DE GRUYTER Stud. Nonlinear Dyn. E. 2015; 19(3): 285–315

Steven M. Fazzari, James Morley and Irina Panovska*

State-dependent effects of fiscal policy

Abstract: We investigate the effects of government spending on US output with a threshold structural vector autoregressive model. We consider Bayesian model comparison and generalized impulse response analysis to test for nonlinearities in the responses of output to government spending. Our empirical findings support state-dependent effects of fiscal policy, with the government spending multiplier larger and more persistent whenever there is considerable economic slack. Based on capacity utilization as the preferred threshold variable, the estimated multiplier is large (1.6) for a low-utilization regime that accounts for more than half of the sample observations from 1967 to 2012 according to the estimated threshold level.

Keywords: Bayesian; government spending; impulse-response comparison; nonlinear dynamics; threshold model; vector autoregression.

JEL Codes: C32; E32; E62.

DOI 10.1515/snde-2014-0022

1 Introduction

The Great Recession and subsequent policy responses have reignited debate, academic and otherwise, about the stabilizing role of discretionary fiscal policy. More broadly, the dramatic economic events of recent years have stimulated new debates about the relevance of aggregate demand and government spending as possible engines of economic activity. In particular, questions have arisen about whether government spending has significant effects on aggregate output and components of output, whether there is evidence of nonlinearity in the responses of output and output components that depend on the state of the business cycle, and the economic mechanisms that drive potential asymmetries and nonlinearities.

This debate is of central importance not only for economic policy, but also for the insights it provides into the underlying structure of modern developed economies. On the one hand, according to equilibrium models in which resources are fully employed a positive shock to government spending affects output only to the extent that it changes inputs or technology. The sign of the effect could go either way.[1] These models often predict some crowding out of private investment or consumption in response to higher government spending. The government spending multiplier could be negative, and if it is positive, it is likely less than unity. [2] Also, these models do not suggest any particular reason for nonlinearities in the responses of output and output

1 For example, the higher interest rate or negative wealth effect (see Parker 2011) induced by a rise in government spending could encourage higher labor supply that raises output, but higher interest rates could also reduce capital accumulation that lowers output in the medium to long run.

2 Gerchert and Will (2012) perform a meta-analysis of fiscal multiplier studies and find that equilibrium models tend to have the lowest multipliers, usually less than unity. That said, multipliers greater than unity can arise in equilibrium models with high degrees of complementarity between government spending and other activities.

***Corresponding author: Irina Panovska,** Department of Economics, Lehigh University, 621 Taylor Street, Rauch Business Center, Bethlehem, PA 18015, USA, e-mail: irp213@lehigh.edu
Steven M. Fazzari: Department of Economics, Washington University in St. Louis, St. Louis, MO 63130, USA
James Morley: School of Economics, University of New South Wales, Business School, Sydney, NSW 2052, Australia

components to government spending. On the other hand, Keynesian models predict that the economy will not always fully employ available resources, possibly for extended periods of time, because of insufficient demand. If output is below its potential level, an increase in government spending can directly employ idle resources and raise output. If government spending raises resource use through demand channels, consumption and investment should respond positively to spending shocks possibly leading to a government spending multiplier that exceeds one. Traditional Keynesian models imply that the spending multiplier could be large much of the time, whenever there is economic slack (and not just in recessions). But when the economy is near full employment and operating with little slack, higher government spending may well crowd out private output, leading to a smaller multiplier. Keynesian models therefore predict that the multiplier is nonlinear and state dependent, and that the appropriate threshold variable would be a measure of under-utilized resources or economic slack.

Many DSGE models with Calvo-type price rigidities emphasize somewhat different sources of state dependence: either the importance of monetary policy or the importance of the share of rule-of-thumb consumers in determining the size of the government multiplier. In DSGE models in which the size of the spending multiplier depends on the stance of monetary policy, multipliers are large only when the zero lower bound binds. The effects of a government spending shock die out as soon as the interest rate reverts to its natural level. In models that incorporate rule-of-thumb consumers, such as Galí, López-Salido, and Vallés (2007), the multiplier is large when the share of rule-of-thumb consumers is high.

To investigate the possibility of state-dependent effects of fiscal policy, we estimate a nonlinear structural vector autoregressive model that allows parameters to switch when a specified variable crosses an estimated threshold. As candidate threshold variables, we consider several alternative measures of economic slack, as well as the debt-to-GDP ratio and a measure of the real interest rate. Various statistical and economic criteria identify capacity utilization (adjusted for a structural break) as the best threshold variable, but the results are robust to the other measures of slack.

Our empirical results provide strong evidence in favor of state-dependent nonlinearity; specifically, government spending shocks have larger effects on output when they occur with relatively low resource utilization than when they occur at times of high resource use. Furthermore, threshold estimates for capacity utilization place half or more of the historical observations from 1967 to 2012 in the low-utilization regime. This evidence implies that the state of the US economy is often one in which fiscal shocks have large positive and persistent effects on output and its components most of the time, not just in deep recessions or when interest rates are pinned against the zero bound. We also employ simulation-based impulse-response functions to isolate the different effects of fiscal policy under particular economic conditions, and we introduce a formal impulse response comparison method that allows us to directly compare the impulse responses across different states of the economy. We find that the responses of output and output components depend crucially on the state of the economy when a policy shock occurs. The response of output to a positive shock in government spending is much larger during periods of slack than during periods when the economy is close to the capacity constraint.

The rest of the paper is organized as follows. Section 2 provides background and motivation for our analysis. Section 3 introduces the baseline empirical model and the estimation method. Section 4 presents the empirical results and extends the baseline model to models that include consumption, investment, and other variables of interest. Results from an extended model that includes both government spending and real interest rates are provided in Section 5. Section 6 concludes.

2 Background and motivation

The multiplier estimates obtained with different estimation techniques and calibrated dynamic stochastic general equilibrium models vary widely, from −4 to 4 (see Ramey 2011a; Parker 2011; Leeper, Traum, and Walker 2011; Van Brusselen 2009; and Leigh et al. 2010, for extensive surveys of the literature). Previous studies that examine state-dependence of the spending multiplier almost exclusively focus on the size of the multiplier in deep recessions compared with expansions. Based on the seminal work of Auerbach and Gorodnichenko (2012a), most studies that allow for nonlinear responses use a version of a threshold vector autoregression. They start with the baseline linear model introduced by Blanchard and Perotti (2002), and extend it to allow the economy to evolve between different regimes. Threshold models provide a natural econometric framework for exploring the state dependence of fiscal multipliers. For example, if government spending shocks affect output through demand channels, we expect such effects to be larger when the economy has resource slack than when it is operating at or near full capacity. If there is nonlinearity in the response of output and output components that is driven by the responses of monetary policy, the interest rate is the threshold variable that triggers the different regimes, and the multiplier will be large only when the interest rate is sufficiently low.

Auerbach and Gorodnichenko (2012a) estimate a smooth transition threshold vector autoregressive (VAR) model for government spending, taxes, and output, in which they impose the restrictions that government spending has different effects during recessions and expansions, and they calibrate the smoothness parameter based on US data so that the economy spends about 20% of the time in recessions. They estimate that the effects of government spending are large and positive when the economy is in a recession and smaller when the economy is not in a recession.[3] They control for the state of the business cycle by using a moving average of output growth as the threshold variable, and they impose that the threshold around which the behavior changes is equal to the mean of output growth.[4] Mittnik and Semmler (2012) estimate a bivariate threshold model for output and employment where the switching variable is lagged output growth and the threshold is predetermined and equal to the mean of output growth. In their model, the responses of employment to output shocks are much larger in the low regime than in the high regime. Candelon and Lieb (2013) extend the model used by Auerbach and Gorodnichenko by imposing long run equilibrium conditions and using sign restrictions to identify fiscal shocks. They also find strong evidence of state-dependence, but the estimated output multipliers are smaller. Baum and Koester (2011) find strong evidence in favor of state-dependent effects of fiscal policy in Germany, but the multipliers are smaller than the estimated multipliers for the US. Shoag (2013) obtains much higher multipliers for state-level government spending during periods of slack in the labor market than during normal periods. Owyang, Ramey, and Zubairy (2013), and Ramey and Zubairy (2013) combine the approach used by Auerbach and Gorodnichenko (2012b) and Ramey (2011a) by imposing a threshold and augmenting the model with a narrative measure of military spending, but they do not estimate the threshold from the data. Instead, they assume that the threshold is equal to a fixed natural rate of unemployment. They find state-dependent effects for Canada, but no significant evidence of state-dependence for the US when the threshold is restricted to be equal to the natural rate of unemployment.

Our analysis differs from many other nonlinear studies of fiscal policy and aggregate demand in some important ways. Importantly, we consider a wide variety of possible threshold variables rather than choosing one variable “a priori”. Statistical criteria select capacity utilization as the switching variable that best describes the nonlinearities in the data for the sample that we consider (with the output gap a close second). Capacity utilization appears to encapsulate much of the information about economic slack from other macroeconomic data. However, capacity utilization is survey-based, so it is not subject to significant revisions, unlike, for example, employment growth or the CBO output gap, for which there are often large revisions

3 Bachmann and Sims (2012) estimate a very similar nonlinear VAR model to Auerbach and Gorodnichenko (2012a) and find the same result that government spending shocks have larger effects during recessions than during expansions. Their additional insight is that these larger effects during recessions appear to operate largely through consumer confidence. In particular, if the response of consumer confidence to government spending shocks is shut down in the calculation of impulse-response functions, the effects are much smaller and similar to the estimated effects in expansions (with or without the consumer confidence channel).

4 In a follow-up to their original study, Auerbach and Gorodnichenko (2012b) find that their results for the US data are largely robust across a large number of OECD countries given the same restrictions to identify recessions, but considering a panel structure and direct multi-period single-equation projections to calculate impulse-response functions. Their consideration of a panel structure and single-equation projections rather than an VAR model is motivated in part by a lower frequency of available data for many countries, making statistical identification of a nonlinear VAR model challenging.

around the NBER turning points (see, for example, Billi 2011, on the CBO output gap, and Orphanides and van Norden 2003, on other measures of the output gap). Also, many of the commonly used measures of slack, including the CBO output gap, require estimating the natural level of output or the state of the economy, which is, of course, subject to estimation error. Morley and Piger (2012) compare many different measures of the business cycle and slack obtained from a wide range of linear and nonlinear time series models. They find that, as an observable time series, capacity utilization is particularly highly correlated with a composite measure of slack that best matched the NBER business cycle chronology and was estimated by averaging across different time series models in order to reduce estimation error. Specifically, capacity utilization serves as a particularly convenient observable proxy for their more complicated forecast-based estimate of slack (see also Morley 2014, who finds a strong relationship between the forecast-based estimate of slack and capacity utilization for a number of economies in Asia and the Pacific).

In addition, we estimate the threshold that determines state-dependent effects from the data, and formally compare the linear model to the nonlinear alternative of state-dependent responses. Most previous studies impose the threshold and the variable that determines the prevailing regime *a priori* and do not formally test for nonlinearity. Three exceptions are Candelon and Lieb (2013), Baum and Koester (2011), and Baum et al. (2013). Candelon and Lieb extend Auerbach and Gorodnichenko's model by allowing for long-run equilibria and selecting a threshold variable based on minimizing MSE. Their preferred variable is Stock and Watson's coincident index, and the estimated threshold splits the sample into downturns and upswings. Baum and Koester (2011) and Baum et al. (2013) estimate the threshold when the switching variable is the output gap. Both studies consider a classical hypothesis test and find supportive evidence for nonlinearity with larger multipliers in recessions than in expansions.

Finally, because we estimate the threshold, our model allows the data to sort observations into possibly different multiplier regimes, and we explore different potential sources of nonlinearities. In contrast to Auerbach and Gorodnichenko (2012a), we find evidence that the US economy spends the majority of its time in the low-utilization/high-multiplier state.[5]

3 Empirical methods

3.1 Model

A basic VAR model is linear, and cannot capture nonlinear dynamics such as regime switching and asymmetric responses to shocks. For our analysis, we consider a nonlinear version of a VAR model that extends the threshold autoregressive model of Tong (1978, 1983) to a multivariate setting. This model splits a time series process endogenously into different regimes. Let Y_t denote a vector containing the endogenous variables in the VAR. Within each regime the stochastic process for Y_t is linear. Let superscripts 1 and 2 denote the regimes. Then, within each regime the dynamics of Y_t follow:

$$Y_t = \Phi_0^1 + \Phi_1^1(L)Y_{t-1} + \varepsilon_t \tag{1}$$

5 Another difference from previous studies is that we consider a threshold VAR model with a discrete change in regime instead of the smooth transition specification considered by Auerbach and Gorodnichenko (2012a). Although the smooth transition specification is potentially more general, estimating the smoothness parameter for such a model can be challenging, as evidenced by the fact that Auerbach and Gorodnichenko (2012a) fix this parameter (as well as the threshold) in their estimation. The difficulty is that the likelihood function for a smooth transition model is flat when the true smoothness parameter is large in the sense of implying a relatively discrete threshold, making maximum likelihood estimation and even Bayesian estimation unreliable. We circumvent this econometric problem by considering a discrete threshold only, which still allows us to focus on the primary question of whether there are state-dependent effects of fiscal policy. The consideration of whether capacity constraints bind or not also provides a possible economic justification for the discrete threshold specification.

for values of t when output is in regime 1 and

$$Y_t = \tilde{\Phi}_0^2 + \tilde{\Phi}_1^2(L)Y_{t-1} + \varepsilon_t \tag{2}$$

for t values in regime 2.

Let q_{t-d} denote the threshold variable that determines the prevailing regime. If q_{t-d} is less than or equal to the threshold c, the variables in Y_t follow the dynamics of regime 1 and Y_t is in regime 2 otherwise. The integer d is the delay lag for the threshold switch. If the threshold variable q_{t-d} crosses c at time $t-d$, the dynamics of output actually change at time t. Define an indicator function $I[\cdot]$ that equals 1 when the q_{t-d} exceeds the threshold c and equals 0 otherwise. The full model can then be written in a single equation as:

$$Y_t = \Phi_0^1 + \Phi_1^1(L)Y_{t-1} + (\Phi_0^2 + \Phi_1^2(L)Y_{t-1})I[q_{t-d} > c] + \varepsilon_t. \tag{3}$$

Where $\Phi_0^2 = \tilde{\Phi}_0^2 - \Phi_0^1$ and $\Phi_1^2 = \tilde{\Phi}_1^2 - \Phi_1^1$. This is our empirical model. The endogenous, data-driven switches between regimes make the full model nonlinear. The constants in each regime, Φ_0^1 and Φ_0^2, the lag polynomial matrices Φ_1^1 and Φ_1^2, the threshold (c), and the delay lag (d) are estimated from the data. In the baseline version of the model, the vector Y_t includes the first difference of the logarithm of real government spending, the first difference of the logarithm of net taxes, the first difference of the logarithm of real GDP, and a measure of economic slack, as discussed in more detail below. We also consider alternative versions of the model that incorporate the private-sector components of real GDP (i.e., consumption, investment, exports, and imports) or other variables such as the unemployment rate, employment, a real interest rate, and inflation, again discussed in more detail below.

