in March 1951, since the Federal Reserve was very accommodative of fiscal policy during that time.

5.2 Results

To determine whether multipliers are different at near the ZLB, we estimate our state-dependent model just as we did for the slack indicator. In this case, however, we define the state as being at or near the zero lower bound. We consider our full sample spanning 1889-2011 for the U.S.[28] Figure 18 shows the impulse responses. The results suggest that government spending responds more slowly, but more persistently during ZLB states than in normal states. The difference in GDP responses follow this pattern, but in a muted way. Table 5 shows the multipliers in each state, computed the three different ways. For the peak multiplier, we find that the multiplier is slightly higher in the normal state, though both are less than unity. For the two-year integral, the multipliers are very similar. Oddly, for the 4-year integral, the multiplier is very high at 1.6 in the normal state, but less than one in the ZLB state. Thus, in no case do we find evidence of significantly higher multipliers during periods at the zero lower bound or constant interest rates.

6 Conclusion

In this paper, we have explored the idea that 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 100 years for the U.S. and Canada. We considered two possible indicators of the state of the economy: the amount of slack and whether interest rates were being held constant close to the zero lower bound. Using a more robust method for estimating state-dependent impulse responses and better ways of calculating multipliers from them, we provided numerous estimates of multipliers across different specifications.

Our results can be summarized as follows. We find no evidence of significant differences in multipliers according to whether the U.S. economy is experiencing substantial slack as measured by the unemployment rate. Most U.S. multipliers are slightly below unity with a few slightly above unity. For Canada, we find substantial differences in multipliers depending on the unemployment rate. For Canada, multipliers are between 0.5 and 1.1 during periods of high unemployment but only -0.1 to 0.2 during periods of low unemployment. In our robustness checks, we obtain a few large negative and positive values of multipliers, but we ascribe those to the effects of shortening the sample on the precision of the estimates.

In our analysis of states distinguished by interest rates being near the zero lower bound

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

in the U.S., we found that multipliers were never higher at the zero lower bound. Thus, we found no support for the prediction of recent New Keynesian models.

Of course, our results come with many caveats. As discussed in the introduction, we are forced to use data determined by the vagaries of history so we do not have a controlled experiment. Because we use news about future military spending as our identified shock, our results do not inform us about the size of multipliers on transfer payments or infrastructure spending, or whether they vary across states. The results might be affected by other aspects of war-time periods. Moreover, because the episodes we studied were characterized by certain paths of taxes, the results are not immediately applicable to the case of deficit-financed stimulus packages or fiscal consolidations.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Ehrmann, Michael, Martin Ellison, and Natacha Valla, 2003. "Regime-Dependent Impulse Response Functions in a Markov-Switching Vector Autoregression Model." *Economics Letters* 78(3): 295–299.

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

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

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

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

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

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

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

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

Marcellino, Massimiliano, James H. Stock, and Mark W. Watson, 2006. "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series." *Journal of Econometrics* 135(1-2): 499–526.

McInnis, Marvin, 2001. *Historical Canadian Macroeconomic Dataset 1871-1994*. http://hdl.handle.net/1974/7669: Queens University.

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

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

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

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

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

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

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

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

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

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

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

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

Table 1. Tests of Instrument Relevance

	F-statistic	Number of observations
U.S.		
1891:1 - 2011:4	12.32	484
1891:1 - 2011:4 - Slack	7.57	172
1891:1 - 2011:4 - No slack	9.40	312
1891:1 - 2011:4 - ZLB	2.25	89
1891:1 - 2011:4 - Normal	19.49	395
1948:1 - 2011:4	20.26	256
1948:1 - 2011:4 - Slack	1.10	74
1948:1 - 2011:4 - No slack	18.32	182
1948:1 - 2011:4 - ZLB	15.2	26
1948:1 - 2011:4 - Normal	10.1	230
Canada		
1913:2 - 2011:4	9.60	395
1913:2 - 2011:4 - Slack	9.31	210
1891:1 - 2011:4 - No slack	4.55	185
1948:1 - 2011:4	10.41	256
1948:1 - 2011:4 - Slack	1.16	146
1948:1 - 2011:4 - No slack	5.43	110

Note: The F-tests are the joint significance of news variables in a regression of log real per capita government spending on its own four lags, four lags of log real per capita GDP, current and four lags of news (scaled by lagged GDP), and a quartic time trend.

Table 2. Estimated multipliers: Considering slack state

	Linear Model	High Unemployment	Low Unemployment	P-value for difference in multipliers across states
U.S.				
Peak	0.92	0.82	1.15	
	(0.462)	(0.351)	(0.696)	0.645
2 year integral	0.78	0.79	0.87	
	(0.118)	(0.131)	(0.184)	0.758
4 year integral	0.87	0.80	1.11	
	(0.109)	(0.095)	(0.181)	0.209
Canada				
Peak	0.42	0.51	0.21	
	(0.326)	(0.172)	(0.305)	0.355
2 year integral	0.50	1.08	0.09	
	(0.122)	(0.222)	(0.111)	0.000
4 year integral	0.58	0.87	-0.07	
	(0.112)	(0.096)	(0.099)	0.000

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

Table 3. Robustness checks of multipliers for U.S.: Considering slack state

	Linear Model	High Unemployment	Low Unemployment
HP filtered time-varying threshold (with $\lambda = 10^6$)			
Peak	0.92	0.87	1.08
2 year integral	0.78	0.89	0.82
4 year integral	0.87	0.82	0.96
Subsample: 1948-2010			
Peak	0.94	0.99	1.10
2 year integral	0.58	-2.34	0.78
4 year integral	0.79	18.51	1.05
	Linear	Recession	Expansion
7 qtr. moving avg. output growth, $F(z)$			
Peak	0.92	0.58	0.98
2 year integral	0.78	-0.49	0.80
4 year integral	0.87	1.34	0.90

