Business School / School of Economics

UNSW Business School Working Paper

UNSW Business School Research Paper No. 2017 ECON 04

When Do Discretionary Changes in Government Spending or Taxes Have Larger Effects?

Steven M. Fazzari
James Morley
Irina B. Panovska

When Do Discretionary Changes in Government Spending or Taxes Have Larger Effects?

Steven M. Fazzari
Washington University
St. Louis MO 63130
USA

James Morley
University of New South Wales
Sydney NSW 2052
Australia

Irina B. Panovska*
Lehigh University
Bethlehem PA 18015
USA

Abstract

We investigate when discretionary increases and decreases in government spending or taxes have larger effects using a nonlinear vector autoregressive model with fiscal shocks identified via sign restrictions. We confirm previous empirical findings of state dependence in the relationship between fiscal policy and aggregate output, with the nonlinearity related to a broad measure of economic slack that displays strong asymmetry across the business cycle. This state dependence has important implications for the timing of stimulus or austerity measures. We find that tax cuts and spending increases have similarly large stimulative effects in periods of excessive slack, but are much less effective, especially in the case of spending increases, when the economy is close to or above potential. In terms of austerity measures designed to reduce the debt-to-GDP ratio, we find that tax increases and spending cuts are most contractionary and largely self defeating in periods of excessive slack, while only spending cuts lead to any significant reduction in the debt-to-GDP ratio when the economy is close to or above potential. The effectiveness of discretionary spending, including its state dependence, appears to be due almost entirely to the response of aggregate consumption, while the responses of both consumption and investment to discretionary taxes are state dependent, but investment appears to play the larger role in terms of their effectiveness.

JEL Classification: E32, E62, C32. Key words: Government spending, austerity measures, nonlinear dynamics, Bayesian, sign restrictions, vector autoregression
*Corresponding author. Email irp213@lehigh.edu. This draft: February 14, 2017. We thank the conference and seminar participants at the 2016 Symposium of the Society of Nonlinear Dynamics and Econometrics, the Sydney Macroeconomics Readings Group, Monash University, the Reserve Bank of New Zealand, and the Bank for International Settlements. Morley acknowledges financial support from the Australian Research Council (Discovery Grant DP130102950 on "Estimating the Effects of Fiscal Policy").

1 Introduction

Since the Global Financial Crisis, there has been an increased focus on discretionary fiscal policy. Both stimulus and austerity measures have been considered and enacted in many countries. The academic literature has increasingly documented possible state dependence in the relationship between fiscal policy and aggregate output; that is, the effects of a fiscal shock may depend on macroeconomic conditions when the policy is undertaken. Because "discretionary" implies choice, including choice about timing, state dependence opens up important questions about when it is best to implement which policy, questions that would simply be irrelevant in a linear world. In particular, in this paper, we consider the following two policy questions that become important given state dependence:

- When are discretionary spending increases or tax cuts the more effective stimulus?
- When are discretionary spending cuts or tax increases the more effective austerity measure in terms of reducing the debt-to-GDP ratio?

We make two contributions to the growing empirical literature on state-dependent effects of fiscal policy. First, we determine the source of nonlinearity driving possible state dependence by examining both small and medium scale models and different measures of economic slack. In particular, using U.S. data and a small threshold vector autoregression (TVAR) model, we find strong support for nonlinearity in terms of recurrent discrete changes in the relationship between fiscal policy and aggregate output depending on the level of a broad measure of economic slack that displays asymmetry across the business cycle.

Second, based on this form of nonlinearity, we estimate a larger-scale TVAR model for the U.S. economy with fiscal shocks identified via sign restrictions in order to address a number of concerns previously raised in the literature. Using generalized impulse response analysis for different initial conditions, we answer the questions raised above about when it is most effective to use different types of discretionary fiscal policy.

The measure of economic slack that we find drives the nonlinear relationship between fiscal policy and aggregate output is a model-averaged estimate of the output gap developed by Morley and Piger (2012) and Morley and Panovska (2017). In contrast to more symmetric measures of slack such as the CBO output gap, this model-averaged estimate displays much larger negative movements during recessions than positive movements in expansions. For the small TVAR model, we find very different dynamic cross correlations between fiscal policy and aggregate output in those periods of excessive economic slack than when the economy is close to or above potential. Notably, marginal likelihood analysis selects this measure of economic slack ahead of other broad measures of economic slack such as the CBO output gap or narrower measures such as the unemployment rate or capacity utilization. However, once we account for apparent structural breaks in the other measures of slack to ensure estimates for the threshold model are not just picking up permanent level shifts in the threshold variable, we find broadly robust results in terms of the timing and implications of the nonlinearity. Given the asymmetry in our preferred measure of slack across the business cycle, the results are also reasonably similar to those based on the related question considered in much of empirical literature following Auerbach and Gorodnichenko (2012, 2013a,b) of whether the economy is in recession versus expansion and allowing a smooth transition between regimes. However, we formally test for nonlinearity and find support for discrete transitions between regimes that are tied to the level of slack rather than the direction of movements in economic activity.

For our larger-scale TVAR model with fiscal shocks identified via sign restrictions, we find that tax cuts and spending increases have similarly large stimulative effects in periods of excessive slack, but they are much less effective, especially in the case of spending increases, when the economy is close to or above potential. In terms of austerity measures designed to reduce the debt-to-GDP ratio, tax increases and spending cuts are most contractionary and largely self defeating in periods of excessive slack, while only spending cuts lead to any significant reduction in the debt-to-GDP ratio when the economy is close to or above potential. The effectiveness of discretionary spending shocks, including its state dependence, is due almost entirely to the response of aggregate consumption, while the responses of both consumption and investment to discretionary tax changes are state dependent, but investment plays a larger quantitative role in the effectiveness of tax changes for output.

Our analysis builds on and merges different strands of the voluminous empirical literature on fiscal policy. Most closely related, a number of studies with smaller-scale nonlinear vector autoregressive models find state-dependent effects of discretionary changes in government spending--see, for example, Auerbach and Gorodnichenko (2012, 2013a,b), Bachmann and Sims (2012), Baum and Koester (2011), Baum et al. (2012), Cagianno et al. (2016), Candelon and Lieb (2013), Fazzari, Morley, and Panovska (2015, FMP henceforth), and Morita (2015). However, other studies that use a narrative approach to construct government spending shocks and employ a local projection method to compute possible nonlinear responses find less evidence of state-dependent effects--see, for example, Owyang, Ramey, and Zubairy (2013) and Ramey and Zubairy (2016). Because our larger-scale TVAR model has more fiscal variables and uses sign restrictions for identification, we are able to address potential problems in identifying discretionary spending shocks separately from built in responses to economic conditions and reconcile conflicting results in the previous literature.