The disturbances ε_t are assumed to be independent and Gaussian with mean zero. Rather than assuming that the disturbances are strictly i.i.d., we set the covariance matrix of ε_t equal to Ω until 1984Q1 and equal to $\lambda\Omega$ afterwards to capture the Great Moderation. Because the focus of this paper is not on determining the exact break date in volatility and because there is near consensus in the literature about the general timing of the volatility break (see, for example, Kim and Nelson 1999, or McConnell and Perez-Quiros 2000), we set the break date exogenously. By using a scale factor λ and a constant variance-covariance matrix Ω, we allow the size of the shocks to change with the Great Moderation, but the correlations between the disturbances do not change over time, and, implicitly, the impact responses are consistent over states. This assumption means that differences in the impulse responses will capture differences in the transmission mechanism, not different identification of structural shocks. Although the assumption of constant correlations may appear restrictive at first sight, it is consistent with other threshold VAR studies (see, for example, Auerbach and Gorodnichenko 2012a), who find no significant difference in the impact responses in the initial period.[6] For the threshold variable, q, we consider capacity utilization, other measures of economic slack, and a selection of other macroeconomic variables, as discussed in more detail below.

3.2 Data

In addition to capacity utilization, we also consider the output gap estimated by the CBO, the unemployment rate, output growth, and employment growth to measure economic slack. The traditional Keynesian theory summarized above implies that the threshold variable should measure the level of economic activity and intensity of resource use. For this purpose, the output gap, the level of capacity utilization or the unemployment rate would seem to be good choices. However, we also consider first differences of these variables and output and employment growth to check the robustness of the results and to explore whether threshold

6 Allowing for the variance-covariance matrix to vary over states to allow for different correlations leads to very similar responses, but less precise inference. There is no statistically significant evidence that the correlations between the disturbances vary over time. Results are available from the authors upon request.

effects might relate to growth (as in Auerbach and Gorodnichenko 2012a, and Mittnik and Semmler 2012) rather than to levels.

Government spending and net taxes are defined as in Blanchard and Perotti (2002). The full sample period is for the baseline estimation is 1967Q1–2012Q4, since the capacity utilization series is available starting from 1967. All output components are measured in real terms and are seasonally adjusted by the source. The series for output, its components, including government spending, and tax revenues were obtained from NIPA-BEA, and the capacity utilization series was obtained from the Federal Reserve Statistical Releases website. We also consider data for US federal government debt, the Federal Funds Rate, inflation based on the CPI (seasonally adjusted), and non-farm payroll employment, which were all obtained from the Federal Reserve Bank of St. Louis FRED website. The monthly series for capacity utilization, the unemployment rate, the Federal Funds Rate, CPI, and employment are all converted to a quarterly frequency by using simple arithmetic means.

All of the models discussed in Sections 3–4 include four variables in the VAR specification. The baseline model includes the growth rate of government spending, the growth rate of tax revenues, the growth rate of GDP, and capacity utilization. The models used to explore the responses of output components, prices, or labor market variables include government spending, taxes, the variable of interest, and capacity utilization. The extended model in Section 5 augments the baseline model by including interest rates as an additional variable: i.e., Y_t includes the growth rate of government spending, the growth rate of tax revenues, the growth rate of GDP, capacity utilization, and the real interest rate. The identification and the specification for the extended model are discussed in more detail in Section 5. We use growth rates rather than log-levels in the VAR model because the logarithms of real GDP and output components appear to have stochastic trends according to standard unit root and stationarity tests, but there is no support for common trends amongst the variables in any version of the VAR model under consideration based on Johansen cointegration tests.[7]

3.3 Specification issues

The lag length for the VAR model is chosen based on AIC (for the baseline linear VAR model, estimated using maximum likelihood as a starting point), and the lag length is imposed on all nonlinear specifications. While considering a model in which the lag length differs across regimes would allow for potentially richer dynamics, doing so would entail comparing a very large number of models. Our approach is in line with the approach used by related non-linear studies, for example, Auerbach and Gorodnichenko (2012), Owyang, Ramey, and Zubairy (2013), and Ramey and Zubairy (2013).[8]

To solve for the structural VAR (SVAR) model given the reduced-form VAR parameters, we impose short-run zero restrictions with government spending ordered first and taxes ordered second in all models, i.e., government spending is assumed to respond to economic conditions only with a lag, but economic conditions are allowed to respond immediately to government spending. Implicitly, our approach to solving the SVAR model assumes that the impact matrix identifying structural shocks remains the same across regimes and throughout the entire sample period, with only the size of structural shocks allowed to undergo a structural break in 1984. This approach avoids any ambiguity about whether the dynamic effects of government

7 This approach follows the suggestion in Hamilton (1994, chapter 18) when unit roots are assumed to present and in the absence of cointegration. The results obtained when estimating the model in levels or imposing cointegration between spending and taxes are qualitatively similar, albeit less precise, in comparison to those for our baseline model and are available from the authors upon request.

8 One exception is Mittnik and Semmler (2012), who estimate a bivariate model for output and employment, and allow the number of lags to vary across the regimes. It is certainly possible to extend the nonlinear VAR model to accommodate a different number of lags across regimes, to allow for more more than two discrete regimes, or even an infinite number of regimes (by using a smooth transition model). However, it would make computation very burdensome and possibly imprecise, both because of the larger number of parameters that would need to be estimated and because of the identification issues for smooth transition models discussed in the previous section.

spending shocks appear state-dependent because of a change in their identification rather than their propagation, as mentioned above.

Economic theory implies several possible choices for the threshold variable. As also discussed above, traditional Keynesian theory suggest that the dynamics may depend on the state of the economy, while some DSGE models imply that the effects of government spending depend on the interest rate. A recent literature suggests that the dynamics may also depend on the level of government debt (see, for example, Reinhart and Rogoff, 2009, and Eggertsson and Krugman, 2012, for two very different views on the impact of debt on the efficacy of fiscal policy). Because we do not want to impose the threshold variable *a priori*, we consider a large set of possible threshold variables and select the preferred threshold variable using Bayesian model comparison. The threshold variables that we consider are

1. lagged output: output growth Δy_{t-d} for d=1, 2, 3, 4), long differences in the natural log of output ($y_t - y_{t-4}$), moving averages of differences in the natural log of output[9]
2. CBO output gap: gap_{t-d} for d=0, 1, 2, 3, 4
3. lagged capacity utilization: level, level adjusted for long-run change in mean, first differences. The series considered are cap_{t-d} for d=1, 2, 3, 4, Δcap_{t-i}, $\widehat{cap_{t-d}}$ for d=1, 2, 3, 4 where $\widehat{cap_t} = cap_t - \mu_t$ with $\mu_t = \mu_1$ before 1974 and $\mu_t = \mu_2$ after 1974, Δcap_{t-i} for i=1, 2, 3, 4 and $\Delta\widehat{cap_{t-d}}$ for d=0, 1, 2, 3, 4.
4. unemployment rate: level, differences, mean-adjusted level for d=0, 1, 2, 3, 4
5. debt-to-GDP ratio: total Federal debt and total Federal debt held by the public as a percent of GDP for d=0, 1, 2, 3, 4
6. real interest rate: level and change in the ex ante real interest rate based on the Federal Funds Rate and CPI inflation under the assumption of static expectations for d=0, 1, 2, 3, 4

Both capacity utilization and the unemployment rate appear to have changes in their long-run mean levels, which would make those series unsuitable for use in a stationary VAR model. Standard tests for a structural break at an unknown break date reject the null of no break in mean for both capacity utilization and the unemployment rate. Meanwhile, there is some debate about whether the unemployment rate has a unit root or whether there were just exogenous structural breaks in its mean (see, for example, Papell, Murray, and Ghiblawi 2000). For both series, therefore, we consider the level, first differences, and the mean-adjusted levels as possible threshold variables.

To determine the delay lag, we estimated a TVAR where d is fixed for different values of d, (d=0, 1, 2, 3, 4) and picked the model with the largest marginal likelihood. This approach is similar in spirit to the standard maximum likelihood estimation of TVAR and threshold models, in which mazimization is performed over c and d. Note that in the cases when the switching variable enters the TVAR (in our case, when q is a function of output or capacity), Y_t depends on c, so in the case when d=0, the error and Y_t are correlated. For those cases, we only estimate the model for $d>0$. In the cases where the switching variable does not enter the VAR directly (beyond the index function), we estimate the model for d=0, 1, 2, 3, 4.[10]

Table 1 summarizes the results of the test for structural breaks in mean for capacity utilization and the unemployment rate. A structural break test for capacity utilization identifies a highly significant break (F statistic of 41.7) in the level of capacity utilization in 1974Q1, which coincides with the well-known productivity slowdown. Structural break tests also identify three breaks in mean for the unemployment rate. The left

9 Because we already estimate a large number of parameters, the weights for the moving averages were fixed exogenously. We considered an arithmetic mean of the past 4 differences, and $q_{l,t-d} = \frac{1}{l-d+1}\sum_{j=d}^{l} threshold\_var_{t-j}$ for l=1, d=4.

10 It is important to note that in the cases where d=0, we would have to assume that the switching variable is exogenous in order to justify the model as specified. Because output is highly correlated with the switching variables, even when they do not directly enter the VAR, and there are many economic reasons to believe that there is causality between output and the switching variables considered, so there still may be an endogeneity problem when d=0. However, the maximum marginal likelihood still chooses a lag of at least one quarter for every threshold variable we considered except for the mean-adjusted unemployment rate, and the marginal likelihood for our preferred model is substantially higher then for the model where the switching variable is the unemployment rate with lag 0, therefore an endogeneity problem when d=0 does not drive our main results.

Table 1 Structural breaks in capacity utilization and the unemployment rate.

	Break date	*F*–statistic	*p*-Value
Capacity utilization	1974Q1	41.66	<0.01
Unemployment rate	1974Q3	32.34	<0.01
	1981Q4	16.43	<0.01
	1994Q4	6.21	0.03

Break dates were obtained using sequential Quandt-Andrews tests. The estimated break dates coincide with the break dates obtained using Bai-Perron's sequential procedure under the assumption that the mean is the only parameter that has structural breaks.

panel of Figure 1 shows the mean-adjusted capacity utilization series and the estimated threshold, and the right panel shows the original capacity utilization data. The mean adjusted capacity utilization series is the threshold variable preferred by the data.

3.4 Estimation and inference

Because the threshold VAR model is highly parametrized, we make inferences about the threshold and the coefficients using Bayesian methods. We use a multi-block Metropolis-Hastings (MH) algorithm, described in detail in the appendix, to sample from marginal posterior distributions for parameters and calculate marginal likelihoods for models. The advantages of using a Bayesian approach in this setting are two-fold.

First, Bayesian estimation allows us to capture the uncertainty about the parameter values when constructing the impulse-reponse functions. When using a frequentist approach to estimate impulse responses for multivariate threshold models, the simulated IRF procedure produces a consistent estimate that is conditional on the initial state, assuming that the parameters c, Φ, and Ω are fixed (the true parameter value is considered to be equal to the maximum likelihood estimate). The impulse responses for the endogenously evolving system have non-standard asymptotic distributions that are usually not Gaussian and depend on the history and the size of the shock. Therefore, studies that use frequentist TVAR model typically report either just the mean response for the evolving states, or the IRF for the piece-wise linear model that assumes that the economy stays in one state forever. Since the Bayesian approach produces the entire posterior distribution for c, Φ, and Ω conditional on the data, we directly account for dispersions in the posterior distribution of the parameters by simulating the impulse responses for each iteration of the MH sampler.

Second, despite the presence of nuisance parameters in the nonlinear models, comparing the linear to the nonlinear model and examining the presence of nonlinear effects is relatively straightforward in the Bayesian framework by comparing marginal likelihoods or the impulse response functions.

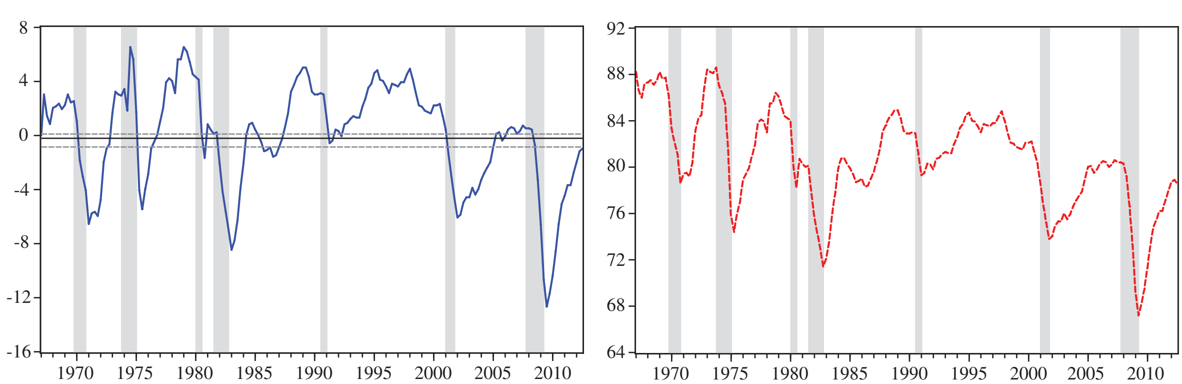

Figure 1 Left Panel: Mean-adjusted capacity utilization and estimated threshold posterior mean with 90% credibility interval. Right Panel: Unadjusted capacity utilization data. The shaded areas are NBER peak-to-trough dates.

To provide an accurate approximation of the target posterior distribution of the parameters, we follow the standard approach in the applied literature and we use a tailored multivariate Student's-t distribution as the proposal distribution. Our prior for the autoregressive parameters Φ is a normal distribution, truncated to ensure stationarity. The prior distribution for the variance-covariance matrix of shocks Ω is an inverse-Wishart distribution, the prior for the scaling parameter for the variance-covariance matrix λ is gamma, and the prior distribution for the threshold c is uniform over $[q_l, q_h]$ where q_l and q_h are the highest and the lowest observed values of the the threshold variable.[11] The full technical details of the posterior sampler and the priors are relegated to the appendix.

There are two ways to compare whether the effects of government spending differ across regimes in this framework. First, we can evaluate if the model exhibits state dependence by using Bayesian model comparison. Second, we can explore state dependence, and size or sign asymmetries in the effects of shocks by directly comparing impulse responses. A crucial empirical question is whether the effects of government spending really do differ across regimes defined by economic slack.[12] To compare the linear model to the nonlinear alternative, we estimate the threshold VAR model using the MH algorithm and then we compare its marginal likelihood to that for a restricted linear version of the VAR model in (1) for which $\Phi_0^2=0$ and $\Phi_1^2=0$. Marginal likelihoods are calculated using Chib and Jeliazkov's (2001) algorithm and we compare models based on Bayes factors, which are the ratio of marginal likelihoods and are equal to posterior odds ratios under even prior odds (i.e., equal prior probabilities on all models under consideration).

Rejecting linearity using Bayesian model comparison implies that at least one of the impulse responses to at least one of the structural shocks is necessarily different across regimes, but the degree of this asymmetry can only be evaluated by looking at the impulse response functions themselves. This approach is appropriate for examining the question at hand for two reasons. First, the impulse responses give us the magnitude of the response of output and its components to any kind of government spending shock for any history of interest, so they can be used both to define the multiplier in the usual sense and to examine the response to cuts and to increases of different sizes. Second, when it comes to designing policies, the response of output is much more important than the coefficient estimates, and policy makers are usually more concerned with the response of output or another variable of interest conditional on current economic conditions, rather than with the response averaged over all historical conditions. The impulse-response comparison approach allows us to compare both the average responses and precisely estimated responses conditional on particular initial conditions.