Table 4. Robustness checks of multipliers for Canada: Considering slack state

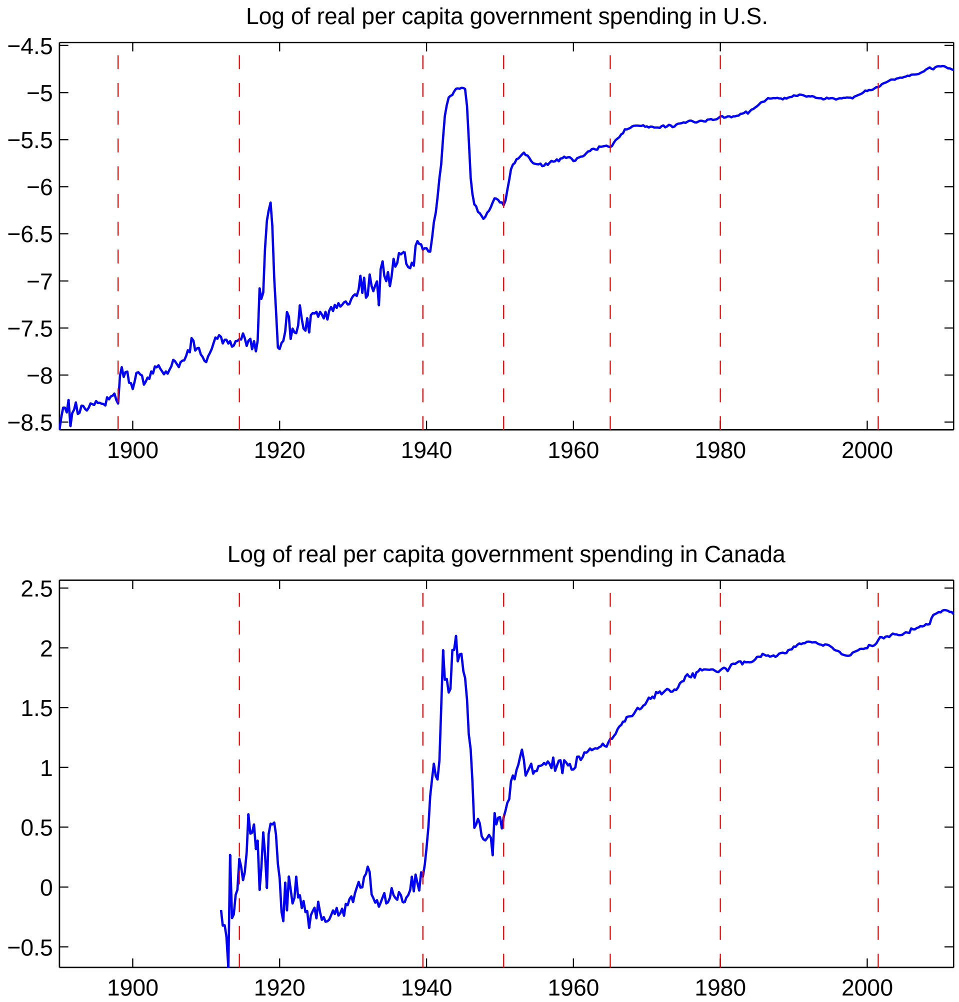
	Linear Model	High Unemployment	Low Unemployment
HP filtered time-varying threshold (with $\lambda = 10^6$)			
Peak	0.43	0.57	0.30
2 year integral	0.50	1.39	0.19
4 year integral	0.58	1.00	0.12
Subsample: 1948-2010			
Peak	1.31	5.49	0.99
2 year integral	1.95	-4.44	1.41
4 year integral	1.24	14.36	0.40
	Linear	Recession	Expansion
7 qtr. moving avg. output growth, $F(z)$			
Peak	0.44	0.58	0.27
2 year integral	0.54	1.40	0.23
4 year integral	0.59	0.90	-1.36

Table 5. Estimated multipliers: Considering zero lower bound

	Linear Model	Near Zero Lower Bound	Normal
U.S.			
Peak	0.92	0.71	0.87
2 year integral	0.78	0.78	0.73
	(0.118)	(0.172)	(0.130)
4 year integral	0.87	0.73	1.60
	(0.109)	(0.113)	(0.304)

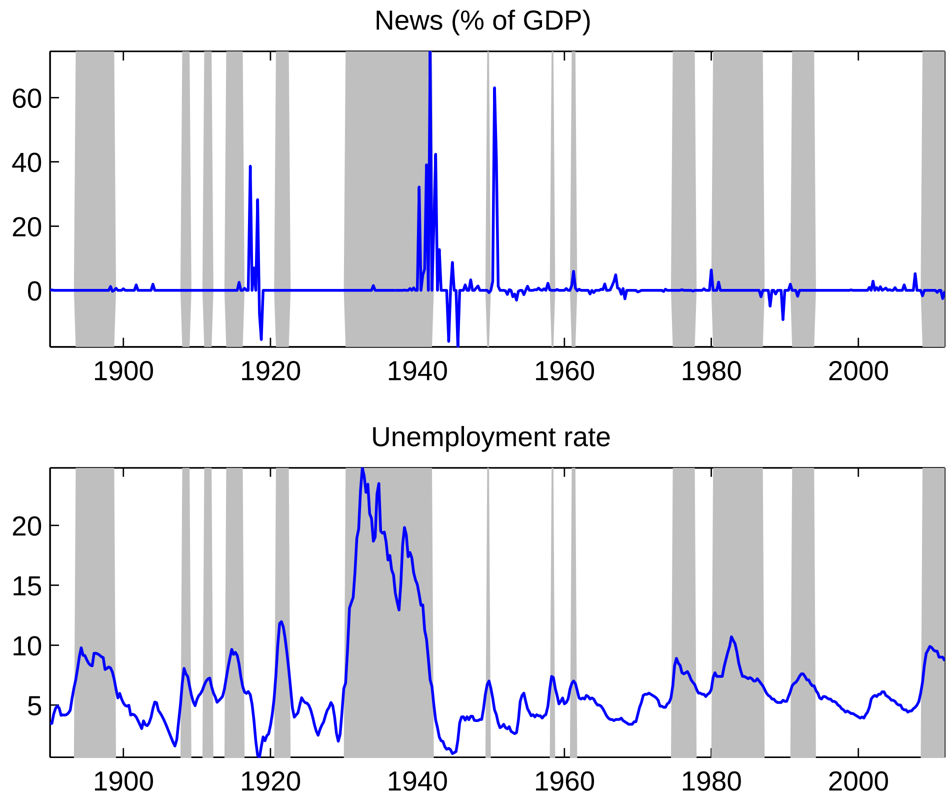
Note: The values in brackets under the multipliers give the standard errors.

Figure 1. Government Spending for US and Canada



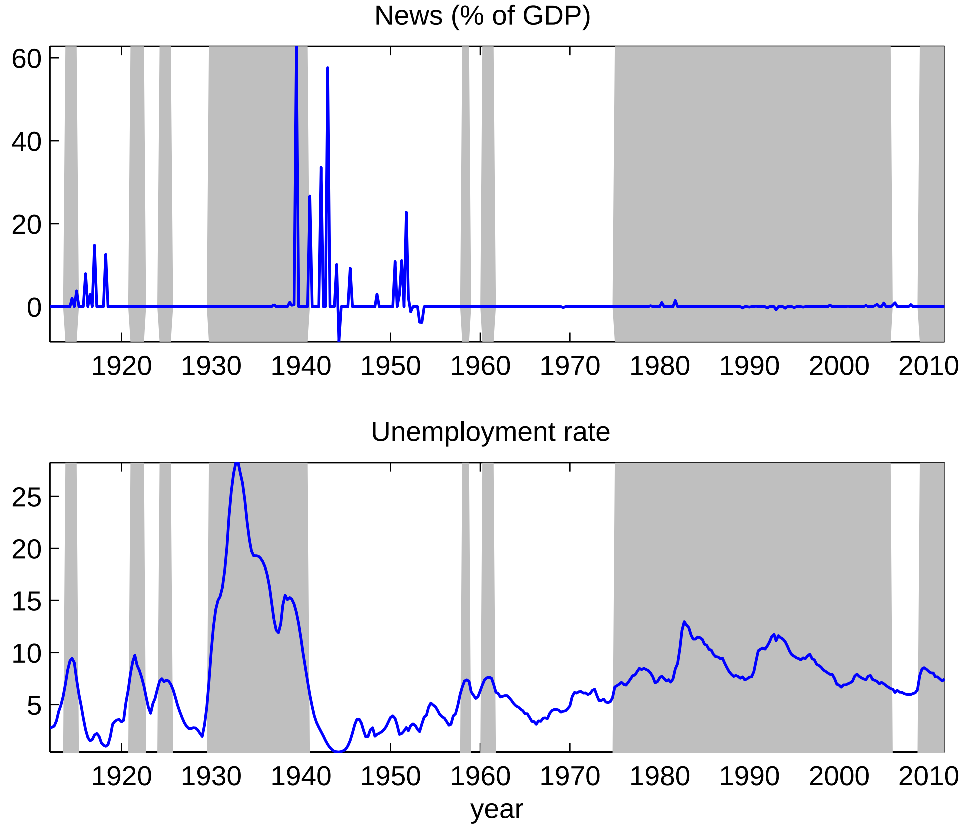
Note: The vertical lines indicate major military events: 1898q1(The Spanish-American War starts with the sinking of the USS Maine), 1914q3 (WWI starts), 1939q3 (WWII starts), 1950q3 (Korean War starts), 1965q1 (Vietnam War starts), 1980q1 (Buildup in response to Soviet invasion of Afghanistan), 2001q3 (9/11).