A larger number of variables and use of sign restrictions also allows us to consider discretionary changes in taxes. Other approaches, such as Wold causal ordering, tend to have a difficult time identifying tax shocks because so much of the movements in tax revenues reflect endogenous responses to economic conditions. Our approach allows us to build on the linear vector autoregressive model with sign restrictions developed by Mountford and Uhlig (2009). Meanwhile, a number of recent empirical studies have considered asymmetries in the effects of stimulus versus austerity measures. Jones, Olson, and Wohar (2015) extend Romer and Romer's (2010) and Cloyne's (2013) findings by exploring whether tax cuts have different effects from tax increases when using narrative measures of taxes for the US and the UK, respectively. They find that tax cuts have significant positive effects on US output, but not on UK output, and that tax increases have no substantial effect on US output, but they have contractionary effects on UK output. Barnichon and Matthes (2015) apply a new empirical methodology to a small fiscal vector autoregressive model and find that spending cuts have larger effects than spending increases, with these increases driven primarily by very strong negative responses of output to austerity measures implemented in recessions. Guajardo et al. (2014) construct a narrative measure of fiscal consolidations and they find large decreases in output in response to these exogenous consolidations. Similarly, Jorda and Taylor (2016) find very large decreases in output following fiscal consolidations. Klein (2016) explores nonlinearity in the responses of output to the Guajardo et al. (2014) narrative measure of consolidations and finds that austerity measures have large negative effects on output when the level of private debt is high. Fotiou (2016) uses the same data set, and shows that austerity measures implemented through tax increases are self-defeating. The generalized impulse response analysis for our model allows us to investigate these and other possible asymmetries.

The rest of the paper is organized as follows. Section 2 examines the existence and source of nonlinearity in the relationship between fiscal policy and aggregate output. Section 3 presents the larger-scale threshold vector autoregressive model. Section 4 reports the generalized impulse response analysis to investigate when discretionary changes in government spending or taxes are most effective. Section 5 explores the roles of aggregate consumption and investment in driving the state-dependent effects of fiscal policy on aggregate output. Section 6 concludes.

2 Revisiting the Evidence for Nonlinearity and State Dependence

In FMP, which uses a small scale four-variable threshold model, capacity utilization with a structural break in 1973 as the threshold variable, and generalized impulse responses, we find strong state dependence in the responses of output spending. However, a recent study by Ramey and Zubairy (2016) raises the concern that our results may not be robust to considering alternative threshold variables. They also point out that there is lower correlation between capacity utilization with a structural break and the CBO output gap compared to the correlation between raw capacity utilization data and the CBO output gap. They bring up similar concerns about the sensitivity of the results presented in Auerbach and Gorodnichenko (2013 a,b).

Because a major goal of this paper is to explore when it is most effective to implement different types of discretionary fiscal policy conditional on state dependence, we revisit the evidence for nonlinearity to address Ramey and Zubairy's (2016) concerns. We start by discussing potential issues when selecting switching variables in a threshold model and then evaluate the evidence of nonlinearity and state dependence when we use different switching variables.

To investigate the evidence for nonlinearity, we consider a small threshold vector autoregression (TVAR) model. We focus on a small TVAR model to make it transparent what the possible source of nonlinearity is. However, we consider a richer large TVAR model in the next section to investigate the state-dependent effects of discretionary fiscal shocks.

Let Y_t denote the vector containing the endogenous variables in a vector autoregression (VAR) model. The TVAR model splits the process endogenously into different regimes. Within each regime, the stochastic process for Y_t is linear, but the process can evolve endogenously between regimes. Let q_{t-d} denote the switching variable that determines the prevailing regime. 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 the VAR change at time t. Define an indicator function $I[]$ that equals 1 when q_{t-d} exceeds the threshold c and 0 otherwise. The full model can be written in a single equation as

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

For the small TVAR model, Y_t contains a measure of log real government spending, log real tax revenues, log real output, and a measure of economic slack. The dynamics of the system when q_{t-d} is below c are given by Φ_0^1 and the lag polynomial matrix $\Phi_1^1(L)$, and by Φ_0^2 and the lag polynomial matrix $\Phi_1^2(L)$ when q_{t-d} is above c. The disturbances ϵ_t are assumed to be *nid* with mean zero and variance-covariance matrix Σ. The parameters Φ_i^j, the threshold c, the delay lag

d, and the number of lags included in the TVAR are estimated from the data using Bayesian methods. The technical details of the estimation are relegated to Appendix A.

A Bayesian approach to inference has two advantages in highly parametrized models such as the TVAR. In a frequentist setting, one can use the joint hypothesis $\Phi_0^2 = 0, \Phi_1^2 = 0$ to test for presence of threshold effects. However, because all threshold type VAR models are highly parametrized, conventional tests can be severely oversized (see, for example, Terasvirta and Yang, 2016). Using a Bayesian approach allows us to circumvent this problem. We directly compare the linear to the nonlinear model by using marginal likelihoods and highest posterior densities (HPD). The marginal likelihoods are calculated using Chib and Jeliazkov's (2001) algorithm and we compare models based on Bayes factors. In addition, motivated by the concerns described by Campolieti et al. (2014), we also report the expected posterior likelihoods and the HPD values for all models, which lead to very similar inference as the Bayes factors. Second, the impulse responses for the endogenously evolving system will have nonstandard asymptotic distributions that are usually non Gaussian and may depend on the history and the size or sign of the shocks, even when the true value of the parameters is known. The Bayesian sampler produces the entire posterior distribution for c, Φ_i^j and Σ conditional on the data, and we can directly account for both kinds of dispersion in the posterior distribution of the parameters by simulating the impulse responses for each iteration of the Bayesian sampler.

From a macroeconometric point of view, there are three main issues that arise in this kind of framework that could complicate selecting which measure of slack is the best switching variable, and that could complicate evaluating whether there is nonlinearity. First, the measure of slack should accurately capture the true degree of economic slack. Second, the switching variable needs to be stationary (see, for example, Hansen, 1997, or Koop and Potter, 1996, 2001). Third, the true nonlinear impulse responses may not be accurately approximated by different linear approximation methods if the estimation methods do not include higher order terms.

Unfortunately, there is no consensus in macroeconomics about how best to measure economic slack. Even settling on the output gap (i.e., the difference between actual and potential log real GDP for an economy), there are large discrepancies that arise when using different methods (see, inter alia, Morley and Piger, 2012, Morley and Panovska, 2017, or Perron and Wada, 2016). The nonlinear fiscal spending literature has used different observed variables as reasonable proxies of slack. In FMP, we used capacity utilization with an imposed structural break in 1973Q4. The capacity utilization series is survey-based, and it is not subject to significant revisions, unlike, for example, employment growth or the CBO output gap, for which there are often large revisions around the NBER turning points (see, inter alia, Billi, 2011, on the CBO output gap, and Orphanides and van Norden, 2003, on other measures of the output gap). In addition, Morley and Piger (2012) compare many different measures of the business cycle and slack obtained from a wide range of linear and nonlinear models. As an observable data series, capacity utilization is particularly highly correlated with a composite measure of slack that best matched the NBER cycle chronology and was estimated by averaging across different time series models in order to reduce estimation error. In FMP, we found that, when using formal model selection criteria, marginal likelihood comparisons very strongly preferred capacity utilization with the imposed break date as the switching variable. Meanwhile, a large number of studies use the CBO output gap (see, for example, Baum and Koester, 2011, or Baum et al. 2012) and find evidence of state dependence similar to the evidence of state dependence found in our previous study. Auerbach and Gorodnichenko (2012, 2013a,b) use moving averages of output and find state dependence, and a large number of other related studies that use the unemployment rate also find evidence of state dependence.