For the nonlinear model, we construct two sets of impulse responses. In the first case, the economy is assumed to remain in a given state forever. Because the model is linear within a state, the IRFs can be obtained using the estimated VAR coefficients for the given regime. In the second case, the state of the economy is allowed to evolve because the threshold variable itself responds to government spending shocks. When we allow the system to evolve and switch between regimes, the IRFs depend on the initial state and possibly on the size and the sign of the shock. Following Koop, Pesaran, and Potter (1996), we consider simulation-based IRFs in order to measure the responses when the threshold variable is allowed to respond endogenously. The

11 Using a truncated univariate Student's-t prior for c with mean equal to the maximum likelihood estimate and 5 degrees of freedom (relatively flat over the observed values) leads to very similar posterior estimates.

12 In a frequentist setting, to test for the presence of nonlinear effects, we would want to consider the null hypothesis $H_0:\tilde{\Phi}_0^2=\Phi_0^1, \tilde{\Phi}_1^2=\Phi_1^1$ that the coefficients are equal against the alternative that at least one of the elements of the matrices Φ_0^2, Φ_1^2 is not zero. This testing problem is tainted by the fact that the threshold c is not identified under the null. If the errors are i.i.d., a test with near-optimal power against alternatives distant from the null hypothesis is the supLR test, but the asymptotic distribution of the test statistic is nonstandard and has to be approximated using Hansen's (1996, 1997) bootstrap procedure. Because the model is very parameter-rich, bootstrapping the asymptotic distribution is computationally prohibitive. Also, it should be noted that the 1984Q1 structural break in the variance-covariance matrix of the disturbances makes it unclear how well Hansen's procedure would perform in this setting. The Bayesian approach circumvents such problems by providing a direct method for comparing models based on the posterior odds ratios. It should be noted, however, that a bootstrap version of the supLR test for a simpler version of the model with only government spending, net taxes, and real GDP as endogenous variables and still using capacity utilization as the threshold variable is significant at the 5% level (under the assumption that the structural break does not distort the test). The results are available from the authors upon request.

impulse responses are defined as the change in the conditional expectation of Y_{t+k} as a result of a shock at time t:

$$IRF[shock_t, \Psi_{t-1}] = E[Y_{t+k} | shock_t, \Psi_{t-1}] - E[Y_{t+k} | \Psi_{t-1}] \tag{4}$$

where Ψ_{t-1} is the information set at time t–1. Calculating the IRFs requires specifying the nature of the shock and the initial conditions Ψ_{t-1}, and then the conditional expectations $E[Y_{t+k}|shock_t, \Psi_{t-1}]$ and $E[Y_{t+k}|\Psi_{t-1}]$ are computed by simulating the model. We consider an orthogonal exogenous shock identified from the SVAR model rather than a forecast error from the reduced-form VAR, as considered in Koop, Pesaran and Potter (1996). Because threshold models imply that the predicted responses from the model to a shock depend on a particular history, we can simulate the responses for the evolving model for a particular history of interest, or averaging over all histories when the threshold variable is above or below the estimated threshold.

In practice, the simulation-based structural IRFs are computed as follows (a detailed version of the algorithm is presented in the appendix): First, shocks for periods 0–20 are simulated using the estimated variance-covariance matrix for the threshold SVAR model and, for given initial values of the variables, fed through the estimated model to produce a simulated data series. The result is a forecast of the variables conditional on initial values and a particular sequence of shocks. Next, the same procedure is repeated with the same initial values and shocks, except that the shock to government spending in period 0 is fixed at 1% of GDP (for that particular starting value of GDP). The shocks are fed though the model and a forecast is produced just as above. The difference between this forecast and the baseline forecast is the IRF for a particular sequence of shocks and initial values. This computation is repeated for 500 draws of the shocks and averaged to produce IRFs conditional only on a particular history. These IRFs are then averaged over a particular subset of initial values.[13]

Because threshold models imply that the predicted responses from the model to a shock depend on a particular history, we first simulate the responses for the evolving model, averaging over all histories when the threshold variable is above the estimated threshold and averaging over states when it is below. Then we compare those results to those obtained when we simulate the IRFs for the recent histories between 1984 and 2011 when the threshold variable is above the threshold and when it is below, including the "New Economy" rapid expansion in the late 1990s and the "Great Recession." To capture the uncertainty about the parameter values, the credibility intervals for the impulse-response functions are obtained by simulating the IRFs for all iterations of the MH algorithm. As discussed in detail in the Appendix, in addition to producing a measure of the multiplier for any kind of spending shock that directly accounts for parameter uncertainty, this approach allows us to compare the impulse responses across states or the responses to different kinds of spending shocks by looking at the posterior of the distribution of the difference $\Delta IRF[shock_{1t_1}, shock_{2t_2}, \Psi_{t_1-1}, \Psi_{t_2-1}] = IRF[shock_{t_1}, \Psi_{t_1-1}] - IRF[shock_{t_2}, \Psi_{t_2-1}]$.

We can evaluate if the difference between the responses is significant simply by checking if zero is within a given quantile of the posterior. This approach is similar in spirit to the approach used by Kilian and Vigfusson (2011), who test for size and sign asymmetry in a frequentist setting by looking at the distribution of the impulse responses, but it is slightly more general, because it allows both for state-dependence and for sign and size asymmetry within a state. The simulation approach for constructing the IRFs used here is different

13 It is necessary to calculate the evolving responses by simulation. For a fixed parameter draw, if the model starts in regime 1, the response in the initial period will be the response from the linear data generating process for regime 1 (DGP1) in period 0. However, the response to the shock may move the model to regime 2 (DGP2). If the shock does move the model to the second regime, the response in period 2 will be governed by DGP2. If the economy initially started close to the threshold, the shock may move the economy from regime 1 to 2 right away. If the economy starts far from the threshold, it may take a long time to move above the threshold, and the response will look more like the response for DGP1. Because the response depends on the initial condition, it has to be calculated separately for each individual initial condition. These responses are conditional on the history and on the parameter draw. When we are interested in the average response for a subset of initial conditions (for example, all periods when the economy was below the threshold), we average the responses for all of those histories. Because the simulated responses depend on the parameters, they are different for different draws of the MH sampler. To obtain the entire posterior distribution for the IRFs, we calculate the average IRF for the histories of interest for each parameter draw.

from Jorda's (2005) projection method used by Auerbach and Gorodnichenko (2012b) and Owyang, Ramey, and Zubairy (2013). While the simulation approach is more time consuming than Jorda's projection method, it is more general when evaluating responses in endogenously evolving states, since it directly allows the responses to depend on the intial state.

4 Empirical results

As discussed in Section 3.4, our formal model comparisons are based on marginal likelihoods and implied Bayes factors. Table 2 reports marginal likelihood values for the baseline model with different threshold variables, including the restricted case of no threshold effect.

The implied Bayes factors strongly favor nonlinearity when threshold variables relate to economic activity. Direct measures of economic slack perform the best.[14] Based on the results in Table 2, the preferred threshold variable for the baseline model is the first lag of capacity utilization (adjusted for a one-time structural break in the mean, as discussed in Section 3.3). The marginal likelihood with the output gap as the threshold variable is very close to that with capacity utilization. [15] Because adjusted capacity utilization is the preferred choice based on statistical criteria, the results that follow are based on estimation with this threshold variable, but similar results would be obtained with any of the measures of economic slack.

This support for a threshold based on economic slack has important implications. The relevance of slack is consistent with has been called an "Old Keynesian" interpretation that the real effects of government spending on output work through a demand channel. Moreover, it is level of slack, rather than some measure of the change or growth of output that the data choose as the best variable to define the regimes.

The hypothesis that high levels of government debt reduce the real effects of fiscal policy is inconsistent with our results for the US. There is no support for nonlinearity with the debt-to-GDP ratio as the threshold variable. The estimated threshold in this case is near the boundary of the parameter space considered, so the lack of support for nonlinearity might reflect the relatively low levels of the debt-to-GDP ratio in the US economy since 1967, at least compared to the levels observed in other countries that have suffered debt crises.

For the real interest rate, interest rate, there is no evidence of nonlinearity. If we impose nonlinear model nonetheless, the estimated threshold is about 2%, which is close to typical estimates of the long-run "neutral"

Table 2 Marginal likelihoods and estimated thresholds for baseline model with different threshold variables.

Threshold variable	Lag	Marginal likelihood	Threshold estimate
None	–	−997.18	–
Output growth	2	−720.69	1.33 (0.12)
Output gap	1	−682.52	−0.59 (0.41)
Capacity utilization	1	−800.52	81.10 (1.42)
Capacity utilization (adjusted)	1	−673.69	−0.21 (0.37)
Unemployment rate	1	−703.25	4.83 (0.33)
Unemployment rate (adjusted)	0	−732.26	0.92(0.29)
Debt-to-GDP ratio	2	−1020.23	47.2 (1.15)
Real interest rate	2	−1004.42	2.10 (1.35)

The threshold estimate is the posterior mean (with standard deviation in parentheses). Preferred variables for each category listed in Section 3.3 are as stated, with the preferred debt measure being total federal debt outstanding.

14 A higher marginal likelihood for model 1 versus model 2 indicates that the data support model 1 over model 2, given the prior distribution of the parameters and the prior probability that we put on each model. With equal prior probabilities on the models the ratio of the marginal likelihoods is equal to the posterior odds ratio; that is, it gives the relative probability of one model versus another given the data and the priors about the parameters. For more technical details, we refer the reader to the Bayesian Estimation Appendix.

15 The correlation between adjusted utilization and the output gap is 0.63.

rate. However, this estimate is quite imprecise, consistent with the lack of support for a threshold effect relating to the interest rate. Our findings therefore do not support the idea that the response of the economy to government spending depends in an important way on the interest rate. Again, this result conforms better with "Old Keynesian" models than with the "New Keynesian" view that fiscal policy has much larger effects when interest rates are pinned at or near the zero lower bound. It is possible, however, that the outcome would be different if the sample included more observations when nominal interest rates were near zero.

The estimated threshold for the baseline model is slightly below the mean of the adjusted capacity utilization series.[16] The mean-adjusted capacity utilization series and its estimated threshold are plotted in Figure 1. Our estimated threshold estimate has quite different implications from the split into "recession multipliers" and "expansion multipliers" found in other studies. Notably, more than 60% of the historical observations for mean-adjusted capacity utilization fall below the mode of the posterior distribution for the threshold parameter, while close to 50% of observations fall below the posterior mean. This result is important because it implies that, for a majority of the time since the middle 1960s, the US economy has operated in a regime in which, as shown below, government spending shocks have relatively large effects on output. Since 2000, almost all observations have been in this regime. This result also distinguishes our approach from Auerbach and Gorodnichenko (2012a), as their approach imposes that only 20% of the observations fall in a recessionary regime. As discussed in Section 2, capacity utilization appears to be a representative measure of economic slack. The turning points of utilization track the NBER turning points quite closely, and it takes several quarters for capacity utilization to return to its pre-recession level following the trough in GDP.

Although the marginal likelihood results in Table 2 strongly favor nonlinearity, it is important to address Sims' (2001) concern that evidence for time-varying parameters in VAR models may be the spurious result of failing to fully account for heteroskedasticity. Therefore, we consider diagnostic tests for our preferred baseline model with mean-adjusted capacity utilization. The model allows for some heteroskedasticity given that it incorporates a one-time structural break in the scale of the variance-covariance matrix for the VAR residuals corresponding to the Great Moderation in 1984Q1. For this model, the standardized residuals based on the parameter values at the posterior mean pass the Jarque-Bera test for normality of the individual residual series and the Doornik-Hansen test statistic for multivariate normality is 10.54 (p-Value 0.23). Also, there is no evidence of serial correlation in the standardized residuals based on Ljung-Box Q-tests and the ARCH-LM test does not reject the null of a constant variances for the individual residual series. Thus, the evidence for nonlinearity does not appear to be an artifact of unmodeled heteroskedasticity. Instead, it appears that we have successfully captured any heteroskedasticity by allowing for a structural break in the scale of the variance-covariance matrix for the VAR residuals.

When we estimate the effects of government spending on output components and other variables, we substitute the outcome variable of interest (i.e., consumption, investment, exports, imports, the unemployment rate, employment, and inflation) for output in the baseline VAR model, using the first lag of mean-adjusted capacity utilization as the threshold variable.[17] As with the baseline model, we find strong evidence of nonlinearity for these models. Table 3 reports the marginal likelihood values for linear and nonlinear specifications of these alternative VAR models. In every case, the nonlinear specification is preferred. In particular, the implied Bayes factors always favor the nonlinear specification, with posterior odds only tipping in favor of linear specifications given extremely high prior odds of more than 10–1 for the linear specifications. Table 3 also reports the estimated thresholds in mean-adjusted capacity utilization for these alternative VAR models and shows that they are quite consistent across the different models, as is evident by looking at the various threshold estimates in the third column of Table 3 in the context of the variation of adjusted capacity utilization plotted in Figure 2.

16 The sample mean for unadjusted capacity utilization is 85.2 for 1967Q1–1974Q1 and 79.9 for 1974Q2–2012Q4.

17 In preliminary analysis, we also considered these effects by adding each series as a fifth variable to the baseline model. The point estimates for the threshold and the median impulse responses were very similar for both specifications, but the 95% (and even the 75%) credibility intervals were very wide in the specification with five variables because there are too few observations per regime to precisely estimate a threshold VAR model with so many variables without imposing very tight priors. Thus, the results presented in the rest of this paper are based on four-variable versions of the VAR model.

Table 3 Marginal likelihoods for linear and nonlinear specifications with different outcome variables and estimated thresholds.

Outcome variable	Linear model ML	Nonlinear model ML	Threshold estimate
Output	−997.18	−673.69	−0.21 (0.37)
Consumption	−851.95	−576.91	−0.54 (0.37)
Investment	−2011.98	−1473.50	−1.39 (1.32)
Exports	−6414.28	−4060.67	−1.00 (0.12)
Imports	−7011.89	−4311.90	−0.18 (0.37)
Unemployment rate	−813.43	−549.33	−0.46 (0.39)
Employment	−802.77	−544.12	−0.51 (0.45)
Inflation	−1563.58	−1156.22	−0.52 (0.35)

"ML" denotes the natural logarithm of the marginal likelihood. The threshold variable is always mean-adjusted capacity utilization.