Figure 2. Military spending news and unemployment rate for US



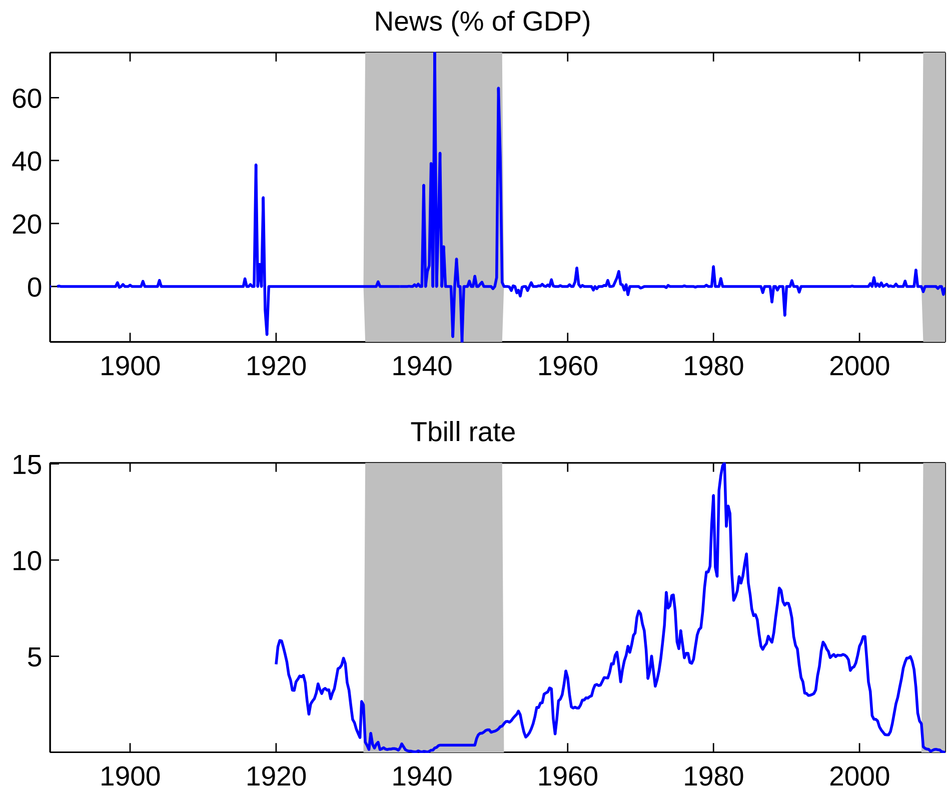
Note: Shaded areas indicate periods when the unemployment rate is above the threshold of 6.5%.

Figure 3. Military spending news and unemployment rate for Canada



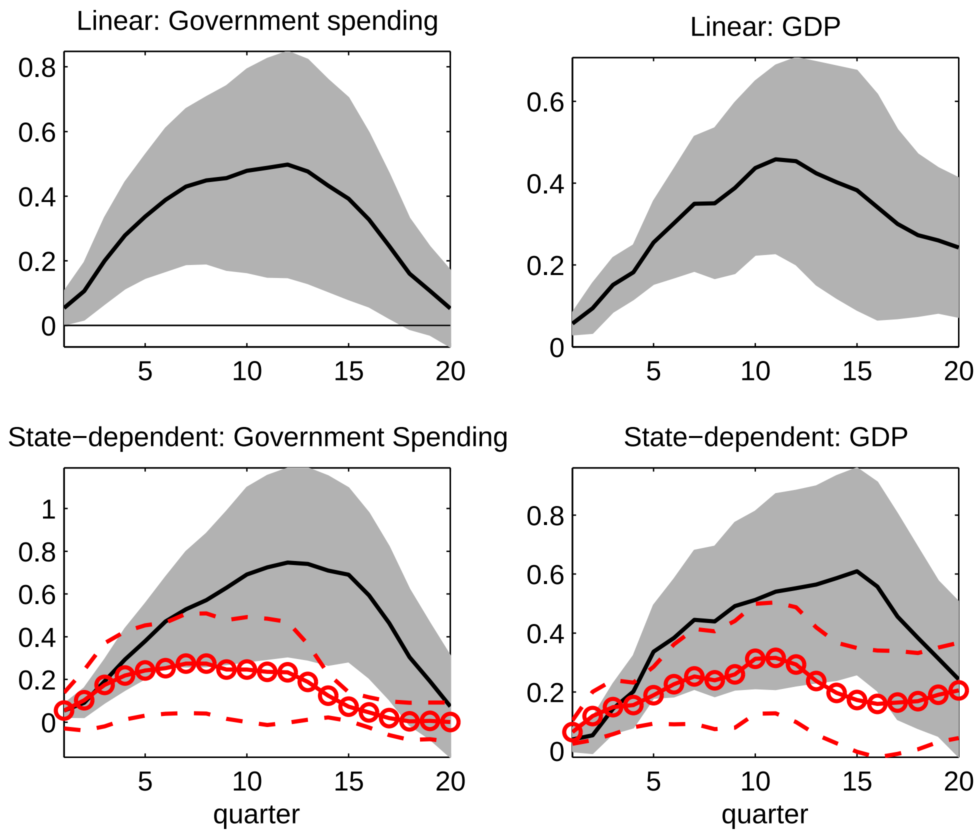
Note: Shaded areas indicate periods when the unemployment rate is above the threshold of 6.5%.