However, even though the evidence of state dependence in the response of output to government spending shocks is relatively well established in the previous literature, in order to formally evaluate any additional evidence for or against nonlinearity obtained by using a TVAR model, it is crucially important to ensure that the switching variable is stationary and without structural breaks, or else explicitly account for any structural breaks. This is an issue that pertains to any kind of threshold-type model. Koop and Potter (2001) show that if the switching variable has structural breaks, conventional model comparisons can erroneously identify the structural break as evidence in favor of nonlinearity.

In FMP, we imposed a structural break that matches the productivity slowdown identified by Perron and Wada (2009).[1] However, when applying tests for multiple breaks to all of the commonly used measures of slack, we find evidence of up to 5 structural breaks in capacity utilization, 5 breaks in the unemployment rate (consistent with the findings of Ghiblavi, Murray, and Papell, 2000), and up to 4 breaks in the CBO output gap. Table 1 summarizes the results of conventional tests for structural breaks in capacity utilization, the CBO output gap, and the unemployment rate.

[1] Conventional unit root and stationarity tests indicate that the mean-adjusted series with an imposed break used in FMP is stationary.

Table 1: Structural Breaks

		Break Dates and Test Statistics		
Maximum Number of Breaks	Imposed or Estimated	Capacity Utilization	CBO Output Gap	Unemployment Rate
1	Imposed (Perron and Wada Date)	1973$Q4$ $F = 29.68$ (< 0.001)	1973$Q4$ $F = 64.30$ (< 0.001)	1973$Q4$ $F = 35.78$ (< 0.001)
1	Estimated	2001$Q1$ $F = 104.95$ (< 0.001)	1974$Q3$ $F = 74.11$ (< 0.001)	1974$Q3$ $F = 38.85$ (< 0.001)
2	Estimated	2001$Q1$ 104.84 (< 0.001) 1974$Q1$ 39.62 (< 0.001)	1974$Q3$ 74.11 (< 0.001) 2008$Q4$ 67.48 (< 0.001)	1974$Q3$ 38.85 (< 0.001) 1987$Q1$ 41.27 (< 0.001)
5	Estimated	2001$Q1$ 104.84 (< 0.001) 1974$Q1$ 39.62 (< 0.001) 1987$Q1$ 23.61 (0.02) 2008$Q4$ 16.45 (0.04)	1974$Q3$ 74.11 (< 0.001) 2008$Q4$ 67.48 (< 0.001) 1984$Q2$ 28.77 (< 0.001)	1974$Q3$ 38.85 (< 0.001) 1987$Q1$ 41.27 (< 0.001) 2008$Q4$ 94.30 (< 0.001) 1996$Q1$ 28.77 (< 0.001)

The break dates were estimated using a Bai-Perron test assuming no changes in autoregressive dynamics.

The fact that there is very strong evidence in favor of a structural break in the CBO output gap around the time of the productivity slowdown explains, quite mechanically, the observation by Ramey and Zubairy (2016) that the correlation between capacity utilization that accounts for a structural break in mean and the CBO output gap is lower than the correlation between the CBO output gap and the raw capacity utilization data. While all series have a break very close to the Perron-Wada productivity slowdown break, given the evidence of multiple additional structural breaks in all series, we start by carefully re-examining the evidence in favor of nonlinearity in a small model similar to the model used by FMP, and considering different combinations and permutations of variables as the measure of slack and as the switching variable.

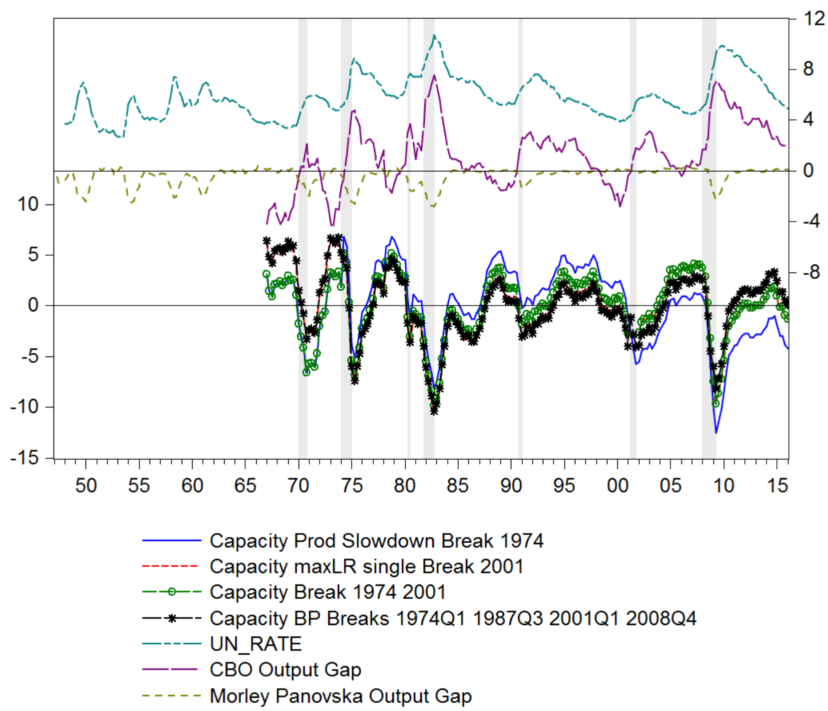
For the measure of slack in the small four variable TVAR model, we consider the following: capacity utilization with a break in 1973 (the measure used by FMP), capacity utilization with a break in 2001 (the first structural break if the break date is estimated, not imposed), capacity utilization with two breaks (1973 and 2001, both obtained by estimating the break dates),[2] the unemployment rate, the CBO output gap, and the model-averaged output gap (MAOG) from Morley and Panovska (2017).

The MAOG is obtained using equal-weights on different time series models of real GDP, as in Morley and Panovska (2017). Following that paper, we use a wide set of empirical models that are commonly used in the empirical macroeconomic literature to model the quarterly real GDP. We estimate a total of 29 different models, both linear and nonlinear, that use different definitions for the long term trend in output, and then average the estimates for the output gap across the different models. The MAOG is estimated using the full available data sample for US real GDP (1947Q1-2016Q1) and is treated as data in the TVAR.

[2] We also included models that allowed for 5 structural breaks in capacity utilization, but the results look virtually identical to the reported results for two breaks, and are available upon request from the authors.