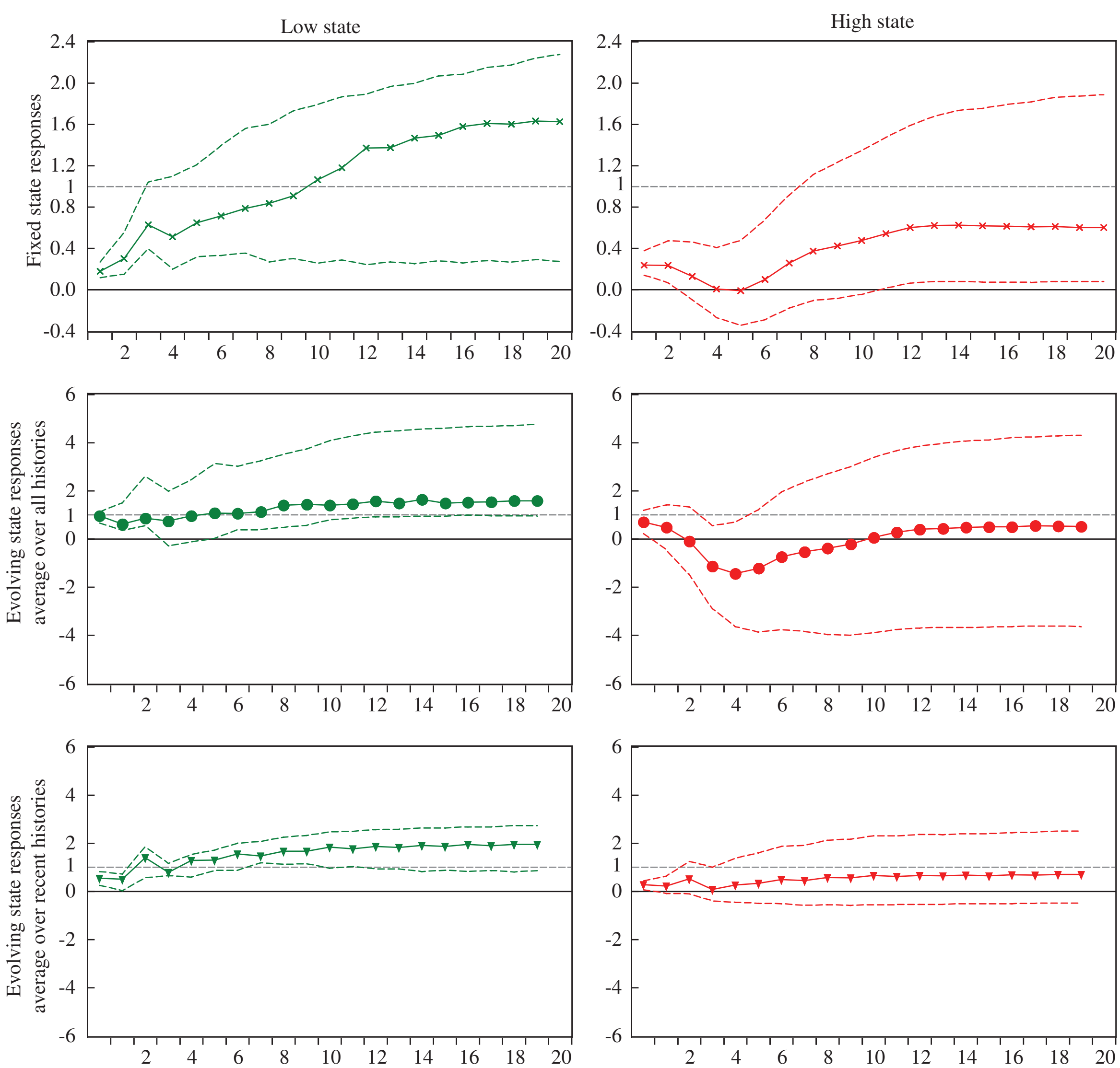

Figure 2 Responses of output to a government spending shock.
Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: low initial state, right: high initial state. Top: fixed states, middle: evolving states, averages over all histories (1967–2011), bottom: evolving states, averages over recent histories (1984–2011). Note that the vertical scale is different for the first row to highlight the distinction between the fixed low and fixed high regimes.

4.1 Responses of output to a government spending shock

As discussed above, we identify government spending shocks in the SVAR model by assuming output and its private-sector components can respond to government spending within a quarter, but government spending does not respond to output within the same quarter.[18] The results are similar when we consider alternative identification schemes; specifically, we obtain almost identical results when we reorder taxes and government spending so that spending can respond to tax shocks, when we use Blanchard and Perotti's (2002) identification scheme that imposes short-run tax elasticities, or when we add Ramey's narrative spending variable and order it first so that the rest of government spending can respond to military spending within a quarter.[19] Following the convention in the fiscal VAR literature, the responses of output are presented as cumulative level dollar-to-dollar responses, and the size of the spending shock is fixed to be equal to 1% of GDP. When we calculate the responses, we look at the responses in accumulated levels. This approach allows for direct comparison of our results with Auerbach and Gorodnichenko's (2012a) results. The responses of output components, presented in Section 4.4, are also given as dollar-for-dollar level responses. The responses of employment, unemployment, inflation, and interest rates, presented in section 4.5 are given as cumulative level responses in percentage points, and the shock to government spending is again fixed to be equal to 1% of GDP.

When constructing the impulse responses to government spending, the shock to government spending is set to be equal to 1% of GDP in the initial period. This shock initiates a dynamic path of adjustment for both government spending and other variables of interest.

Our primary results appear in Figures 2 and 3. Figure 2 shows the responses of output to fiscal spending, and Figure 3 compares the impulse response of output in different regimes. The top row of Figure 2 shows the impulse responses of output to a government spending shock for the two capacity utilization regimes, in both cases assuming that the economy remains in the same state forever. The response of output to spending shocks depends strongly on the regime. An increase in government spending pushes output up immediately in both the high and the low utilization regimes. However, in the low regime, output rises almost monotonically to a cumulative change in output equal to 1.6 times the cumulative change in government spending.

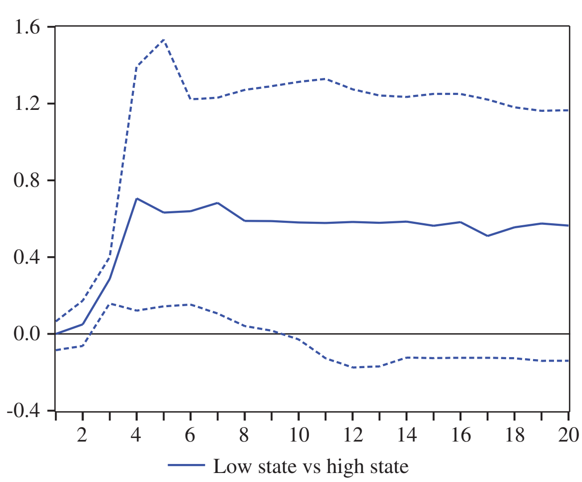

Figure 3 Difference between evolving states. The simulated difference between the evolving low state and the evolving high state, and the 90% credibility interval for the difference.

18 Note that government spending, following Blanchard and Perotti (2002), and Auerbach and Gorodnichenko (2012a) is consumption and investment. Transfers, and the associated automatic variations in spending linked to economic activity, are excluded.

19 Owyang and Zubairy (2013) also find IRFs for SVAR models are broadly robust when considering different identification schemes, including sign restrictions. They consider a linear VAR model that includes US state-level data and separates out military spending, as in Ramey (2011b).

Most of the effect takes place in the first 3 years (although the top of the credibility band hits 1.0 in just three quarters). In the high-utilization regime, the pattern is substantially different. After the initial positive response, the cumulative change in output falls back towards zero. The long-term response is positive, but the multiplier is less than half of that when output is in the low-utilization regime.

When the economy is allowed to evolve from one state to another, the magnitude of the multiplier varies depending both on the state of the economy at the time of the government spending shock and on the actual history of other shocks. As shown in the bottom two rows of Figure 2, the output response for all low states peaks at 1.6 after 2 years and then the effects of the spending shock die out. The lower bound of the credibility interval for the low-regime impulse response is strongly positive, despite the fact that we use a fairly conservative 90% credibility interval. In comparison, the average response for all high states peaks at 0.8 after 2 years, and then it remains stable, but the credibility interval always covers zero.

Figure 3 shows the estimate for the difference in the impulse responses between the evolving low regime and the evolving high regime (i.e., the difference between the left middle panel and the right middle panel in Figure 2). As shown in Figure 3, when the shock to government spending is fixed to equal 1% of GDP in both states, the difference between the mode of the average response in the low state and the mode of the average response in the high state is 0.8, and the 90% credibility interval does not include zero during the first 2 years.[20]

Thus, our estimates clearly imply that the effects of government spending on output are larger and more persistent when capacity utilization is low. In the following subsections, we examine the source of this nonlinearity in more detail. In particular, we look at the responses of output components in order to determine whether the state dependence comes from difference in the response of fiscal variables to the government spending shock or if it is due to different responses in the components of private spending.

4.2 Responses of fiscal policy to a government spending shock

From Figure 4, it is clear that the response of government spending to its own shock does not depend very strongly on the prevailing regime. In this case, the IRFs are shown as cumulative dollar-for-dollar changes in government spending relative to the size of the initial shock, because the ratio of government spending to itself is necessarily equal to one. For both regimes, the peak cumulative dollar-for-dollar change is consistent with the results obtained in the linear case by Blanchard and Perotti (2002) and similar to the results obtained by Auerbach and Gorodnichenko (2012a). Both the credibility intervals for the regimes overlap and the actual estimated responses are similar across regimes. The similar responses across regimes clearly indicate that the asymmetric response of output is not due to higher or more persistent government spending in the low-utilization regime.[21]

Figure 5 shows that the peak response of tax revenues to a government spending shock is roughly 0.8 when we account for evolving regimes, with little effect of the initial state of the economy. In the fixed low-utilization regime, tax revenues appear to increase persistently after a government spending shock, while the response of tax revenues is smaller and dies off quickly when the economy starts and remains in the high regime. But, given the wide credibility intervals for the responses at long horizons, there is no obvious evidence of state dependence in the response of tax revenues.[22] Formal impulse response comparison (available from the authors upon request), confirms these findings.

20 When using less conservative 68% credibility intervals, common in the fiscal VAR literature, the credibility interval includes zero only after 12 quarters. Results available upon request from the authors.

21 The fact that the endogenous response of government spending to an exogenous spending shock takes a number of quarters to build up helps explain the rather long response of output to the cumulative increase in government spending initiated by the shock, as shown in Figure 3.

22 It is important to note that these results are for the responses of tax revenues, not tax rates. Tax revenues are correlated with income, so part of the increase in revenues comes from increases in income due to the positive government spending shock, indicating that spending could be partially self-financing (although further analysis would be necessary to examine this possibility

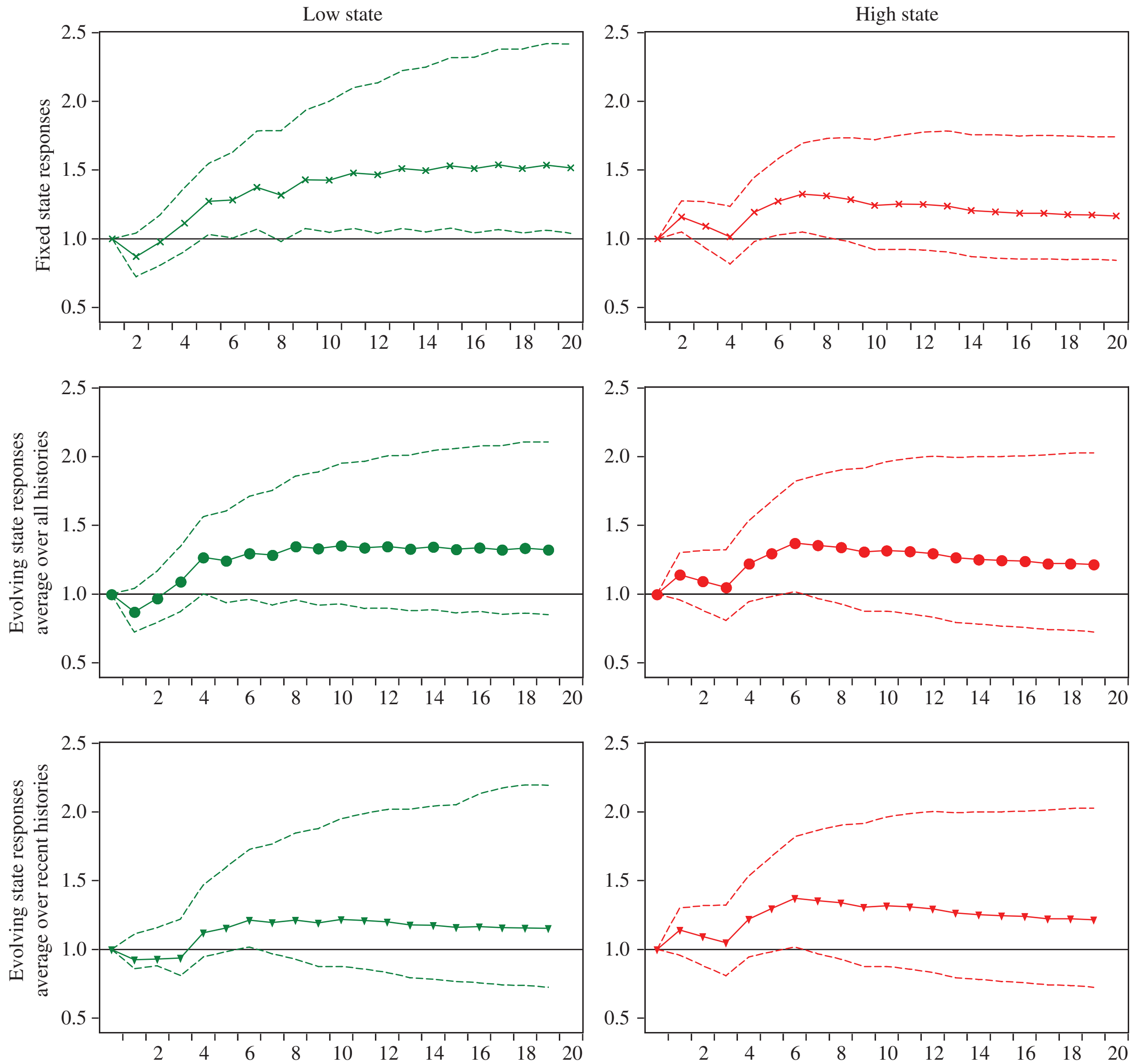

Figure 4 Responses of government spending to a government spending shock.
Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: low initial state, right: high initial state. Top: fixed states, middle: evolving states, averages over all histories (1967–2011), bottom: evolving states, averages over recent histories (1984–2011).

4.3 Responses of consumption and investment to a government spending shock

Figure 6 displays the responses of consumption to a government spending shock. The main result is that consumption increases in both regimes, but the magnitude of the response is much larger when the economy is in the low-utilization regime.[23] When starting from a low-utilization state, but allowing the state to evolve, the long-run response levels off after 3 years at close to 0.8, averaging over all histories. Consumption is much less responsive when the economy starts in a high-utilization state. The peak

given the wide credibility intervals). Another part of the increase in revenues could come from the endogenous response of tax rates to a government spending shock. The use of tax revenues also makes it difficult to interpret the responses of output and its components to changes in taxes because individuals and firms respond to marginal tax rates. Unfortunately, though, reliable data for marginal tax rates are only available at an annual frequency.