Figure 4. Military spending news and interest rate for US



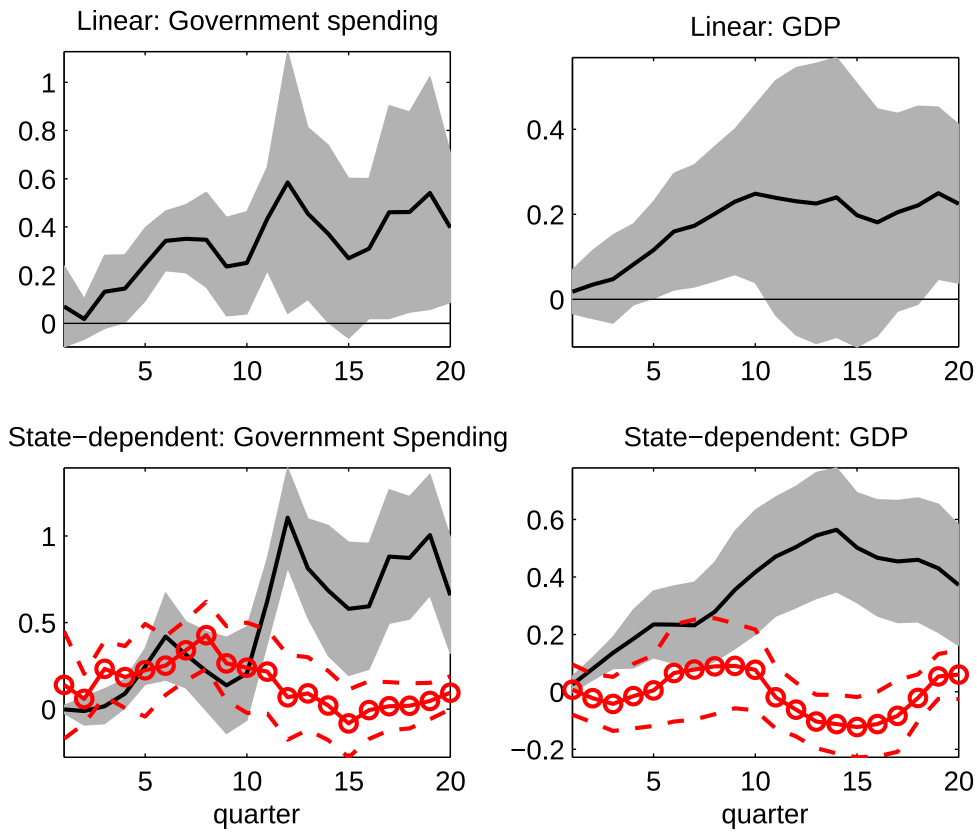
Note: Shaded areas indicate periods when the interest rate is close to or at the zero lower bound.

Figure 5. Government spending and GDP responses to a news shock for US: Considering slack state



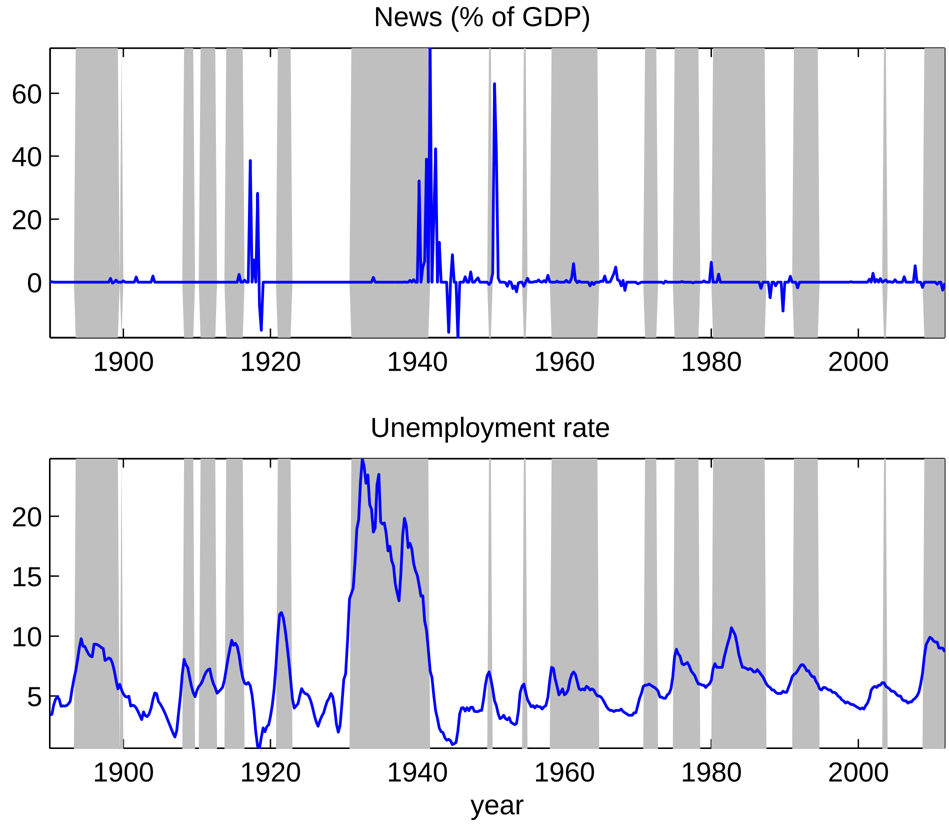
Note: US response of government spending and GDP to a news shock equal to 1% of GDP. The top panel shows the responses in the linear model. The bottom panel shows the state-dependent responses where the black solid lines are responses in the high unemployment state and the lines with red circles are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

Figure 6. Government spending and GDP responses to a news shock for Canada: Considering slack state



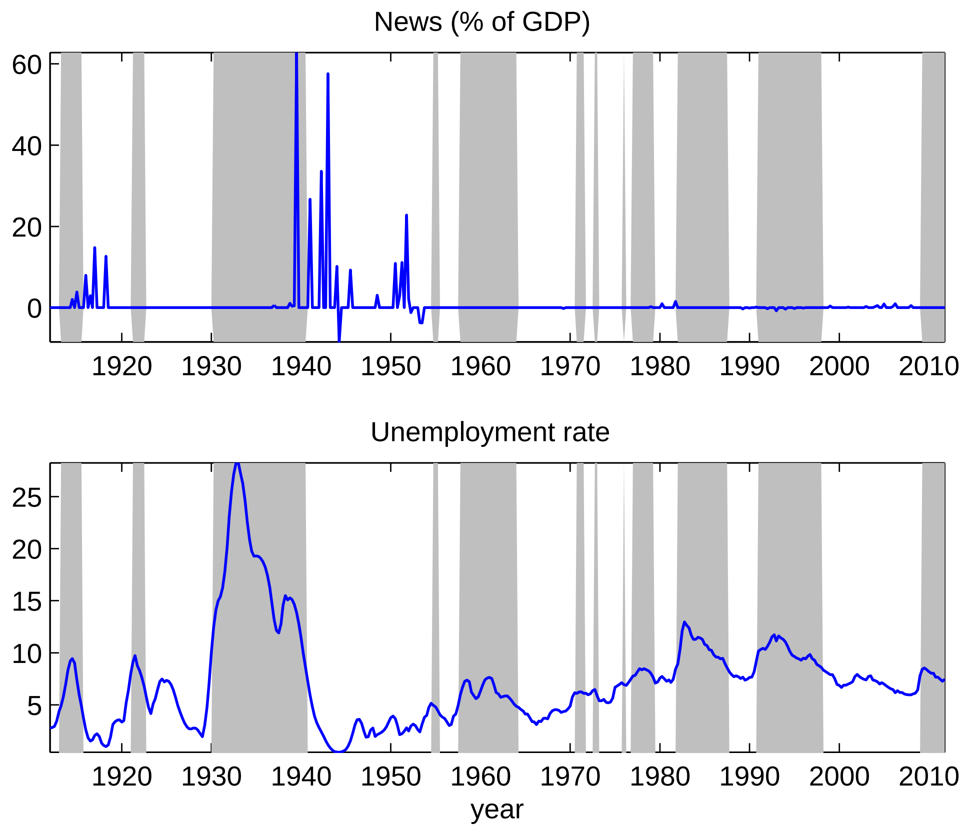
Note: Canada response of government spending and GDP to a news shock equal to 1% of GDP. The top panel shows the responses in the linear model. The bottom panel shows the state-dependent responses where the black solid lines are responses in the high unemployment state and the lines with red circles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