Morley and Panovska (2017) adapt Morley and Piger's (2012) model-averaging approach, and show that the output gaps obtained using their adapted approach perform very well in terms of matching business cycle dates and correlations with narrower measures of slack not just for the US, but for a large group of OECD countries. Full details for the MAOG estimation can be found in the original studies. Figure 1 plots all measures of slack considered here.

Figure 1: Alternative Measures of Slack



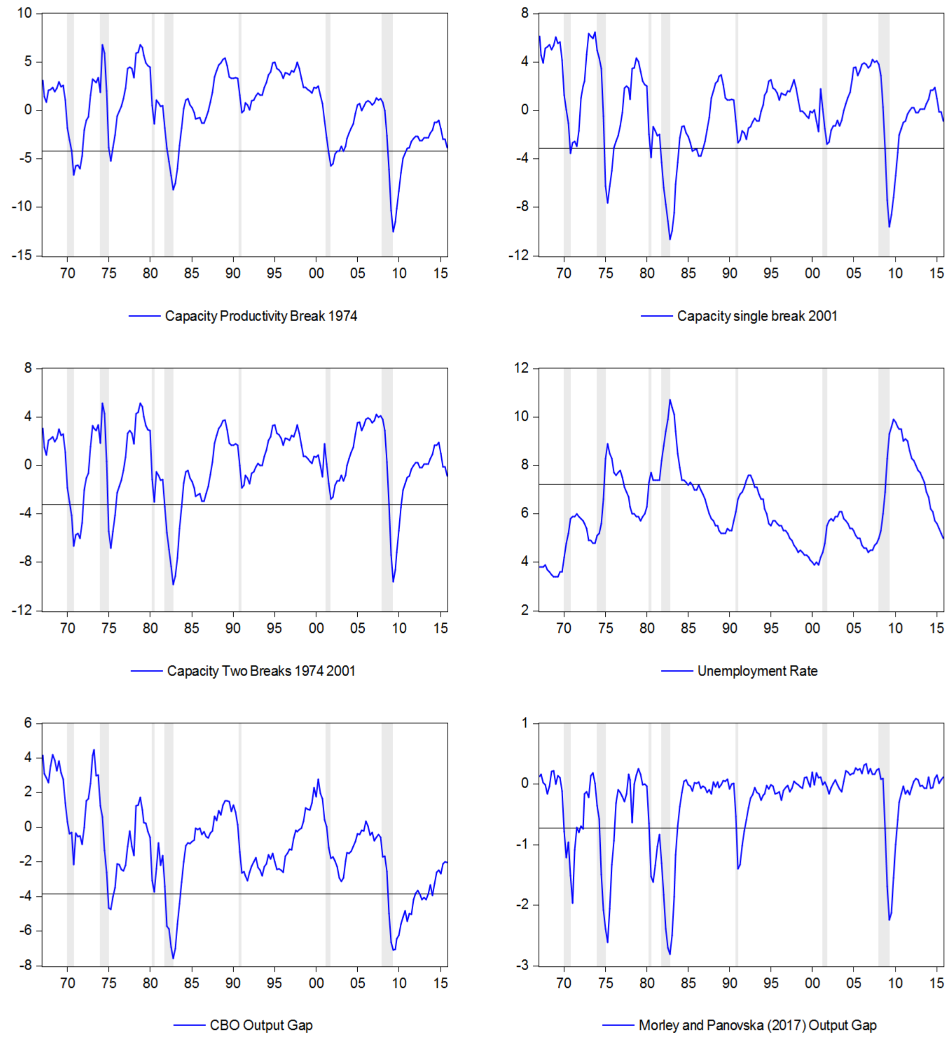
There are three minor differences from the basic model used in FMP: the data are in levels (in FMP we estimate the model using first differences, but find that the results are robust to estimating the model in levels), we use federal variables (instead of adding up federal and state and local fiscal variables), and we extend the sample through 2015Q4 (the FMP sample ends in 2012Q4). The full sample period for the estimation is 1967Q1-2015Q4. All fiscal variables are converted to real terms using the GDP deflator, and all data series were obtained from NIPA-BEA. In the next section, government spending is split into three components: consumption and investment, transfer payments, and debt payments. However, in the small four variable model in this section, government spending is defined as federal consumption and investment, which is the federal equivalent of Blanchard and Perotti's (2002) definition used by a large number of other studies, including FMP. For ease of direct comparison with FMP, the samples for all models considered here start in 1967, unless noted otherwise. Again, we consider 6 different measures of slack in the TVAR (capacity utilization with a break in 1973Q4, in 2001Q1, with breaks in 1973Q4 and 2001Q1, the unemployment rate, the CBO gap, the MAOG), and considered each of those measures of slack as a possible switching variable. Table 2 reports the threshold estimates, the highest posterior likelihood for each model, and the marginal likelihoods. The Bayes factor strongly favors the TVAR model in all cases.[3] Figure 2 plots the switching variables and the estimated thresholds when we allow for different switching variables in the benchmark FMP model. Figure 3 plots the impulse responses of output to a government spending shock for the fixed low and fixed high state each measure of slack for the best threshold variable (selected using the highest marginal likelihood for each column in Table 2).[4] The evidence in favor of state-dependence is similar to the results presented in FMP: in the low state, output responds with a large and persistent increase to an increase in government spending, and the response is smaller in the high

[3]The Bayes factor is the ratio of marginal likelihoods and equal to posterior odds ratios under even prior odds, i.e. equal prior probabilities on all models under the consideration. The ratio of the marginal likelihoods gives the relative probability of one model versus another given the data and the priors.

[4]For this small TVAR model the shocks are identified using a Cholesky decomposition, as in FMP, although we are able to use sign restrictions to identify both spending and tax shocks in the large model in the next section.

state. The pattern is consistent for the different measures of slack, indicating that our previous results were not driven by the use of capacity utilization with an imposed break.

Figure 2: Alternative Measures of Slack as the Switching Variable



Small Model Capacity Utilization from FMP Used as Measure of Slack in the VAR: Threshold Estimates for Different Switching Variables