23 This result is robust to considering consumption of nondurables and services only.

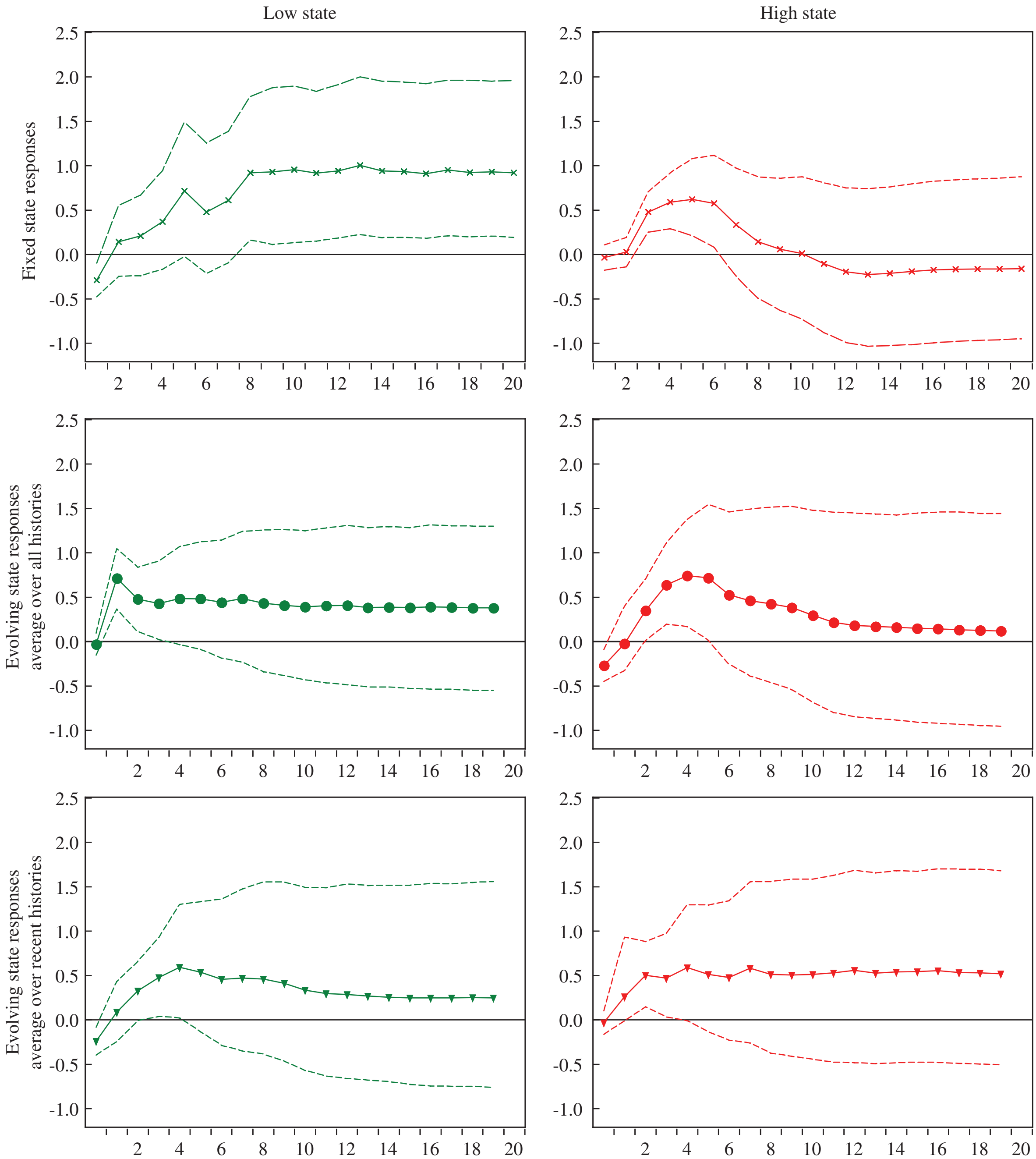

Figure 5 Responses of tax revenues to a government spending shock.
Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: low initial state, right: high initial state. Top: fixed states, middle: evolving states, averages over all histories (1967–2011), bottom: evolving states, averages over recent histories (1984–2011).

response in this case is only around 0.4, and becomes insignificant after a year. Thus, it appears that the state dependence in the response of output to government spending is at least partly due to consumption. The findings of a positive response of consumption in both regimes is consistent with the linear results obtained by Blanchard and Perotti (2002), Pappa (2009), and Woodford (2011). Also, accounting for anticipated government spending by including Ramey's military spending variable and ordering it first in the linear or nonlinear versions of the SVAR model does not change the response or the significance of the response.

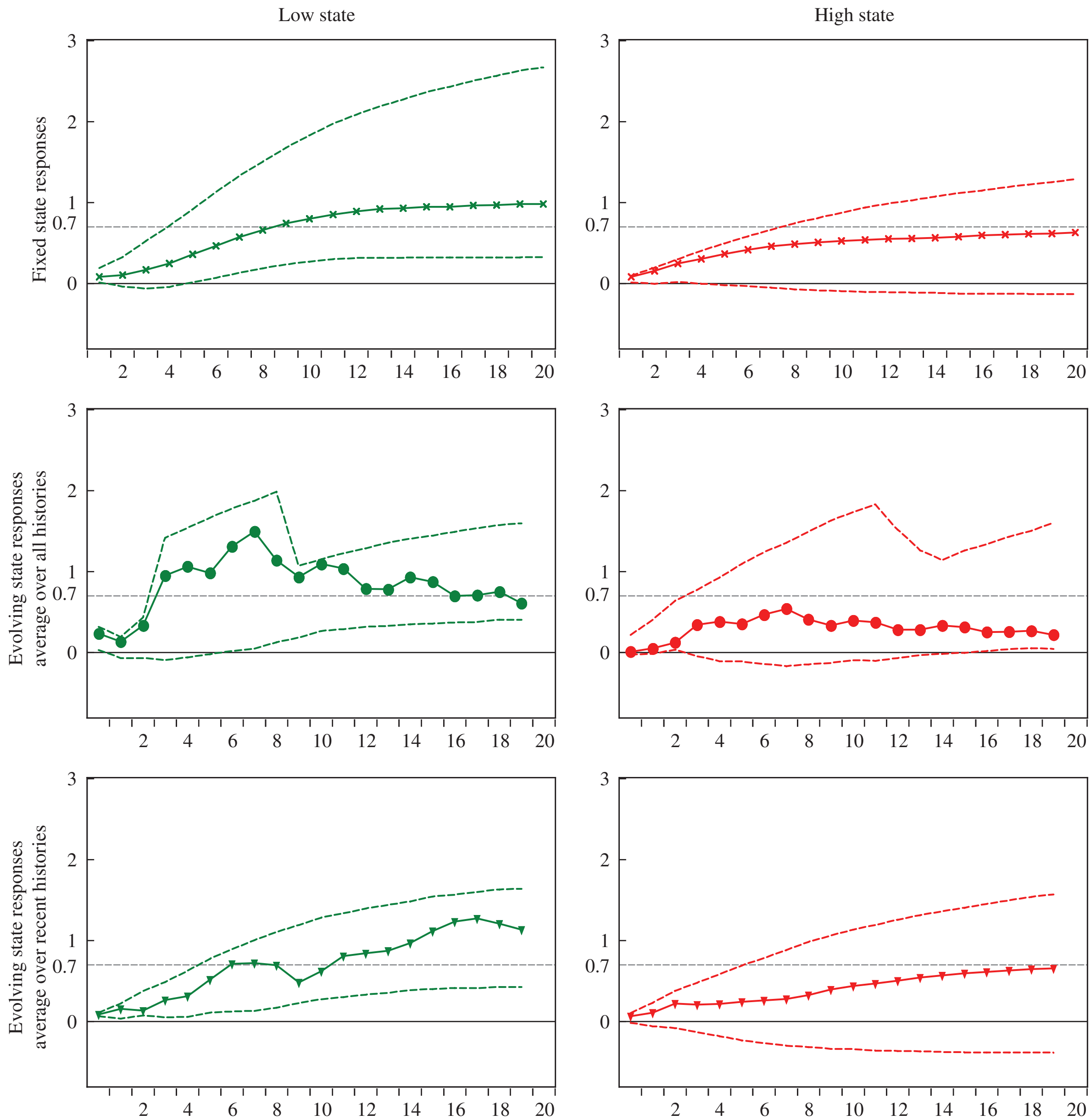

Figure 6 Responses of consumption to a government spending shock.
Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: low initial state, right: high initial state. Top: fixed states, middle: evolving states, averages over all histories (1967–2011), bottom: evolving states, averages over recent histories (1984–2011).

These results provide further support for the traditional Keynesian understanding of fiscal policy that higher government spending brings unemployed resources into use, creates higher incomes, and therefore encourages consumption. But these results are inconsistent with an alternative theoretical explanation for a positive multiplier that higher government spending today raises expected future taxes, reduces wealth, and therefore raises labor supply as workers attempt to offset the wealth shock, at least partially. While the output effect of a fiscal shock in such a model would indeed be positive, one would expect *negative* consumption effects if both consumption and leisure are normal goods.

Figure 7 displays the responses of investment, which also appear to depend on the state of the economy. In the fixed low regime, investment increases in response to government spending, with a peak response of 0.4, although the credibility interval includes zero. When the economy is assumed to remain in the high-utilization

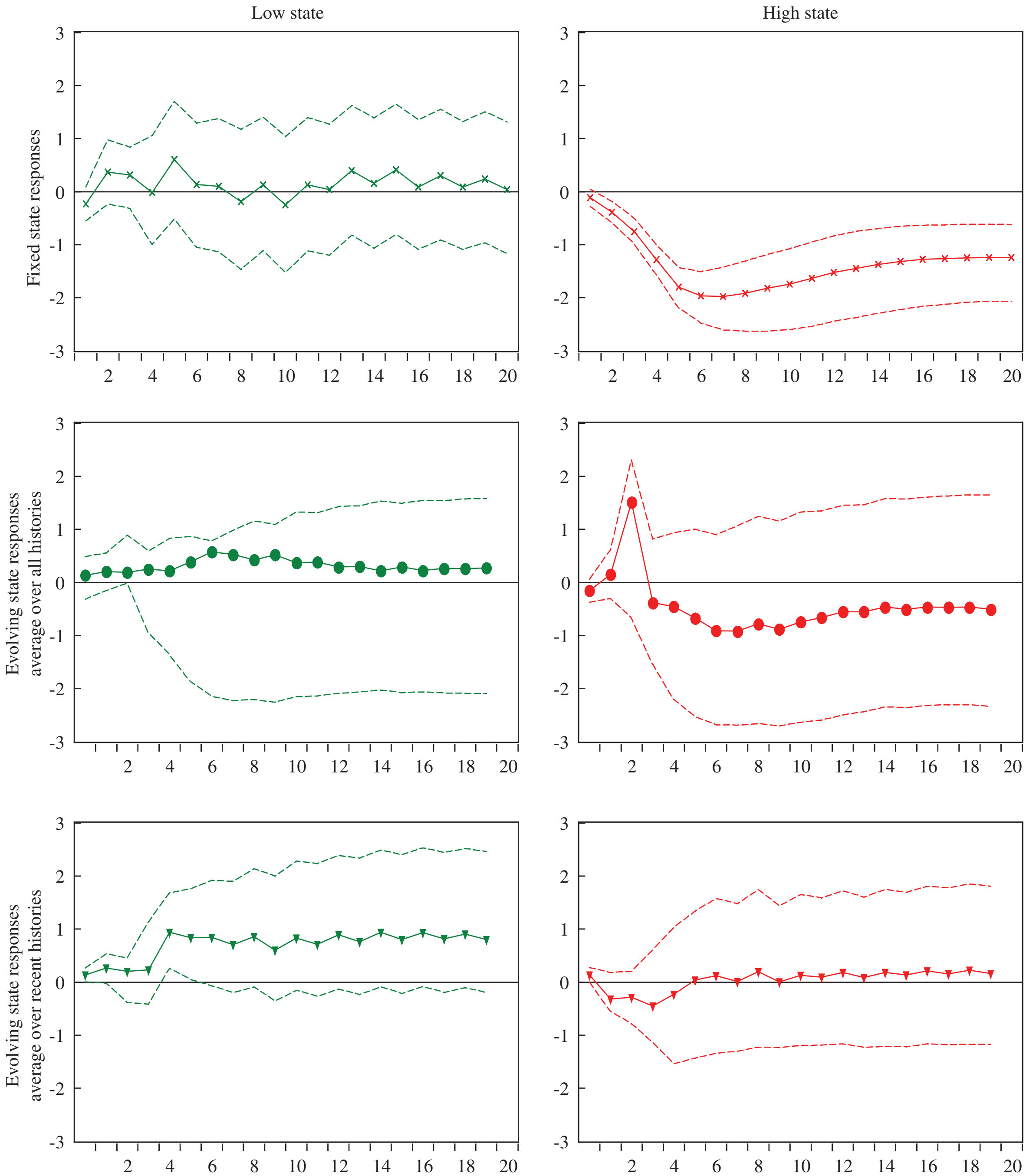

Figure 7 Responses of investment to a government spending shock.
Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: low initial state, right: high initial state. Top: fixed states, middle: evolving states, averages over all histories (1967–2011), bottom: evolving states, averages over recent histories (1984–2011).

state forever, investment drops significantly in response to a spending shock, with a cumulative decline equal to 0.9 after 5 years. Allowing the economy to evolve from one regime to another, the responses of investment are weakly positive when the economy starts from a low-utilization state and not different from zero when the economy starts from a high-utilization state. These results suggest the possibility of investment crowding out in the high-utilization state, but provide no support for crowding out in the low-utilization state. Furthermore, these results may help explain the "investment puzzle" in linear studies such as Blanchard and Perotti

(2002), that is, a negative response of investment when output and consumption respond positively, because the negative response in the linear VAR model is roughly a weighted average of the responses in the nonlinear model. Specifically, the apparent neoclassical behavior of investment found in these studies appears to reflect crowding out only when capacity utilization is high. Overall, the strong state dependence in the responses of consumption and investment suggests that much of the state dependence in the response of output is due to different responses of private spending that depend on the degree of resource utilization.

4.4 Responses of other macroeconomic variables to a government spending shock

Figure 8 shows that the unemployment rate decreases in response to a spending shock in both states. In the low-utilization regime, the unemployment rate decreases monotonically, falling by a total of 2.5 percentage points after 5 years. The effect of a spending shock on the unemployment rate is weaker and less persistent when the economy is in the high-utilization regime. The impact response is essentially zero, and the maximum response (in magnitude) is a 1.3 percentage point decline. When analyzing the magnitude of the responses, it is important to keep in mind that the impulse responses were constructed using a relatively large spending shock (1% of the GDP), which explains the large responses of the unemployment rate.

The responses of employment also exhibit state-dependence that is consistent with the responses of the unemployment rate. In Figure 9, when the economy is in a low-utilization state, employment increases by 1% after 2 years, and the long run response is equal to 0.8%. When the economy starts from a high-utilization state, the effect of a government spending shock on employment is only slightly positive and transitory. The credibility intervals for employment, however, are quite wide, and zero effects are not outside the 90% interval for either regime. This result is due to the fact that we use a conservative 90% interval and the fact that employment only builds up slowly after the shock.

The fixed-regime responses of exports and imports are very similar across regimes, suggesting little support for asymmetric responses of imports and exports to government spending.[24]

Figure 10 displays the response of the real interest rate, and Figure 11 displays the response of inflation. In the low regime, there is little response of either variable. Thus, monetary policy appears to accommodate fiscal policy when capacity utilization is low, with little implication for inflation (perhaps due to a convex Phillips curve). Notably, this accommodation of fiscal policy does not just occur in a zero-lower-bound environment (see Christiano, Eichenbaum, and Rebelo 2011, and Woodford 2011), but is the apparent response of monetary policy whenever the economy is in the low-utilization state.[25] In the fixed high regime, an increase in government spending has a more persistent effect on inflation and triggers a delayed, but large, response of the interest rate. The estimated responses are consistent with the idea that government spending can crowd out resource use, thus increasing marginal costs when the economy is close to capacity, but monetary policy responds to keep inflation under control. The responses of the interest rate and inflation in the evolving high-utilization state is large, but the credibility intervals for the responses for both variables in the endogenously evolving regime case are quite wide (due to the VAR polynomial having a root that was relatively close to 1).

5 Robustness checks

The results from the four variable baseline model presented in the previous two sections imply that there is strong evidence in favor of state-dependent effects of fiscal policy. Furthermore, they are directly

24 Figures available upon request from the authors.

25 It is also notable that the real interest does not appear to be an important variable in linear SVAR models of fiscal policy (for example, it is absent from Blanchard and Perotti's, 2002, model) or as a threshold variable in a nonlinear model (see our results in Table 2). The implication is that the different responses of monetary policy are primarily determined by the state of the economy as captured by capacity utilization regimes, not by other factors such as a binding zero nominal lower bound (which, of course, only occurs near the end of the sample period) that might influence the behavior of real interest rates.