Figure 7. Robustness check: New threshold of unemployment rate based on time-varying trend for US



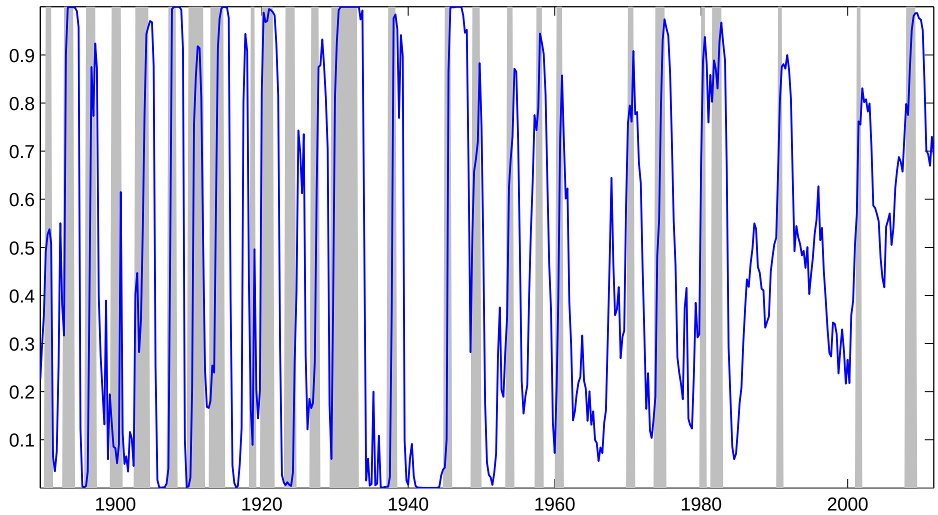
Note: Shaded areas indicate periods when the unemployment rate is above the time-varying trend based on HP filter with $\lambda = 10^6$.

Figure 8. Robustness check: New threshold of unemployment rate based on time-varying trend for Canada



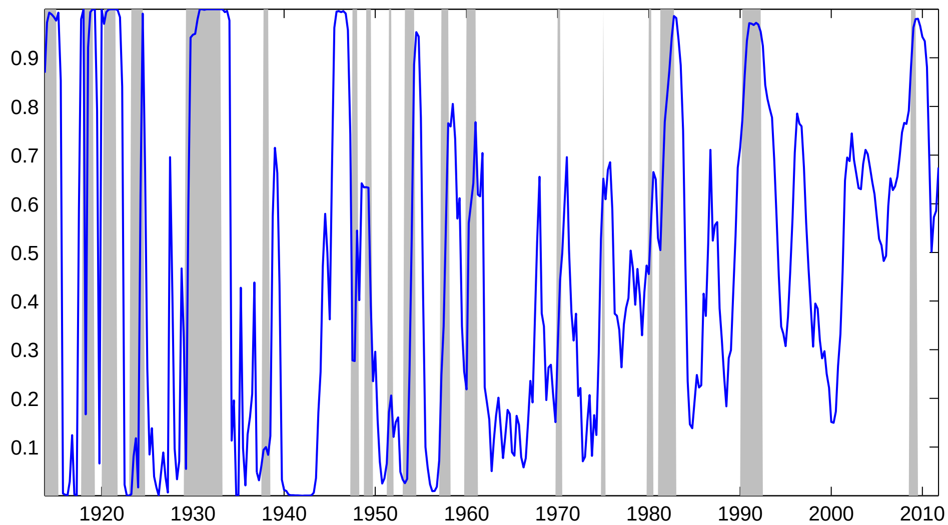
Note: Shaded areas indicate periods when the unemployment rate is above the time-varying trend based on HP filter with $\lambda = 10^6$.

Figure 9. Robustness check: New smooth transition threshold based on moving average of output growth for US



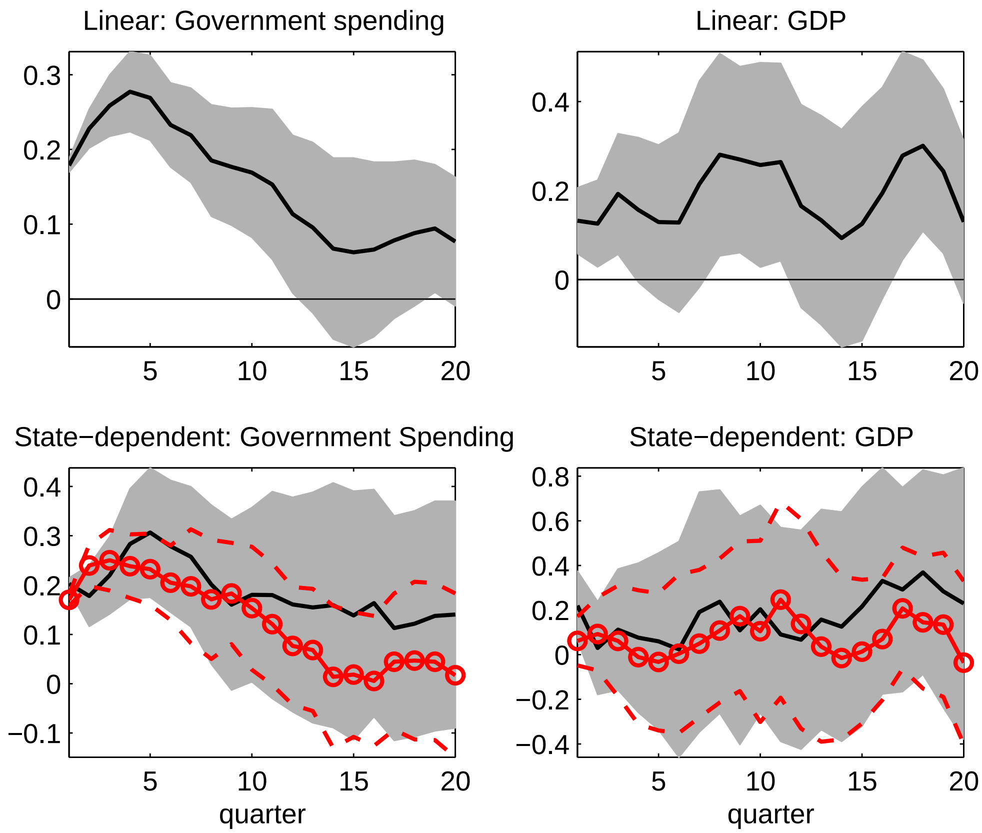
Note: The figures shows the weight on a recession regime, $F(z)$ and the shaded areas indicate recessions as defined by NBER.