Table 2: Model selection: Linear vs Nonlinear Models Four Variate Model

(a) Capacity Utilization With Different Break Dates as a Measure of Slack

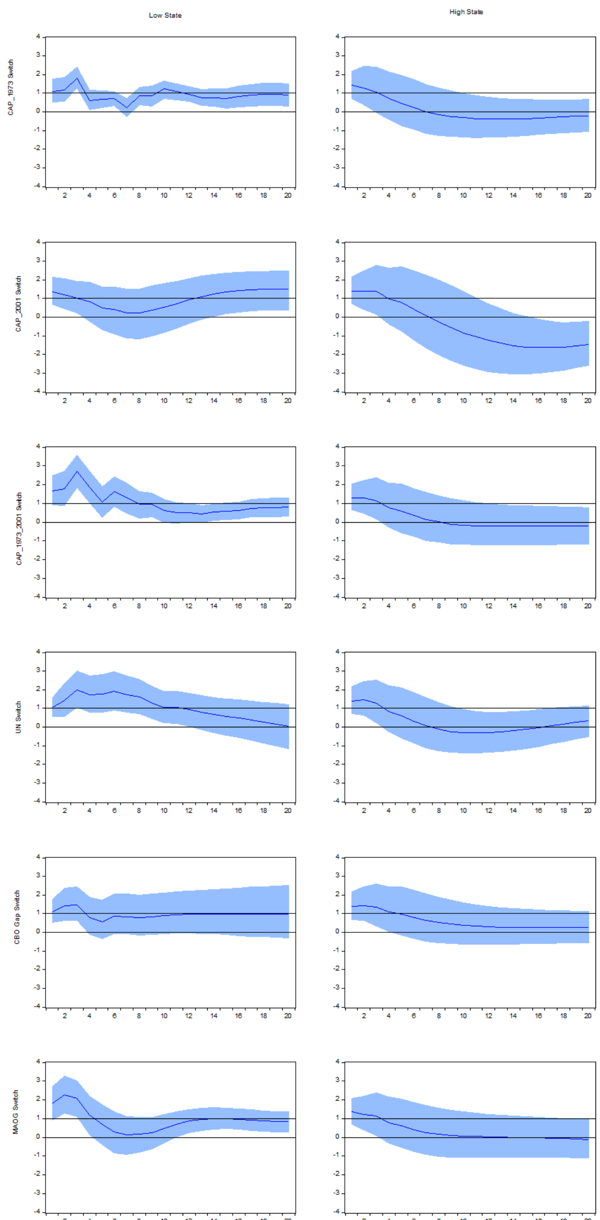
	Measure of Slack in the VAR		
Switching Variable	cap_{1973b}	cap_{2001b}	$cap_{1973,2001}$
Linear model (none)	−1257.27 −1242.82 −822.73	−1279.87 −1280.44 −388.05	−1289.77 −1288.15 −1947.62
cap_{1973b}	−1174.81 −1172.18 −333.89 −4.18 (−5.05, −2.67)	−1187.83 −1184.03 −189.85 −3.67 (−3.96, −3.32)	−1197.84 −1195.68 −683.85 −0.74 (−0.91, −0.49)
cap_{2001b}	−1173.91 −1179.26 −289.13 −3.15 (−3.46, −2.72)	−1187.30 −1183.87 −178.79 −1.54 (−1.78, −1.21)	−1205.19 −1203.58 −819.51 −1.64 (−2.47, −0.48)
$cap_{1973,2001}$	−1169.49 −1169.92 −294.18 −3.23 (−4.11, −2.42)	−1186.89 −1179.05 −178.05 −3.16 (−3.57, −2.58)	−1199.23 −1201.68 −997.34 −3.15 (−3.60, −2.53)
un	−1187.30 −1192.49 −716.72 7.21 (6.17, 8.01)	−1214.26 −1213.00 −216.12 7.20 (5.89, 8.01)	−1230.03 −1231.00 −1007.22 6.90 (5.89, 8.12)
GAP_{CBO}	−1167.52 −1162.49 −616.72 −3.84 (−4.11, −2.41)	−1204.15 −1199.86 −211.58 −3.26 (−3.71, −2.65)	−1215.26 −1220.76 −1100.32 −3.33 (−3.63, −3..05)
$MAOG$	−1168.40 −1163.20 −276.18 −0.73 (−1.02, −0.42)	−1185.61 −1182.06 −166.22 −0.71 (−0.92, −0.34)	−1206.81 −1202.72 −981.71 −0.74 (−0.91, −0.49)

(b) Other Measures of Slack

	Other Measures of Slack in the VAR		
Switching Variable	un	GAP_{CBO}	$MAOG$
Linear model (none)	−1017.77 −1028.30 −731.72	−651.08 −649.55 −1077.92	−995.36 −985.55 −477.92
cap_{1973b}	−931.13 −930.79 −445.57 −3.70 (−4.61, −0.79)	−565.98 −561.43 −671.52 −3.64 (−4.12, −1.99)	−915.19 −918.67 −295.05 −2.34 (−3.77, −0.91)
cap_{2001b}	−922.60 −926.51 −470.78 −3.20 (−3.45, −2.84)	−550.36 −551.43 −662.98 −3.05 (−4.62, −1.62)	−908.91 −911.25 −246.05 −3.11 (−3.56, −2.59)
$cap_{1973,2001}$	−925.24 −923.62 −471.05 −2.99 (−3.45, −2.35)	−557.30 − −557.56 −751.52 −3.03 (−3.69, −0.33)	−906.98 −900.20 −266.95 −3.21 (−3.81, −2.26)
un	−944.15 −943.22 −490.05 7.60 (7.00, 8.12)	−575.39 −569.40 −773.00 7.60 (6.62, 8.13)	−933.27 −938.11 −392.92 7.50 (6.61, 8.19)
GAP_{CBO}	−920.61 −928.15 −401.54 −3.66 (−3.81, −3.64)	−547.52 −547.24 −571.52 −3.33 (−3.81, −1.72)	−927.48 −932.87 −422.45 −3.03 (−3.77, −1.22)
$MAOG$	−947.09 −927.11 −387.51 −0.58 (−1.33, −0.21)	−547.95 −572.43 −542.42 −0.74 (−1.30, −0.22)	−904.48 −903.80 −222.45 −0.82 (−1.42, −0.24)

Each cell reports the likelihood obtained using a frequentist grid search procedure, the expected posterior likelihood obtained Bayesian estimation, and the marginal likelihood (T, M, B).
The second entry is the threshold estimate, including 90% credibility intervals, obtained from the posterior Bayesian distribution.

Figure 3: State Dependence for Small Model



The responses of output to government spending in the fixed low state (L) and high state (R) when using different measures of slack in the VAR. Measures of slack (top to bottom) capacity with an imposed break in 1973, capacity with an estimated break in 2001, capacity with 2 structural breaks (2001, 1973), the unemployment rate, the CBO output gap, Morley and Panovska's (2017) MAOG. Horizontal lines are at 1.

However, Table 3 illustrates the importance of taking structural breaks into account. The table summarizes the results from a model that reestimates the small four-variate model extending the sample period back to 1949. Because capacity utilization is only available after 1967, we only consider the unemployment rate, the CBO output gap, and the MAOG. As shown by the threshold estimates, the posteriors for the likelihood functions when the unemployment rate or the CBO output gap have a global mode at a value for c that is close to the historic averages, thus illustrating the criticism brought up by Koop and Potter (2001) that structural breaks could contaminate any inference about nonlinearity.[5] Meanwhile, because the MAOG already accounts for structural breaks, the threshold estimate is very close to the threshold estimate reported in Table 2. Figure 4 shows the response of output to government spending and taxes in the low state and in the high state for the extended sample when the MAOG is used both as a measure of slack and the switching variable (the MAOG was also the variable that maximized both the likelihood function and the expected posterior likelihood). The responses are in line with the responses for the shorter subsample, indicating that the results are not driven by the sample period selections. Therefore, the differences between the results obtained using GIRFs and the results obtained using the local projection method can be likely attributed to the fact that most studies that use the local projection method use a linear approximation, while a TVAR data generating process would require higher order (quadratic, cubic, or even higher order terms) for the projection to accurately approximate the true impulse responses (which are consistently estimated by the simulation-based approach used here). However, a full formal comparison between the different methods is left for further research.[6] The

[5]The likelihood functions for the longer samples were multimodal with one mode at the mean, and second smaller modes. The full likelihood surfaces are available in a Supplemental Appendix that is available upon request.