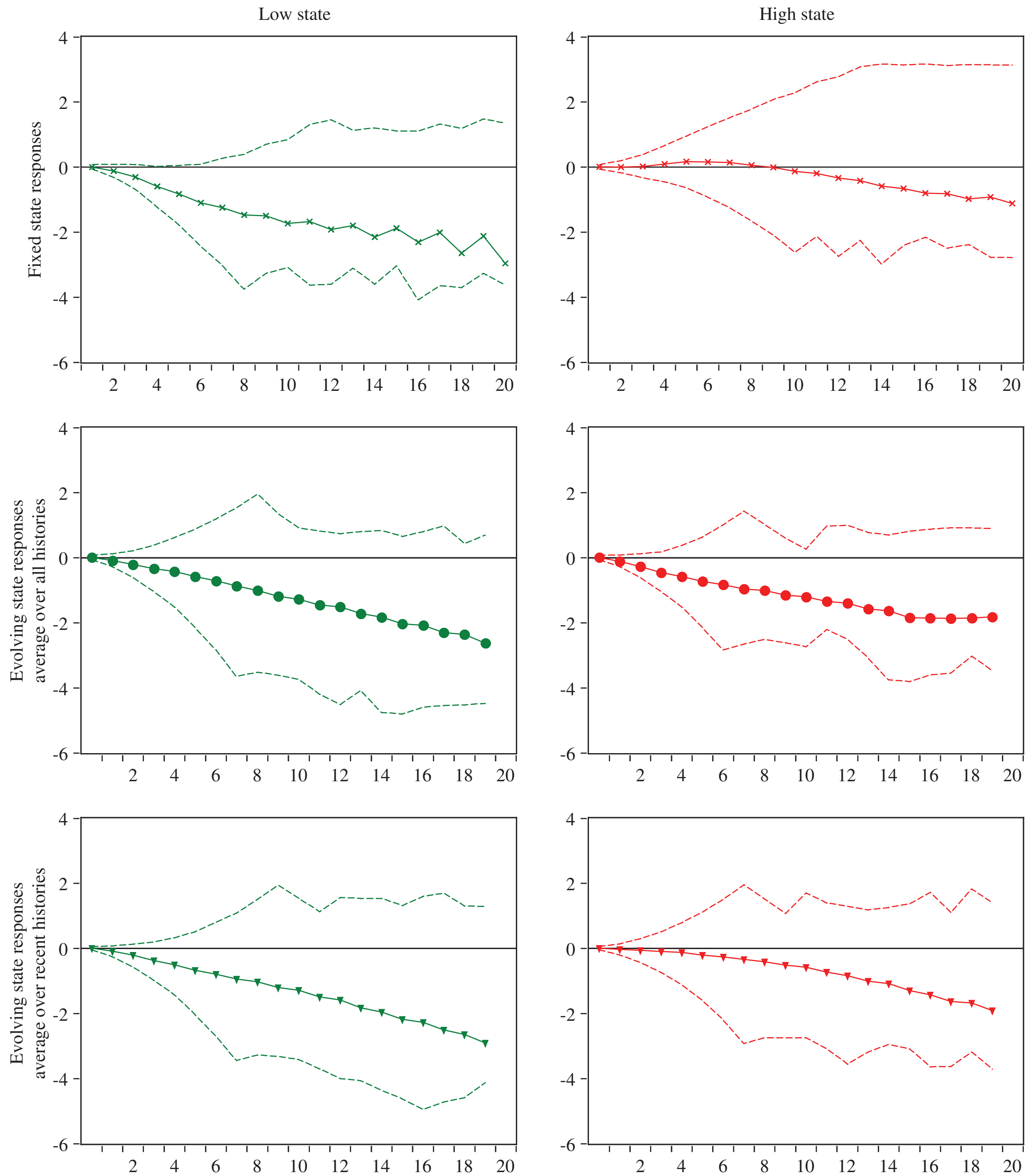

Figure 8 Responses of the unemployment rate to a government spending shock.
Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: low initial state, right: high initial state. Top: fixed states, middle: evolving states, averages over all histories (1967–2011), bottom: evolving states, averages over recent histories (1984–2011).

comparable to the benchmark results obtained by Auerbach and Gorodnichenko (2012a) and other nonlinear models that use a VAR-type model with government spending, a measure of taxes, output, and a measure of slack of the switching variable. However, it is important to note that in many of the theoretical models that allow for nonlinearities in the responses of output and output components to fiscal spending, the nonlinearity works purely through the monetary channel (see, for example, Cogan et al. 2010, or Davig at al. 2012). As soon as the interest rate returns to its natural level, the effects of a government spending stimulus die out.

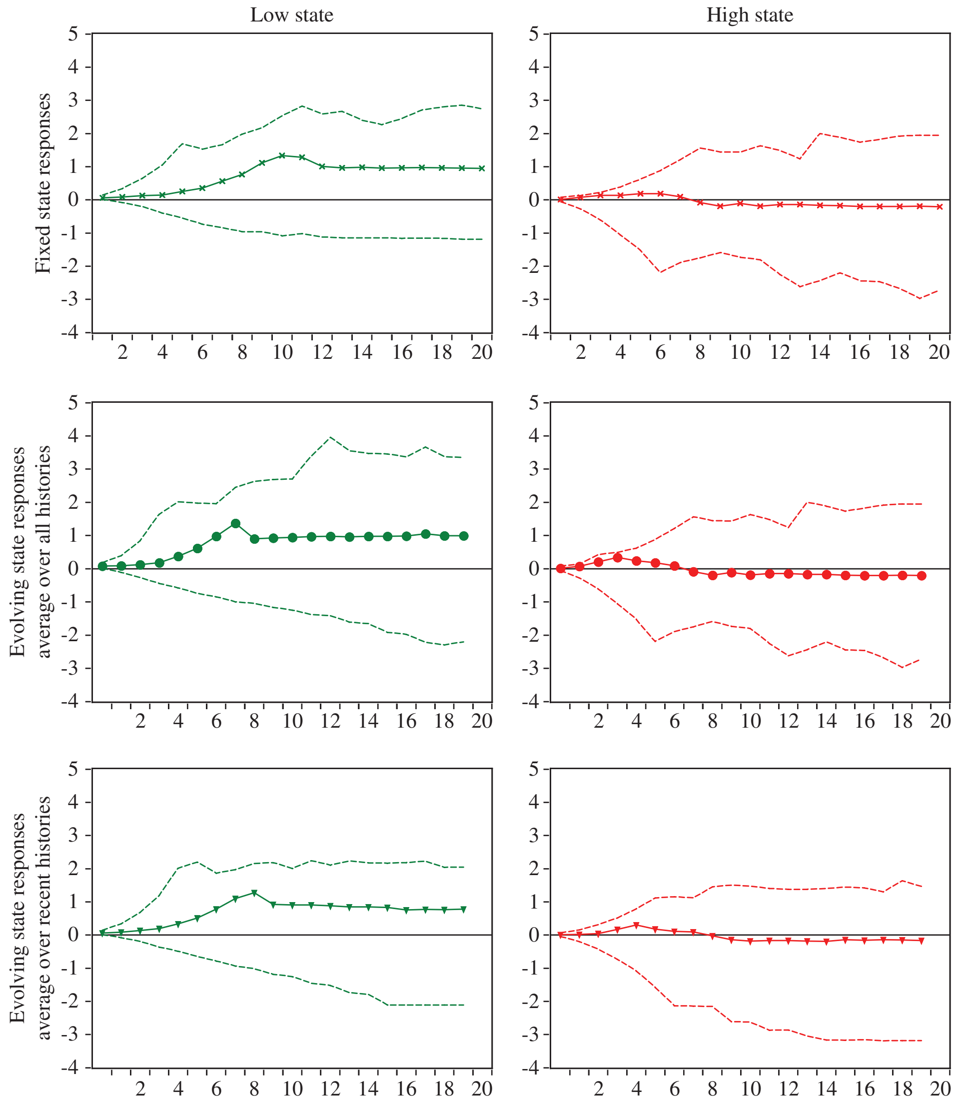

Figure 9 Responses of employment to a government spending shock.
Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: low initial state, right: high initial state. Top: fixed states, middle: evolving states, averages over all histories (1967–2011), bottom: evolving states, averages over recent histories (1984–2011).

The results from the four variable model that include the real interest rate indicate that the nonlinearity in our model does not arise primarily because of a regime switch that depends on the interest rate. However, our baseline model does not directly account for interactions between the interest rate and government spending and the effects of output. To consider the possibility that our results are partially driven by lax monetary policy that is not captured in the baseline model, and to explore the possible interaction between government spending and real interest rate, we perform a robustness check where we extend the baseline model by including the real interest rate as a fifth variable in the VAR model. Following the monetary policy literature, the real interest rate is ordered last. It is important to note that this five variable model does not fully account for all possible fiscal and monetary interactions. A complete model of fiscal and monetary policy interaction would be a much larger model that should also include asset prices, longer-term interest rates, and allow for the possibility of multiple regimes where fiscal and monetary policy have different effects. The five variable model is primarily a robustness check to account for the possibility that the nonlinearity in the response of output arises because the baseline model omits monetary policy variables, and to verify whether the baseline

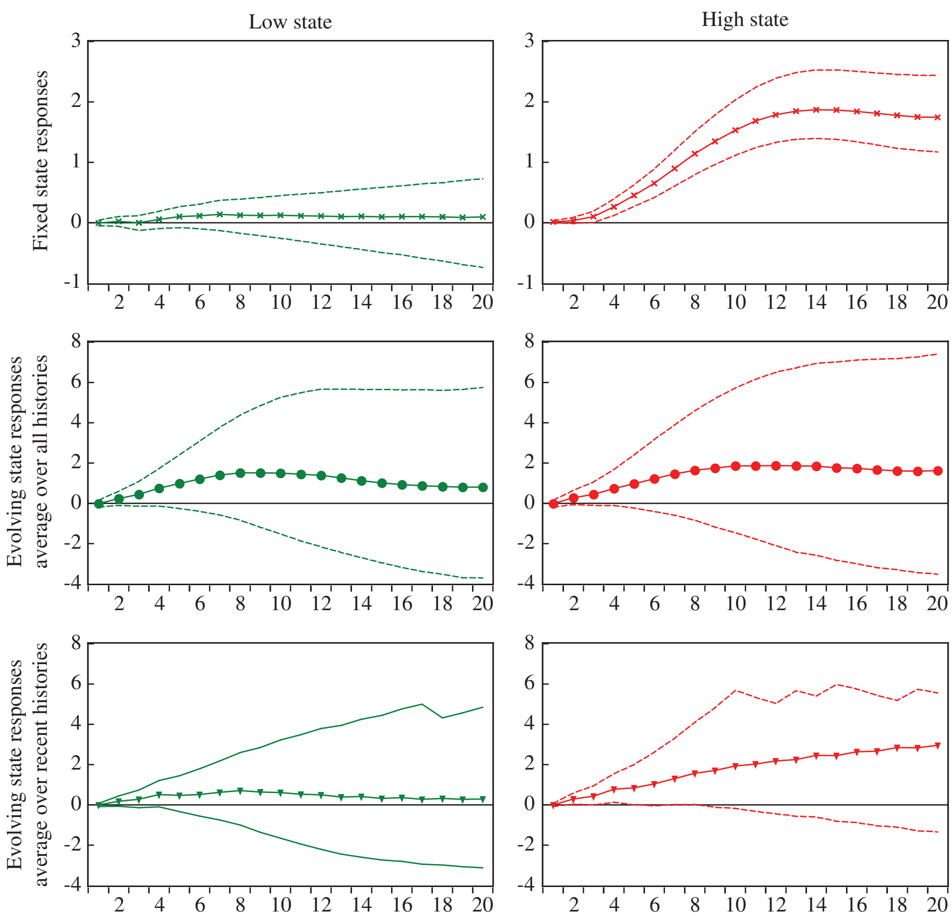

Figure 10 Responses of the real interest rate to a government spending shock.
Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: low initial state, right: high initial state. Top: fixed states, middle: evolving states, averages over all histories (1967–2011), bottom: evolving states, averages over recent histories (1984–2011). Note that the vertical scale is different for the first row to highlight the distinction between the fixed low and fixed high regimes.

model is an adequate tool for evaluating nonlinearities in the relationship between government spending and output.

Table 4 summarizes the results from the marginal likelihood comparisons and the estimated thresholds for different measures of slack for the model that includes interest rates. The marginal likelihood results are almost identical to the results from the four variable model. The credibility intervals for the estimated thresholds are wider than in the smaller model and asymmetric, but the credibility intervals overlap the intervals from the smaller models, and the point estimates are very similar. The wider credibility intervals are not surprising, given the fact that the autoregressive matrices in the larger model have 210 estimate coefficients compared with 136 autoregressive coefficients in the baseline model. It is important to note that while the posterior modes are quite similar to the posterior modes from the smaller model, the posterior densities for the threshold had a smaller second mode that split the sample into recessions and all other periods. The second mode coincided with time periods identified from the literature on asymmetric effects of monetary policy (see, for example, Balke 2000, or Lo and Piger 2005). The increase in the number of parameters and the bimodality help explain the asymmetric credibility intervals for the estimated threshold values, and the wider credibility intervals for all parameters.

Because the five variable model is highly parameterized, rather than averaging over all low histories and all high histories, we test whether the impulse response functions are different for particular histories of interest. In particular, we compare the responses of output and the real interest rate for two histories: starting

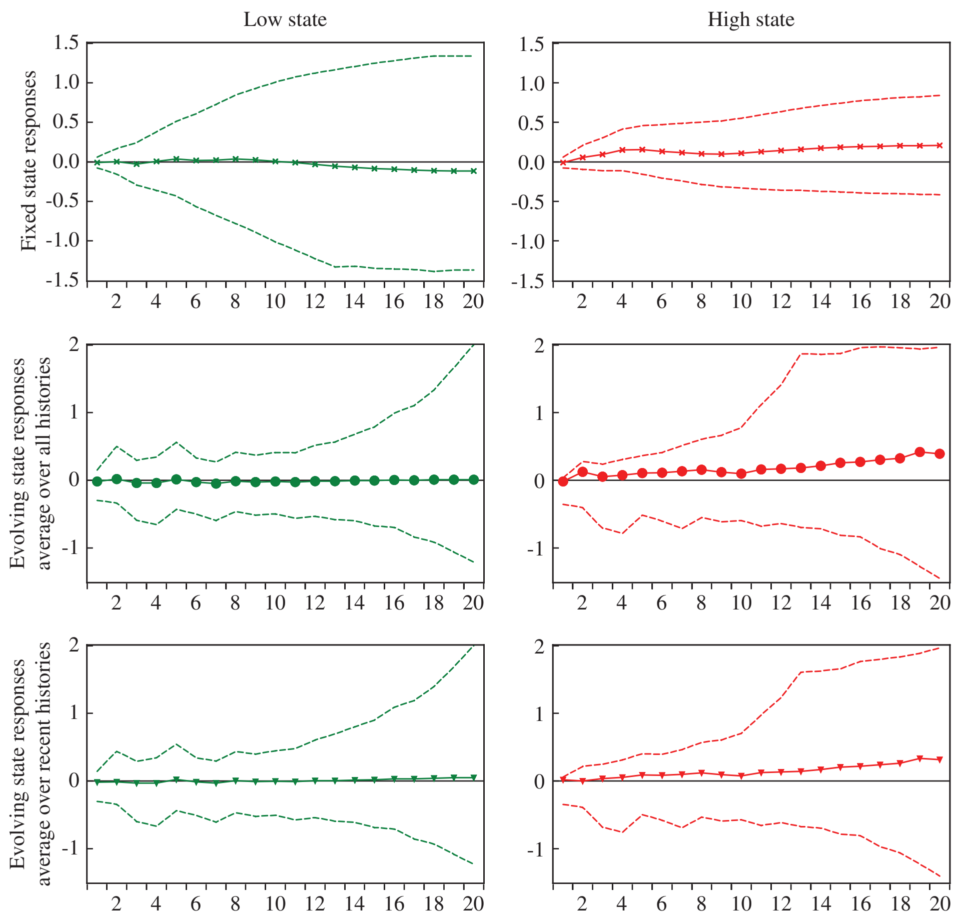

Figure 11 Responses of the inflation rate to a government spending shock.
Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: low initial state, right: high initial state. Top: fixed states, middle: evolving states, averages over all histories (1967–2011), bottom: evolving states, averages over recent histories (1984–2011). Note that the vertical scale is different for the first row to highlight the distinction between the fixed low and fixed high regimes.

in 2008q4, when the economy was in the middle of the Great Recession, and in 1997Q2, a period when the economy was in the middle of a prolonged boom. Even when accounting for uncertainty in the threshold estimate, the first history clearly falls in the low regime, and the second history falls in the high regime.