Figure 10. Robustness check: New smooth transition threshold based on moving average of output growth for Canada



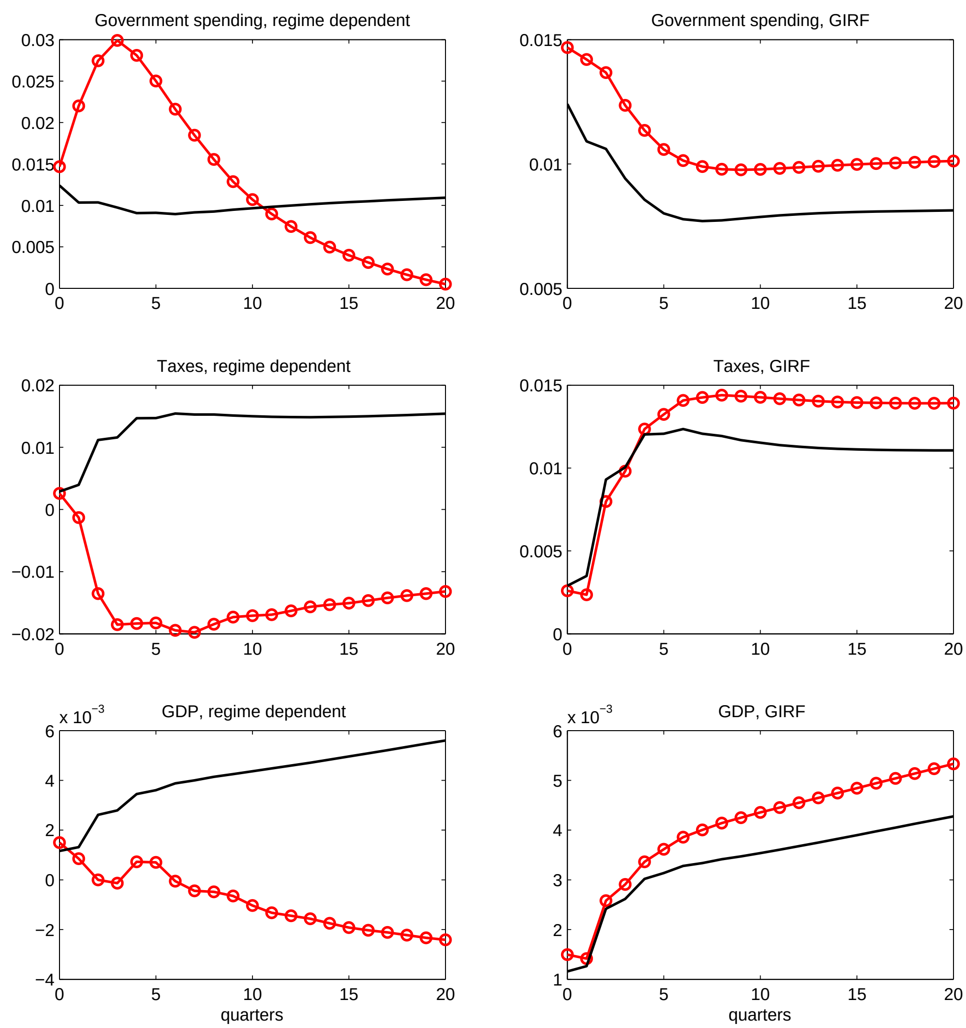
Note: The figures shows the weight on a recession regime, $F(z)$ and the shaded areas indicate official recessions.

Figure 11. Running Auerbach and Gorodnichenko (2012) with the Jorda method



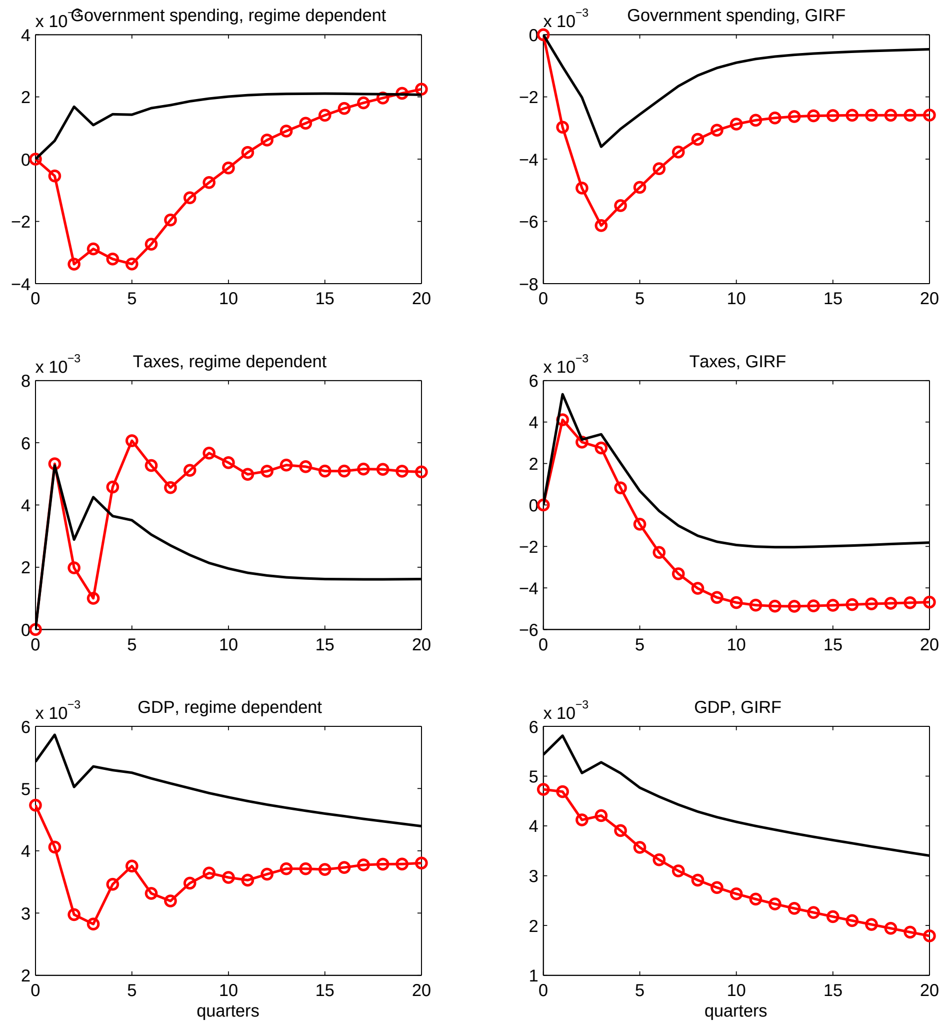
Note: US response of government spending and GDP to a government spending shock equal to 1% of GDP, with the same data, identification scheme and threshold definition as Auerbach and Gorodnichenko (2012). The top panel shows the responses in the linear model. The bottom panel shows the state-dependent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

Figure 12. Comparison of impulse response functions to a government spending shock across methodologies