[6]This concern was first raised by Jorda (2005), in the paper that first proposed using projection method. He cautioned that incorporating higher-order terms might be necessary in

Table 3: Model selection: Linear vs Nonlinear Models Four Variate Model Extended Sample 1949Q1-2015Q4

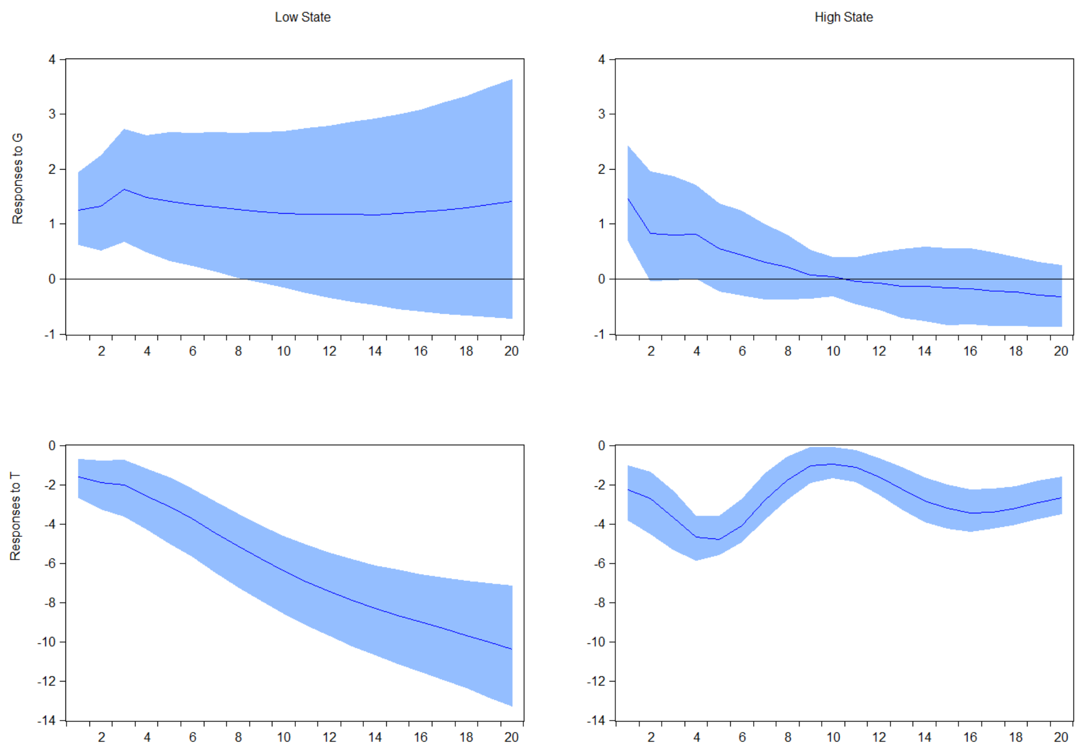
	Measure of Slack in the VAR					
Switching Variable	un		CBO_{gap}		$MAOG$	
Linear model (none)	-1544.17		-1058.14		-1486.86	
un	-1442.51	5.20 $(4.70, 5.73)$	-947.89	5.20 $(4.79, 5.62)$	-1399.34	5.19 $(4.50, 6.13)$
CBO_{gap}	-1427.18	0.94 $(-0.15, 1.65)$	-902.56	1.82 $(0.25, 3.35)$	-1391.86	2.16 $(0.15, 2.35)$
$MAOG$	-1420.53	-0.69 $(-0.98, -0.56)$	-948.60	-0.73 $(-1.02, -0.53)$	-1390.37	-0.69 $(-1.23, -0.42)$

Each cell reports the expected posterior likelihood obtained Bayesian estimation. The second entry is the threshold estimate, including 90% credibility intervals, obtained from the posterior Bayesian distribution.

model selection criteria, the threshold estimates, and the impulse responses for different switching variables and different samples yield similar results in terms of the timing and implications of the nonlinearity, thus assuaging the concerns that previous findings in favor of nonlinearity were driven by idiosyncratic behavior in a specific switching variable or by the choice of the switching variable. Given the strong support in favor of state dependence both from the previous literature and from our results, we move on to our larger benchmark model.

nonlinear models if using the projection method to approximate the true impulse responses. Similarly, LM-type tests for nonlinearity in univariate STAR models include quadratic or cubic terms in order to approximate the smooth transition model, and recently Terasvirta and Yang (2016) use a model with quadratic terms to test a smooth transition LSTVAR model against a linear VAR model. Fotiou (2016) shows that for a simulated STVAR model, the projection method can lead to impulse responses that are substantially different from the true impulse responses in some cases. In smooth transition models, a second-order Taylor approximation can be sufficient to approximate the nonlinearity, and even the linear approximation may be a good approximation if the speed of transition is relatively low. However, given the fact that the transition in a TVAR model is abrupt, an accurate approximation may require many more higher order terms. We therefore focus on using impulse responses obtained from the TVAR rather than on projection-based impulse responses.

Figure 4: State Dependence for Small Model given Long Sample (1949Q1-2015Q4)



Responses of output to G (top row) and T (bottom row). MAOG as a measure of slack and switching variable. Fixed low (L) and high state (R) with 90% CIs.

3 A Relatively Rich Nonlinear Model of Fiscal Policy and Aggregate Output

3.1 Model comparison

For our benchmark specification of the larger-scale TVAR model considered in the rest of this paper, the vector Y_t includes nine variables: log real federal consumption and investment spending, log real federal transfer payments to persons, real federal interest payments on debt, real transfer taxes, other tax revenues in real terms, real GDP, a measure of slack, interest rates (measured using the Federal Funds Rate), and inflation (calculated using the GDP deflator). By focusing only on federal variables, we can trace out the impacts on debt and deficits. In particular, if the total federal debt at time t is $Debt_t$, then

$$Debt_t = Debt_{t-1} + G_t + GTrPay_t + GIntPay_t - TrTax_t - OtherTax_t$$

and the debt to output ratio can be calculated as

$$d_t = d_{t-1} + \frac{G_t + TrPay_t + IntPayments_t - TrTax_t - OtherTax_t}{Y_t}$$

where d_t is the ratio at time t.[7] Most fiscal stimulus or austerity measures that involve discretionary changes in government spending are usually implemented by changes in government consumption and investment. The two additional spending variables are included to ensure that debt is traced out more accurately. Government transfer payments to persons are strongly affected by the state of the business cycle. While changes in transfer payments are occasionally used as a fiscal policy tool (for example, the unemployment benefit extensions during the Great Recession), a lot of the movements in transfer payments are likely to be endogenous.[8] Similarly, government interest rate payments are affected by the historic patch of government spending and taxes, and are endogenous,

[7] An alternative way to track debt would be to account for evolution in interest rates and inflation, as in Favero and Giavazzi (2012). We have also considered their specification in preliminary analysis. However, the responses of inflation and interest rates are small in magnitude, and the results obtained using Favero and Giavazzi's specification for debt applied to our model are very similar to those obtained using our ratio. The results are available upon request.