Table 4 Marginal likelihoods and estimated thresholds for the extended model with different threshold variables.

Threshold variable	Lag	Marginal likelihood	Threshold estimate
None	–	−993.43	–
Output growth	2	−920.69	0.26 (−0.1, 0.4)
Output gap	1	−911.12	1.13 (−0.2, 1.8)
Capacity utilization	1	−988.77	78.33 (76, 82)
Capacity utilization (adjusted)	1	−890.71	0.17 (−0.26, 0.53)
Unemployment rate	4	−903.81	5.31 (4.75, 6.1)
Unemployment rate (adjusted)	3	−894.63	0.09(−0.41, 0.45)
Debt-to-GDP ratio	2	−1004.07	51.1 (38.5, 62.3)
Real interest rate	2	−1000.54	5.10 (1.1,7)

The threshold estimate is the posterior median (with 90% CI in parentheses). Preferred variables for each category listed in Section 2.3 are as stated, with the preferred debt measure being total federal debt outstanding.

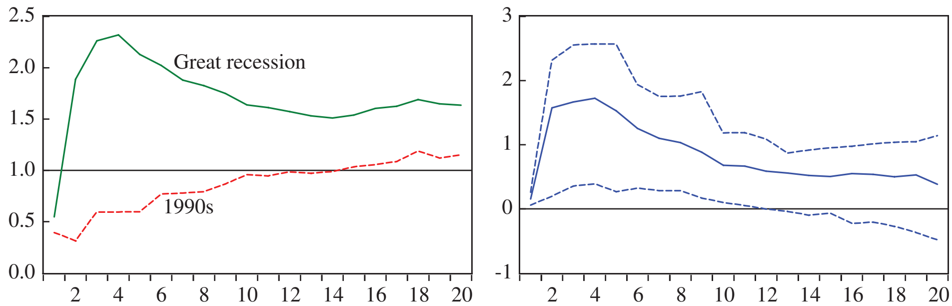

Figure 12 Modal responses of output from five variable model. Left: Great Recession vs Late 1990s. Right: difference between the modal responses (solid) with equal-tailed 90% credibility bands (dashed).

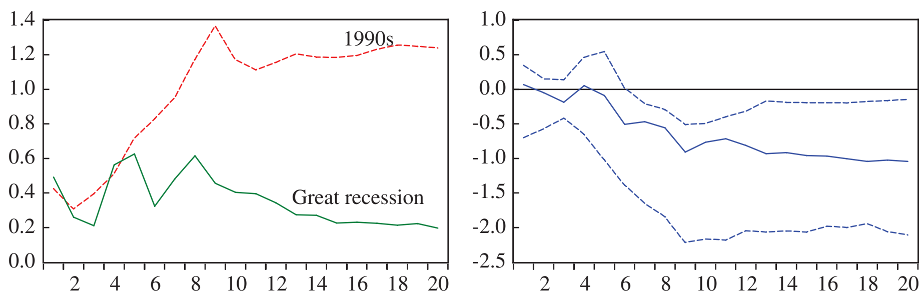

Figure 13 Modal responses of interest rates from five variable model. Left: Great Recession vs Late 1990s. Right: difference between the modal responses (solid) with equal-tailed 90% credibility bands (dashed).

Figure 12 plots the modal responses of output and the difference in the responses between the two histories, and Figure 13 plots the responses of the real interest rate from the five variable model.

The responses of output are significantly different across states in the 3 years following the spending increase. The interest rate increases in the high regime, and this increase is significant, while the increase in the low regime is small and not significant. The difference in the responses of the interest rate is significant, even when using conservative 90% CIs. These results are quite similar to the results presented in Figures 2 and 3 and in Figure 10, confirming that the asymmetry in the baseline model was not driven by the fact that it did not directly allow for the possibility of interaction between fiscal and monetary policy. Because the estimated thresholds, the marginal likelihood comparisons, and the impulse response comparisons confirm the results of the baseline model, the results indicate that the asymmetry in the responses of output is primarily driven by the state of the economy, and that our baseline model is adequate for evaluating the degrees of these asymmetries.

6 Conclusions

We present strong empirical evidence in favor of non-linear, state-dependent effects of fiscal policy. In particular, the estimates from a threshold structural vector autoregressive model clearly identify different responses of the economy to government spending shocks depending on whether the economy has high or low utilization of economic resources. We find that a rise in demand from the government sector causes large and persistently positive effects on output when the economy is operating with low capacity utilization. This effect is much smaller and less persistent when capacity utilization is above an estimated threshold for our model. It is particularly interesting to note that the estimated threshold for capacity utilization is such that a majority of observations for the US economy over the past 40 years are in the regime in which spending shocks have larger and more persistent effects.

We find no evidence that higher government spending crowds out consumption. Indeed, consumption rises after positive government spending shocks in both the high- and low-utilization regimes, but the increase is almost twice as large during low utilization periods. Most of the increase in the private components of output comes from the increase in consumption. These results for consumption are consistent with the linear results obtained by Blanchard and Perotti (2002), Perotti (2008) and Pappa (2009), but are at odds with the simulation results obtained using most calibrated dynamic stochastic general equilibrium (DSGE) models. Only when allowing for a high proportion of rule-of-thumb consumers, Galí, López-Salido, and Vallés (2007) find large responses of consumption in a calibrated DSGE model. In addition, the state-dependent responses of consumption are potentially related to the results obtained by Kaplan and Violante (2011), who develop a life-cycle model that endogenizes the proportion of rule-of-thumb consumers in order to examine the effect of taxes on consumption when a large proportion of the consumers' wealth is tied up in illiquid assets such as real estate. Historically, the number of credit-constrained consumers rises in recessions, and the Great Recession started with the crash of the housing market, which likely implied a large increase in the proportion of credit-constrained consumers in its aftermath. See Anderson, Inoue, and Rossi (2013), for an analysis of fiscal policy given heterogenous credit-constrained consumers. Our findings are also consistent with Canzoneri et al. (2013) who calibrate a New Keynesian DSGE model with costly financial intermediation and show that countercyclical shocks to the spread between rates paid by borrowers and received by depositors implies countercyclical fiscal multipliers.

Regardless of the exact mechanism behind the state-dependent effects of fiscal policy, the implications for policy are straightforward and significant. Higher government spending raises output, but this effect is both larger and more persistent when capacity utilization is low. At these times, including during recessions, higher government spending increases output, consumption, and investment. Although stimulus policy may increase government debt, the effect is smaller than a simple calculation would suggest because higher government spending raises output, income, and therefore tax revenue, and the effect of spending stimulus on public debt is less than dollar for dollar.

Further extensions of this work will explore policy implications more deeply. In particular, because our "low-utilization" regime prevails in at least half of the sample period, it would be interesting to consider whether allowing a third regime would identify recession effects when stimulus policy might be even more effective. Also, beyond the state-dependent nonlinearities found here, there may be additional asymmetries in the response of output to the size and sign of changes in fiscal policy. In addition, we plan to explore the effects of higher government spending on the dynamics of government debt in more detail. Finally, we have made preliminary analysis of tax shocks and found some comparable results to those for government spending shocks. But identifying tax shocks is challenging due to a lack of availability of quarterly data on tax rates instead of tax revenues, for which movements are largely endogenous (see, for example, the May 2012 issue of the *American Economic Journal: Economic Policy* for a number of studies illustrating the challenges in identifying the effects of tax shocks, even within a linear framework). Thus, we leave a more complete analysis of possible state-dependent effects of tax shocks for future research.

Acknowledgments: We thank Alan Auerbach, Rudi Bachmann, Jared Bernstein, Lutz Kilian, Raul Santaeulalia-Llopis, Jeffrey Sheen, Houston Stokes, John Taylor, Pao-Lin Tien, and conference and seminar participants at Deakin University, the International Monetary Fund, University of Melbourne, University of Otago, the 2012 VUW Macro Workshop, the 2012 MIFN Meetings, the 2012 SNDE Meetings, the 2012 CAMA Macro Workshop, the 2012 Berlin Conference of the Macroeconomics and Macroeconomic Policies (FMM) Research Network, the 2011 EC2 Conference, the 2011 SCE Meetings, the 2011 Midwest Econometrics Group Meetings, and the 2010 Washington University Graduate Student Conference for helpful comments. All errors are our own. We also thank Valerie Ramey for providing the narrative military spending data set. Fazzari and Panovska acknowledge the generous support of The Institute for New Economic Thinking. Panovska also gratefully acknowledges the support of the Graduate School of Arts and Sciences and the Olin Fellowship program at Washington University in St. Louis. Morley acknowledges financial support from the Australian Research Council (Discovery Grant DP130102950 on "Estimating the Effects of Fiscal Policy").

Appendix

A Techical Appendix: Bayesian Estimation

A.1 Simulating from the posterior distributions

For the linear version of the baseline model, we assume that the prior for the conditional mean parameters is multivariate normal, the prior for the variance matrix is an inverse Wishart distribution, and the prior for the scale parameter λ is a Gamma distribution. Specifically, the linear model is simply $Y_t=\Phi_0+\Phi(L)Y_{t-1}+\lambda_t\varepsilon_t$, where $\Phi(L)$ is an autoregressive matrix polynomial with roots strictly outside the unit circle, $\lambda_t=1$ for t = 1967q1,...,1983q4, $\lambda_t=\lambda$ for t = 1984q1,... T_{final}, where T_{final} is the final observation, and ε_t is i.i.d. Gaussian random variable with mean 0 and variance- covariance matrix Ω that does not change over time. Then, letting $\Phi=vec(\Phi_0)|vec(\Phi_1)...|vec(\Phi_p)$, we assume that the prior for Φ is a normal distribution, truncated to the stationarity region, with mean equal to 0, and variance-covariance matrix equal to V_n. The scaling parameter λ is assumed to have an inverse gamma prior with parameters $\alpha/2$ and $\beta/2$, and we impose an inverted Wishart prior with v_0 degrees of freedom and a scale matrix R_0. Letting $x_t=[1\; y_{t-1,1}...y_{t-1,k}...y_{t-p,k}]\otimes I_p$, it is straightforward to see that $\Phi|\Omega, \lambda, y$ is Gaussian with variance

$$V=(V_n^{-1}+\sum_{t=p+1}^{T}x_t'(\lambda_t\Omega)^{-1}x_t)^{-1}$$

and mean

$$\mu=V^{-1}(\sum_{t=p+1}^{T}x_t'(\lambda_t\Omega)^{-1}y_t).$$

Similarly, $\Omega|\Phi, \lambda, y\sim IW(v_1, R_1)$ where $v_1=v_0+T-p$ and $R_1=[R_0^{-1}+\sum_{t=p+1}^{T}(y_t-x_t\Phi)'\lambda_t^{-1}(y_t-x_t\Phi)]^{-1}$. The inverse Wishart distribution is a standard distribution, so we can sample Ω conditional on the other parameters directly. Conditional on the other parameters and the data, λ has an inverse gamma distribution with parameters $\alpha_1=(\alpha+T_p)/2$ and $\beta_1=\beta+0.5\sum_{t=t_1}^{T}(y_t-x_t\Phi)'\Omega^{-1}(y_t-x_t\Phi)$ where t_1=1984Q1, and T_p is the number of observations after 1984Q1. Under these assumptions, we can sample the model parameters directly using the Gibbs sampler.

For the threshold model in (1), it is straightforward to show that $\Phi|\Omega, \lambda, \gamma, y$ is Gaussian with mean and variance as before, except now $x_t=[x_t^*\; x_t^* I[q_{t-d}>c]]\otimes I_p$, where $x_t^*=[1\; y_{t-1,1}\ldots y_{t-1,k}\ldots y_{t-p,k}]$ and the distribution is truncated such that the VAR model is stationary in each regime. The conditional distribution of Ω is inverse Wishart and the conditional posterior distribution of λ is gamma, as before.

Conditional on c and the threshold variable, the model is linear in Φ and Ω. Estimating the linear model by splitting the sample into two subsamples yields the conditional estimators $\hat{\Phi}$ and $\hat{\Omega}$. The estimated threshold value (conditional on the threshold variable and the delay lag) can be identified uniquely as

$$\hat{c}=argmax_{c\in\Gamma_n} lik_n(c|q, d)$$

where Γ is approximated by a grid search on $\Gamma_n=\Gamma\cap\{q_1, q_2, ..., q_n\}$ and lik_n denotes the log likelihood. To ensure identification, the bottom and top 15% quantiles of the threshold variable are trimmed. We use the estimated value $\hat{c}$ for constructing the proposal for the first draw of the MH algorithm. Given a sufficiently large burn-in, the value of $\hat{c}$ does not affect the Bayesian estimates, but it provides us with a plausible starting value for the mode and it enables us to easily compare the Bayesian mode with the maximum likelihood estimate.

A potential issue is that the grid search makes it infeasible to obtain the variance of the estimate of c based on numerical derivatives. Instead, we follow the suggestion in Lo and Morley (2013) for constructing a proposal density for a threshold parameter. In particular, we obtain a measure of the curvature of the posterior with respect to c by inverting the likelihood ratio statistics for the threshold parameters based on the assumption that the parameter estimator is normally distributed and the LR statistics is $\chi^2(1)$. We use the 95% CI for the likelihood ratio statistics to obtain a corresponding standard error for c, based on an asymptotic equivalence between the inverted LR and Wald-based confidence intervals. Even though the distributional assumption and equivalence is not correct due to the nonstandard distribution of the threshold parameter and related LR tests, this approach still provides a sense of the curvature of the posterior, which is all that is needed for the proposal distribution for the sampler. Specifically, at the ith iteration of sampler, the transition density for $\gamma^{(i+1)}$ is a Student-t distribution with mean equal to $c^{(i)}$ and variance equal to $\kappa\hat{\sigma}_c^2$, where $\hat{\sigma}_c^2$ is obtained as described above. The parameter κ is calibrated on the fly to ensure acceptance rate between 20 and 60%.