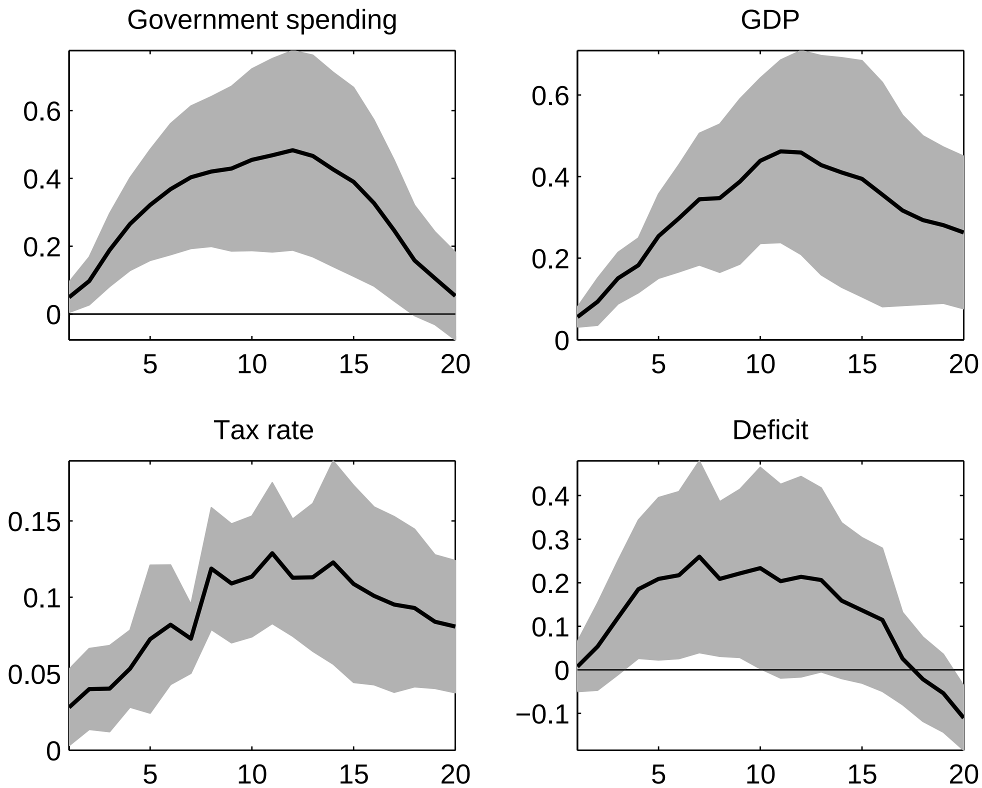
The left panel shows the regime-dependent responses of government spending, taxes, and output to a shock to government spending in an expansion (red circles) or a recession (black solid line). The right panel shows the generalized impulse responses of the same variables to the same shocks. In each case, an expansion (recession) is defined as a period t for which $F(z_{t-1}) < 0.2$ ($F(z_{t-1}) > 0.8$). Once the histories are partitioned into expansion and recession, the regime-dependent responses are computed by fixing $F(z_{t-1}) = 0$ ($F(z_{t-1}) = 1$). Generalized impulse responses as computed using 100 MC replications of future shocks and integrating over all appropriate histories.

Figure 13. Comparison of impulse response functions to a output shock across methodologies



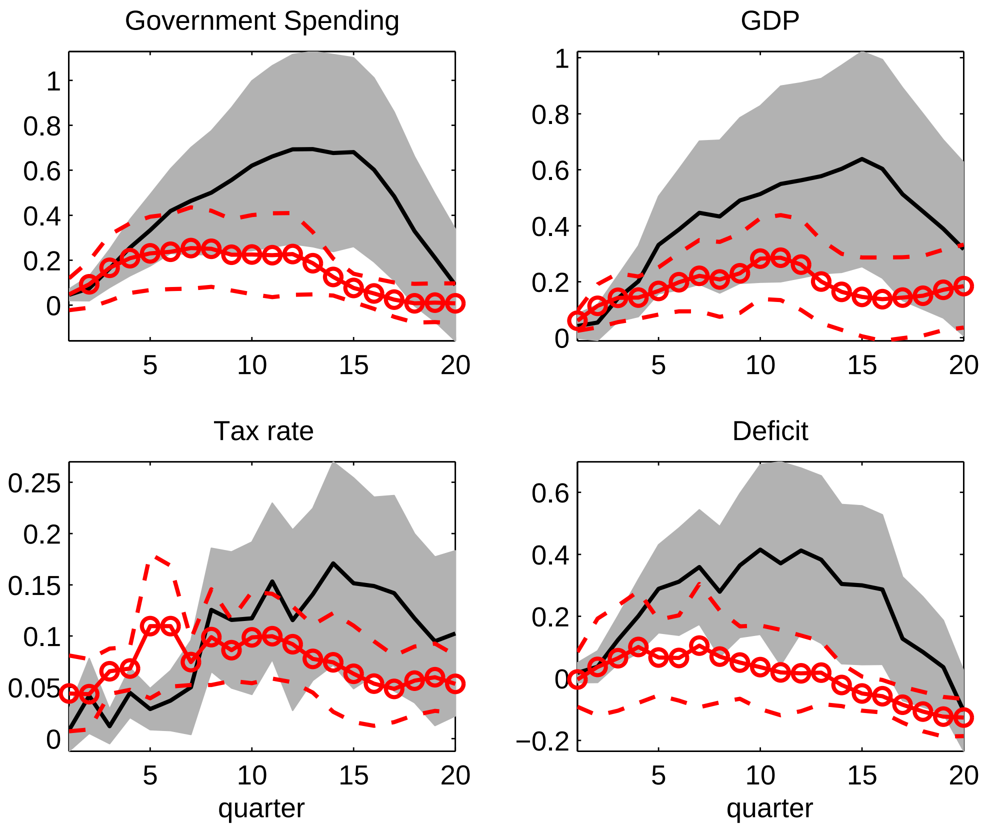
The left panel shows the regime-dependent responses of government spending, taxes, and output to a shock to output in an expansion (red circles) or a recession (black solid line). The right panel shows the generalized impulse responses of the same variables to the same shocks. In each case, an expansion (recession) is defined as a period t for which $F(z_{t-1}) < 0.2$ ($F(z_{t-1}) > 0.8$). Once the histories are partitioned into expansion and recession, the regime-dependent responses are computed by fixing $F(z_{t-1}) = 0$ ($F(z_{t-1}) = 1$). Generalized impulse responses as computed using 100 MC replications of future shocks and integrating over all appropriate histories.

Figure 14. Responses of taxes and deficits for US



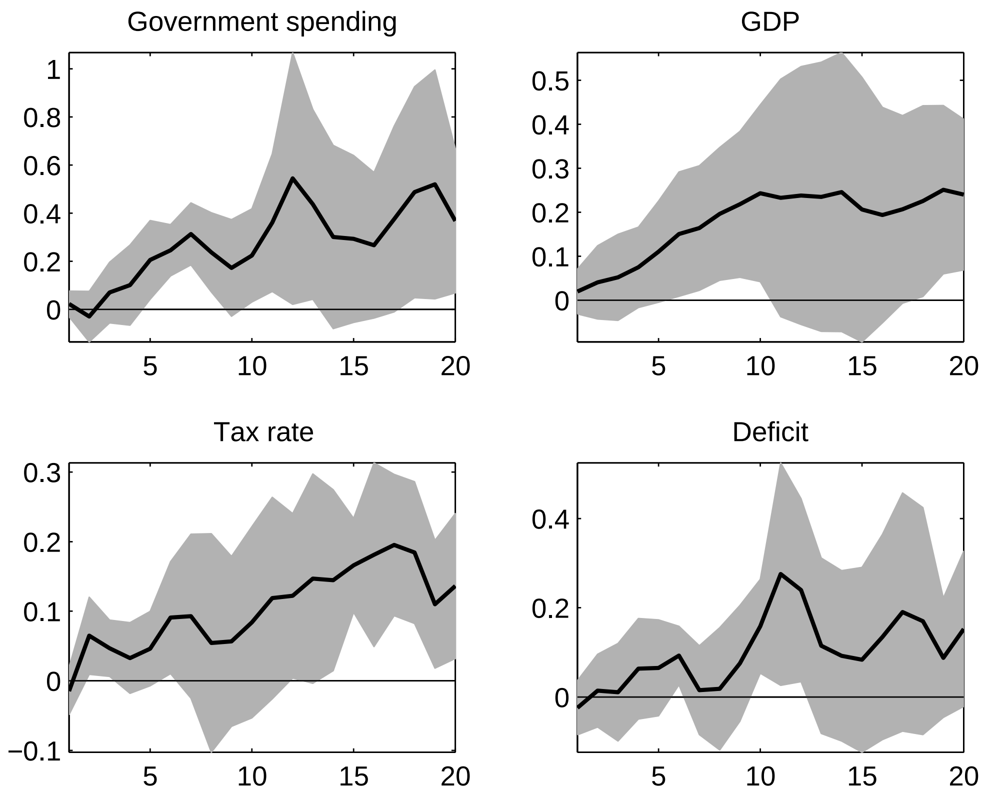
Note: These are responses for taxes and deficits for US in the linear model. The shaded areas indicate 95% confidence bands.