[8] When we perform a variance decomposition using HPD values, the bulk of the variation in transfer payments is explained by business cycle shocks (61% on impact, 91% after 8 quarters), and they explain very little of the variation in output (peaking at 6% on impact, but becoming insignificant after 1 quarter).

but both transfer payments and interest rate payments affect the federal deficit. We also split the tax series into two sub-components. The first one is transfer taxes, which is dependent on the state of the business cycle, and rarely used as a discretionary fiscal policy tool. The second component is tax receipts net of transfer taxes (the federal equivalent of Blanchard and Perotti's tax series). Including inflation and interest rates in the TVAR ensures that the identified fiscal shocks are orthogonal to business cycle shocks and to monetary policy shocks. We use the Wu-Xia (2016) shadow rate to measure the interest rate during the zero-lower bound period.

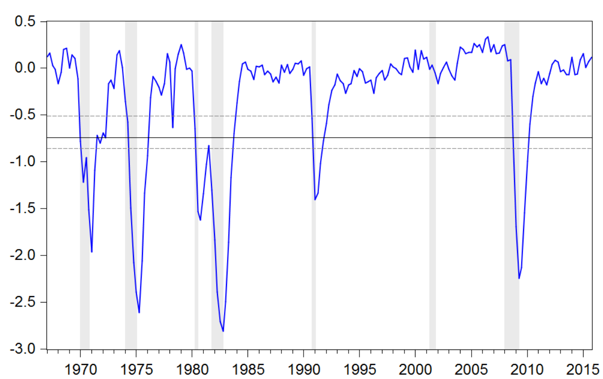
Table 4 reports the marginal likelihoods and the threshold estimates for all combinations and permutations of switching variables and measures of slack. Again, the data strongly favor the nonlinear specifications over the linear specifications in all cases, and the threshold estimates split up the sample into periods of slack (recessions and the recovery periods immediately following the recessions), and normal times. There is conclusive evidence about the existence of at least one structural break in capacity, unemployment, and the CBO output gap, as shown in Table 1. However, the evidence about the exact number and timing of structural breaks beyond the first one is somewhat inconclusive and depends on whether autoregressive terms are included and on whether we allow for heteroskedastic errors. Therefore, we use the specification with the model-averaged output gap as our main model. [9]

Figure 5 plots the switching variable (MAOG lagged 1 period), the estimated threshold, and the 90% posterior credibility intervals for the threshold. The estimated threshold is -0.69, with the 90% credibility interval being equal to

[9] However, in Appendix C (Figure C1) we also present the impulse responses for different combinations of measures of slack and switching variables to illustrate that our results are resonably robust to the choice of the switching variable and measure of slack.

$(-0.74, -0.52)$.[10] The threshold estimate roughly splits the samples into recessions and expansions, with two important exceptions. First, the initial stages of most recoveries are classified as being in the low state (which is consistent with, for example, the unemployment rate and the CBO output gap, which stay high and low during the initial stages of the recoveries). The second exception is an extended period following the Great Recession which was classified as an expansion by the NBER. Taking into account that the unemployment rate was above its mean almost until 2015, and that the CBO output gap was large and significant following the Great Recession, this classification of the post 2009 period as a period of slack is unlikely to be driven solely by some idiosyncrasy in the choice of the switching variable.

Figure 5: The Switching Variable (MAOG) and the Estimated Threshold (90% CI)



[10] As shown in Table B1 in Appendix B, the estimated threshold is very similar when the TVAR is estimated substituting consumption or investment for output.

Table 4: Model selection: Linear vs Nonlinear Models Large Model

(a) Capacity Utilization With Different Break Dates as a Measure of Slack

	Capacity as a Measure of Slack in the VAR		
Switching Variable	cap_{1973b}	cap_{2001b}	$cap_{1973,2001}$
Linear model (none)	−2670.03 −2676.38 −822.73	−2698.34 −2698.08 −627.62	−2705.10 −2704.17 −826.55
cap_{1973b}	−2243.49 −2242.44 −304.15 −0.78 (−1.48, −0.01)	−2262.57 −2256.48 −347.18 −0.78 (−1.42, −0.14)	−2268.04 −2266.71 −376.62 −0.62 (−1.43, 0.18)
cap_{2001b}	−2286.63 −2289.42 −320.18 −2.55 (−2.73, −1.25)	−2284.73 −2284.48 −344.09 −1.26 (−1.35, −1.17)	−2315.76 −2314.41 −378.85 −1.25 (−1.35, −1.17)
$cap_{1973,2001}$	−2238.54 −2239.00 −339.26 −1.35 (−1.66, −1.04)	−2259.92 −2260.41 −349.18 −0.31 (−0.38, −0.12)	−2270.77 −2265.18 −349.58 −0.30 (−0.47, −0.04)
un	−2254.28 −2255.30 −343.30 7.40 (6.20, 7.79)	−2278.88 −2275.82 −368.47 7.40 (6.10, 8.18)	−2289.50 −2289.33 −385.20 7.70 (5.55, 8.11)
GAP_{CBO}	−2240.81 −2241.10 −346.25 −3.09 (−3.17, −2.78)	−2259.91 −2259.89 −370.49 −3.10 (−3.17, −2.80)	−2272.51 −2271.01 −390.64 −3.00 (−3.14, −2.80)
$MAOG$	−2173.03 −2186.07 −281.15 −0.69 (−1.11, −0.56)	−2222.16 −2222.31 −250.31 −0.71 (−0.74, −0.69)	−2224.00 −2224.22 −325.77 −0.69 (−0.71, −0.65)