To ensure that the results are robust to the choice of priors, we estimate the model by using different hyperparameters for the priors, and by using different functional forms for the priors (when the priors are not conjugate to the posteriors, we draw all parameters using a multi-block MH step). Also, to check for convergence for each combination of priors, we start the algorithm from different points, and we use a large burn-in for all runs of the MH algorithm. In particular, we use a burn-in sample of 20,000 draws and make inference based on an additional 50,000 MH iterations The results presented and discussed in the main text are based on the priors in Table 5.

Table 5 Priors.

	Distribution	Mean/Location	Variance/Scale
Φ	Multivariate Gaussian	0	$100{*}I_k$
Ω	Inverse Wishart	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$25{*}\mu$
λ	Gamma	1	0.75
c	Uniform	0.165	0.365^2

A.2 Marginal likelihoods and model comparison

When comparing two models, M_1 versus M_2, in a Bayesian setting, each model consists of the prior probability that we assign to that model $p(M_i)$, which is tells us how likely we believe the model is ex-ante, the prior distribution for all of the parameters of the model, $\pi(\theta_i)$, and the likelihood function for that model conditional on the data and the parameters, $f(data|\theta_i, M_i)$. To compare models, we can compute $Prob(M_i|data)$, which is the probability that model i is the correct model, given the data. This probability can be computed using the Bayes theorem:

$$P(M_i|data)=\frac{P(M_i)*f(data|M_i)}{f(y)}=\frac{P(M_i)*\int f_i(y|\theta_i,M_i)\pi_i(\theta_i)}{f(y)}=\frac{P(M_i)*\int f_i(y|\theta_i,M_i)\pi_i(\theta_i)}{\sum_j P(M_j)*\int f_i(y|\theta_j,M_j)\pi_j(\theta_j)}$$

where j=1, 2, ..., N are all of the models under consideration. The integral

$$\int f_i(y|\theta_i, M_i)\pi_i(\theta_i)$$

is the marginal likelihood for model i. The marginal likelihood can be interepreted as the expected value of the likelihood function with respect to the prior distribution. The higher the odds ratio, the higher the support in favor of model M_1. The Bayes factor is the ratio of the marginal likelihoods for two models (it gives us the ratio of the expected likelihoods, not taking into account any priors we may have put on the models ex-ante, before looking at the data). If two models are considered equally likely ex-ante, that is, if the researcher has no reason to believe that one model is more likely than another before looking at the data, the Bayes factor is simply the ratio of the marginal likelihoods. In that case, the distance between the marginal likelihoods tells us the probability of model 1 relative to model 2, given the data. If we have a model with marginal log likelihood lml_1 and a model with marginal log likelihood lml_2, and they are both equally likely ex-ante, the probability of $modelM_1$ relative to model M_2 is $exp(lml_1)/exp(lml_2)$. If the researcher puts different prior probability on different models, the posterior odds ratio depends on the prior probabilities, but if $lml_1>lml_2$, this implies that the odds ratio in favor of M_2 relative to M_1 is large only if we are willing to put a really high ex-ante probability on M_2 being the true model. It is, however, important to note that a large difference in the marginal likelihoods between the non-linear and the linear model does not directly imply that there is necessarily a difference in the size of the fiscal multipliers. It merely implies that there is strong evidence that at least one of the coefficients in the matrices Φ_0^2 or Φ_1^2 is different from zero. The model comparison is a useful first step that can help us evaluate whether there is any reason to use the nonlinear model at all. To compare whether this nonlinearity that is detected by the model comparison affects the spending multipliers, we look at the impulse response functions directly.

B Simulation-based impulse response function and impulse response comparison

The procedure for computing the generalized impulse response functions (GIRFs) follows Koop, Pesaran and Potter (1996), with the modification of considering an orthogonal structural shock, as in Kilian and Vigfusson (2011). The generalized impulse response is defined as the effect of a one-time shock on the forecasted level of variables in the model, and the response is compared against a baseline “no shock” scenario.

$$GIRF_y(k, shock_t, \Psi_{t-1})=[E[Y_{t+k}|shock_t, \Psi_{t-1}]-E[Y_{t+k}|\Psi_{t-1}]]$$

where k is the forecasting horizon, Ψ_{t-1} denotes the initial values of the variables in the model. The impulse response is then computed by simulating the model. The shock to government spending is normalized to be equal to 1% of GDP (at the time the shock occurs). The $GIRF_y$ response for a given draw $\Theta^{(i)}$ of the MH algorithm is generated using the following steps:

1. Pick a history Ψ_{t-1}. The history is the actual value of the lagged endogenous variable at a particular date.
2. Pick a sequence of forecast errors ε_{t+k}, k=0, 1, ..., 20. The forecast errors are simulated assuming an independent Gaussian process with mean zero and variance-covariance matrix equal to $\lambda_t^{(i)} * \Omega^{(i)}$.
3. Using Ψ_{t-1}^r and ε_{t+k}, simulate the evolution of Y_{t+k} over l+1 periods. Denote the resulting path $Y_{t+k}(\varepsilon_{t+k}, \Psi_{t-1})$ for k=0, 1, ..., l.
4. Using the Cholesky decomposition of Ω_t to orthogonalize the shocks, solve for the government spending shock at time t, replace it with a shock equal to 1% of GDP, and reconstruct the implied vector of forecast errors. Denote the implied vector of forecast errors as $\varepsilon_t^{shock_t}$, the sequence of forecast errors as $\varepsilon_{t+k}^{shock_t}$, and the resulting simulated evolution of Y_{t+k} over l+1 periods as $Y_{t+k}(\varepsilon_{t+k}^{shock_t}, \Psi_{t-1})$ for k=0, 1, ..., l.
5. Construct a draw of a sequence of impulse responses as $Y_{t+k}(\varepsilon_{t+k}^{shock_t}, \Psi_{t-1})-Y_{t+k}(\varepsilon_{t+k}, \Psi_{t-1})$ for k=0, 1, ..., l.
6. Repeat steps 2 to 5 for B times, with B=500, and average the sequences of responses to obtain a consistent estimate of the impulse response function conditional on the history and the size of the shock.
7. To obtain the average response for a subset of histories, repeat steps 1–6 for a the subset of histories of interest, and report the response averaged over all histories.
8. In order to compare the responses for two types of shocks for a fixed history and, or the responses for two different histories, we construct the difference

$$\Delta IRF = IRF_y(k, shock_t^1, \Psi_{t-1}^1) - IRF_y(k, shock_t^2, \Psi_{t-1}^2).$$

Because the impulse responses are nonlinear functions of the parameters, their distribution of both the generalized impulse responses and the significance ΔIRF is nonstandard and it is not necessarily symmetric around the mean. In this case, reporting the median value is unlikely to be adequate, as the median may not be a valid measure of central tendency. In order to circumvent this problem, we adapt the approach proposed by Inoue and Kilian (2013). For a given history, we evaluate the impulse response function for each draw of the MH algorithm, drawing the entire impulse response function for periods 1 through 20. Then we average over the histories of interest, and we evaluate the posterior likelihood of the impulse response for that draw of the algorithm. The impulse response function with the highest average posterior likelihood is then used for inference. To construct the $(1-\alpha)*100\%$ credibility interval, we order the posterior likelihood values, and we include the impulse responses whose posterior likelihood was in the upper $(1-\alpha)*100$ percentile. This method results in a "credibility cloud" with a shot gun pattern because we draw entire impulse responses rather than responses for each individual point in time. For easy interpretation, we report only the outer points of the cloud. To convert the responses to dollar-for-dollar or jobs-for-dollar responses, all the impulse responses are converted to cumulative responses, and then scaled using the ratio $G_t/Variable$ for every t.

References

Anderson, E., A. Inoue, and B. Rossi. 2013. "Heterogeneous Consumers and Fiscal Policy Shocks." Working Paper.

Auerbach, A. J., and Y. Gorodnichenko. 2012a. "Measuring the Output Responses to Fiscal Policy." *American Economic Journal: Economic Policy* 4 (2): 1–27.

Auerbach, A. J., and Y. Gorodnichenko. 2012b. Fiscal Policy after the Financial Crisis. Eds. Alesina, A. and F. Giavazzi. Chicago: University of Chicago Press.

Bachmann, R., and E. R. Sims. 2012. "Confidence and the Transmission of Government Spending Shock." *Journal of Monetary Economics* 59: 235–249.

Balke, N. 2000. "Credit and Economic Activity: Credit Regimes and Nonlinear Propagation of Shocks." *The Review of Economics and Statistics* 82 (2): 344–349.

Baum, A., and G. B. Koester. 2011. "The Impact of Fiscal Policy on Economic Activity over the Business Cycle: Evidence from a Threshold VAR." Deutsche Bundesbank Discussion Paper 03/2011.

Baum, A., M. Poplawski-Ribeiro, and A. Weber. 2013. "Fiscal Multipliers and the State of the Economy." IMF Working Paper WP/12/286.

Billi, R. 2011. "Output Gaps and Monetary Policy at Low Interest Rates." Federal Reserve Bank of Kansas City Economic Review (3): 63–87.

Blanchard, O., and R. Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *Quarterly Journal of Economics* 117 (4): 329–368.

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

Canzoneri, M., F. Collard, H. Dellas, and B. Diba. 2013. "Fiscal Multipliers in Recessions." Bern University Working Paper 1204.

Chib, S., and I. Jeliazkov. 2001. "Marginal Likelihood from the Metropolis-Hastings Output." *Journal of the American Statistical Association* 96: 270–281.

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

Cogan, J. F., T. Cwik, J. B. Taylor, and V. Wieland. 2010. "New Keynesian Versus Old Keynesian Government Spending Multipliers." *Journal of Economic Dynamics and Control* 34: 281–295.

Eggertsson, G. B., and P. Krugman. 2012. "Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach." *Quarterly Journal of Economics* 26 (3): 1469–1513.

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

Gerchert, S., and H. Will. 2012. "Fiscal Multipliers: A meta Regression Analysis." Macroeconomic Policy Institute Working Paper 97.

Hamilton, J. 1994. Time Series Analysis. Princeton: Princeton University Press.

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

Hansen, B. E. 1997. "Inference in TAR Models." *Studies in Nonlinear Dynamics and Econometrics* 2 (1): 1–14.

Inoue, A., and L. Kilian. 2013. "Inference on Impulse Response Functions in Structural VAR Models." *Journal of Econometrics* 31 (1): 78–93.
Jorda, O. 2005. "Estimation and Inference of Impulse Responses by Local Projections." American Economic Review 95 (1): 161–182.
Kaplan G., and G. L. Violante. 2011. "A Model of the Consumption Response to Fiscal Stimulus Payments." NBER Working Paper No. 17338.
Kilian, L., and R. J. Vigfusson. 2011. "Are the Responses of the U.S. Economy Asymmetric in Energy Price Increases and Decreases?" *Quantitative Economics* 2: 419–453.
Kim, C-J., and C. R. Nelson. 1999. "Has the US Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle." *The Review of Economics and Statistics* 81 (4): 608–616.
Koop, G., M. H. Pesaran, and S. M. Potter. 1996. "Impulse Response Analysis in Nonlinear Multivariate Models." *Journal of Econometrics* 74: 119–147.
Leigh, D., P. Devries, C. Freedman, J. Guajarado, D. Laxton, and A. Pescatori. 2010. "Will it Hurt? Macroeconomic Effects of Fiscal Consolidation." IMF World Economic Outlook.
Leeper, E. M., N. Traum, and T. B. Walker. 2011. "Clearing up the Fiscal Multiplier Morass." NBER Working Paper No. 17444.
Lo, C. M., and J. Morley. 2013. "Bayesian Analysis of Nonlinear Exchange Rate Dynamics and the Purchasing Power Parity Persistence Puzzle." UNSW Australian School of Business Research Paper No. 2013-05.
Lo, C. M., and J. Piger. 2005. "Is the Response of Output to Monetary Policy Asymmetric? Evidence from a Regime-Switching Coefficients Mode." *Journal of Money, Credit and Banking* 37: 865–886.
McConnell, M. M., and G. Perez-Quiros. 2000. "Output Fluctuations in the United States: What Has Changed Since the Early 1980s? " *American Economic Review* 90 (5): 1464–1676.
Mittnik, S., and W. Semmler. 2012. "Regime Dependence of the Fiscal Multiplier." *Journal of Economic Behavior & Organization* 83 (3): 502–522.
Morley, J., and J. Piger. 2012. The Asymmetric Business Cycle." *The Review of Economics and Statistics* 91 (1): 208–221.
Morley, J. 2014. "Measuring Economic Slack: A Forecast-Based Approach with Applications to Economies in Asia and the Pacific." BIS Working Paper No. 451.
Orphanides, A., and S. van Norden. 2003. "The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time." *Journal of Money, Credit, and Banking* 37 (3): 583–601.
Owyang, M. T., and S. Zubairy. 2013. "Who Benefits from Increased Government Spending? A State-Level Analysis." *Regional Science and Urban Economics* 43 (3): 445–464.
Owyang, M. T., V. A. Ramey, and S. Zubairy. 2013. "Are Government Spending Multipliers Greater During Periods of Slack? Evidence from 20th Century Historical Data." *American Economic Review, Papers and Proceedings* 103 (3): 129–134.
Papell, D. H., C. J. Murray, and H. Ghiblawi. 2000. "The Structure of Unemployment." *Review of Economics and Statistics* 82 (2): 309–315.
Pappa, E. 2009. "The Effects of Fiscal Shocks on Employment and Real Wages." *International Economic Review* 50 (1): 217–244.
Parker, J. A. 2011. "On Measuring the Effects of Fiscal Policy in Recessions." *Journal of Economic Literature* 49: 703–718.
Perotti, R. 2008. "In Search of the Transmission Mechanism of Fiscal Policy." In *NBER Macroeconomic Annual* 2007, edited by D. Agemogly, K. Rogoff, and M. Woodford, 169–226. Chicago, Illinois: University of Chicago Press.
Ramey, V. A. 2011a. "Can Government Purchases Stimulate the Economy?" *Journal of Economic Literature* 49 (3): 673–685.
Ramey, V. A. 2011b. "Identifying Government Spending Shocks: It's All in the Timing." *Quarterly Journal of Economics* 126 (1): 1–50.
Ramey, V. A., and S. Zubairy. 2013. "Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historic Data." Working Paper.
Reinhart, C. M., and K. S. Rogoff. 2009. *This Time Is Different: Eight Centuries of Financial Folly*. Princeton: Princeton University Press.
Romer, C., and J. Bernstein. 2009. "The Job Impact of the American Recovery and Reinvestment Plan." *Executive Office of The President of the United States, Council of Economic Advisers Report*.
Shoag, D. 2013. "The Impact of Government Spending Shocks: Evidence of the Multiplier from State Pension Plan Returns." Harvard University Working Paper.
Sims, C. A. 2001. "Comment on Sargent and Cogley's 'Evolving US Postwar Inflation Dynamics." *NBER Macroeconomics Annual* 16: 373–379.
Tong, H. 1978. *On a Threshold Model in Pattern Recognition and Signal Processing*. C. H. Chen, Amsterdam: Sighoff and Nordhoff.
Tong, H. 1983. *Threshold Models in Non-linear Time Series Analysis*. New York: Springer-Verlag.
Van Brusselen, P. 2009. "Fiscal Stabilization Plans and the Outlook for the World Economy." *NIME Policy Brief* 01-2009.
Woodford, M. 2011. "Simple Analytics of the Government Expenditure Multiplier." *American Economic Journal: Macroeconomics* 3 (1): 1–35.

Supplemental Material: The online version of this article (DOI: 10.1515/snde-2014-0022) offers supplementary material, available to authorized users.