Figure 15. State-dependent responses of taxes and deficits for US: Considering slack state



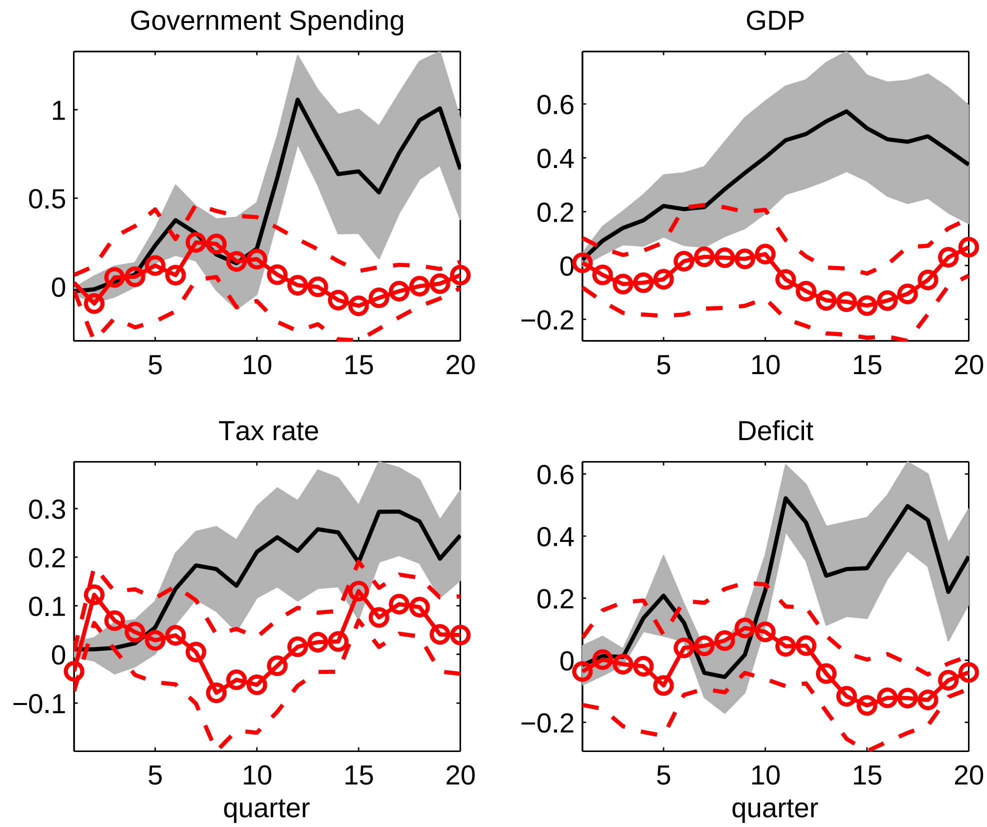
Note: These are state-dependent responses for taxes and deficits for US, where the black solid lines are responses in the high unemployment state and the lines with red circles are responses in the low unemployment state. 95% confidence intervals are also shown.

Figure 16. Responses of taxes and deficits for Canada



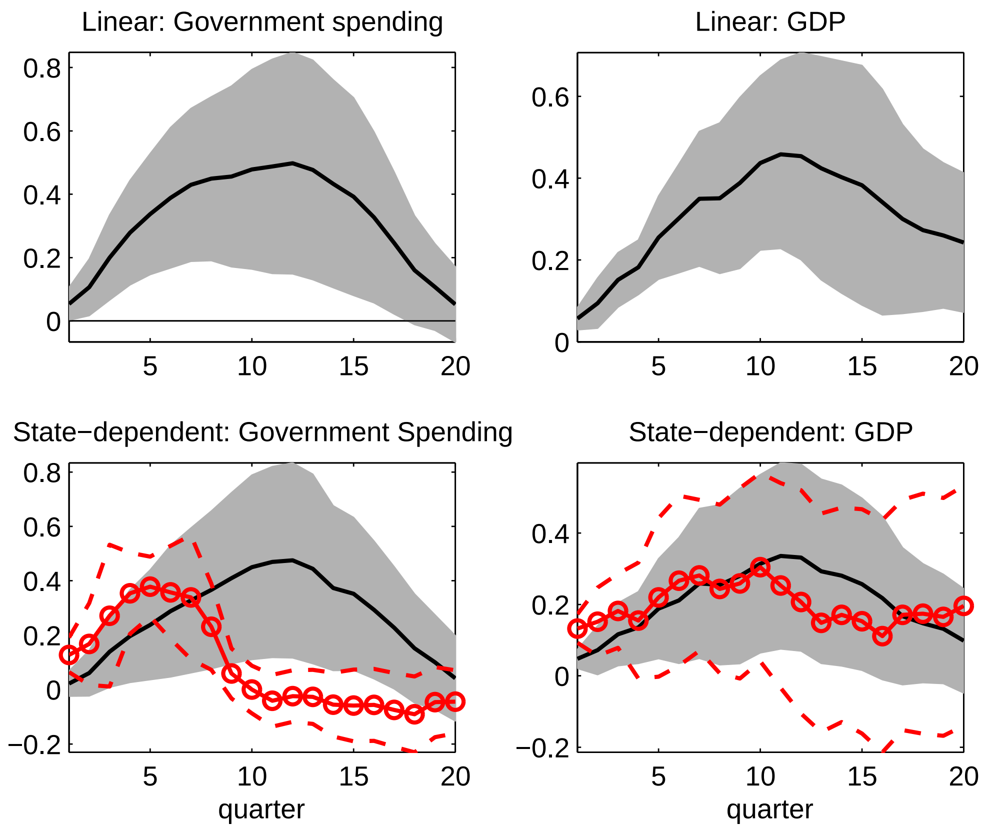
Note: These are responses for taxes and deficits for US in the linear model. The shaded areas indicate 95% confidence bands.

Figure 17. State-dependent responses of taxes and deficits for Canada: Considering slack state



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

Figure 18. Government spending and GDP responses to a news shock for US: Considering zero lower bound



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