(b) Other Measures of Slack

	Other Measures of Slack in the VAR		
Switching Variable	un	GAP_{CBO}	$MAOG$
Linear model (none)	−2420.91 −2417.96 −601.62	−2038.98 −2038.00 −802.45	−2394.05 −2389.91 −833.73
cap_{1973b}	−1961.29 −1960.22 −275.32 −1.34 (−1.49, −0.56)	−1576.15 −1575.60 −576.32 −1.34 (−1.49, −0.56)	−1951.35 −1950.40 −588.44 −0.05 (−1.42, 0.21)
cap_{2001b}	−2029.51 −2028.01 −298.52 −1.22 (−1.41, −1.13)	−1615.23 −1614.00 −632.69 −1.05 (−2.00, −0.80)	−2008.64 −2007.60 −658.02 −2.55 (−3.00, −0.30)
$cap_{1973,2001}$	−1987.31 −1986.01 −292.16 −1.33 (−1.41, 0 − 1.26)	−1583.35 −1584.00 −625.38 −1.35 (−1.70, −0.83)	−1938.70 −1938.02 −568.77 −1.33 (−1.82, −0.91)
un	−1986.71 −1986.60 −304.71 7.30 (6.15, 7.80)	−1589.99 −1587.00 −602.59 7.10 (6.02, 8.83)	−1979.15 −1973.03 −651.32 7.30 (6.00, 7.92)
GAP_{CBO}	−1955.46 −1954.01 −319.34 −3.01 (−3.17, 0 − 2.79)	−1629.17 −1628.05 −633.37 −2.46 (−2.78, −1.08)	−1946.13 −1945.20 −533.26 −3.14 (−3.17, −2.80)
$MAOG$	−1937.81 −1937.00 −269.66 −0.69 (−0.74, −0.52)	−1567.08 −1566.01 −521.45 −0.70 (−0.81, −0.57)	−1873.11 −1870.00 −450.37 −0.74 (−0.86, −0.51)

Each cell reports the likelihood obtained using a frequentist grid search procedure, the expected posterior likelihood obtained Bayesian estimation, and the marginal likelihood (T, M, B).
The second entry is the threshold estimate, including 90% credibility intervals, obtained from the posterior Bayesian distribution.

3.2 State Dependence: Sign Identification and Impulse Responses

A crucial empirical question to consider with this model is whether the effects of government spending really do differ across regimes defined by economic slack, and whether, conditional on that state dependence, there is any evidence that an austerity measure will have effects that are significantly different from the simple mirrored effect of a stimulus of the same size implemented at the same time. Rejecting linearity using Bayesian model comparison directly implies that at least one of the impulse responses to at least one (structural) shock is different across regimes, but the degree of this asymmetry can only be evaluated by looking at the impulse response functions themselves.

In order to construct the impulse responses, the structural shocks have to be identified using a plausible orthogonal decomposition of the variance-covariance matrix Σ. Different strands of the fiscal literature have taken different approaches. The most popular approaches are the timing approach, the narrative approach, and the sign restriction approach. The timing approach is used by, for example, Blanchard and Perotti (2002), FMP, Baum and Koester (2011), and it entails imposing the restriction that government spending does not respond to business cycle shocks within a quarter.[11] While the timing approach can be justified using institutional knowledge in small VARs, it is much more challenging to justify it in larger VARs such at the one used here, because there is no clear guidance about the timing restrictions of the responses to all variables (for example, it is not immediately clear whether transfer taxes respond to endogenous shocks in other taxes within a period).

Ramey (2011), Owyang, Ramey, and Zubairy (2013), Ramey and Zubairy

[11]Blanchard and Perotti (2002) and the subsequent studies use a combination of timing restrictions to identify government spending shocks and structural restrictions that utilize estimated tax elasticities to identify tax shocks.

(2015, 2016), Cloyne (2013), Romer and Romer (2010), Caggiano et al. (2016), and Jones, Olson, and Wohar (2015), inter alia, use the narrative approach, or a combination of the narrative approach with the timing approach. The narrative approach uses a series of government spending shocks or tax shocks that is constructed by examining historic announcements about changes in government spending and taxes. Because our goal is to trace the effects of spending shocks both on output and on combinations of different components of fiscal spending, establishing a one-to-one link between the narrative series and the other components of fiscal spending is not immediately obvious. In addition, because a lot of studies that use the narrative approach identify military spending shocks, this means that a lot of the observations for the spending shocks equal zero, which makes exploring state-dependence challenging from an econometric point of view. Therefore, we use the sign-restriction approach to identify the structural shocks.

The sign restriction approach entails defining the number of structural shocks of interest (which can be smaller than the number of variables in the TVAR), and specifying the sign of the response of variables over a number of horizons. This approach is usually considered more general and agnostic, because it nests the zero restrictions imposed by the timing restrictions. Mountford and Uhlig (2009) use sign restrictions in a linear fiscal VAR to study the effects of fiscal policy and taxes, and they find that deficit-financed tax cuts increase output more than deficit-financed increases in government spending. Candelon and Lieb (2013) extend the model used by Mountford and Uhlig (2009) to a nonlinear model with cointegration. They find that there is strong evidence of nonlinearity in the response of output to government spending shocks, but that the multipliers are always lower than 1. Both Mountford and Uhlig (2009) and Candelon and Lieb (2013) use the penalty function approach to orthogonalize the variance-covariance matrix when identifying the structural shocks. However, recent developments in the time series literature have cast doubt on the results obtained using this approach. In particular, Inoue and Kilian (2013) point out that the means and the medians for the distributions of the impulse responses will not correspond to a single structural model. Arias, Rubio-Ramirez, and Waggoner (2015) show that the penalty function approach imposes complicated nonlinear restrictions not only on the orthogonalization matrix, but also on the VAR parameters, thus biasing the impulse responses and leading to artificially narrow credibility intervals. They propose an efficient algorithm for sampling from the posterior distributions when the structural model is identified using sign restrictions, and demonstrate that the penalty function approach leads to biased impulse responses and artificially narrow confidence intervals when applied to the VAR model used by Mountford and Uhlig (2009). In the remainder of this paper, we focus on 4 structural shocks that are identified using sign restrictions on the impulse responses, and the impulse responses are constructed using the QR decomposition approach and sampler proposed by Arias, Rubio-Ramirez, and Waggoner (2015).[12]

The four shocks of interest are a government spending shock, a tax shock, a "business cycle" shock, and a monetary policy shock. Table 5 summarizes the sign restrictions used to identify the shocks. All shocks are orthogonal to one another, which differs from Mountford and Uhlig (2009) and Candelon and Lieb (2013), who do not impose the restriction that tax shocks are orthogonal to government spending shocks.

[12]Pagan and Fry (2011) also suggest using the QR decomposition in larger VARs where the shocks are identified using sign restrictions.

Table 5: Sign Identification

	Response								
Shock	G	$TransPay$	$IntPay$	$TransTax$	$OtherTax$	Y	$slack$	i	π
G	+++	?	?	?	?	+/?	?	?	?
T	?	?	?	?	+++	?	?	?	?
BC	?	?	?	?	+ ++	+ ++	+++	?	?
MP	?	?	?	?	?	?/- - -	?	+++	- - -

Business cycle shocks increase output, tax revenues, and increase the measure of slack on impact and for 2 quarters following the shock.[13] In the next section where we consider the responses of consumption and investment, business cycle shocks increase consumption and investment.[14] Monetary policy shocks do not increase prices for 2 quarters, while they increase interest rates for 2 quarters. Because there is conflicting evidence from the monetary policy literature (see, for example, Lo and Piger, 2005, and Tanreyro and Thwaites, 2016) that shows the responses of output to monetary policy can vary and be insignificant at some points of the business cycle, and potentially be larger or smaller in recessions, we do not impose the restriction that output falls in response to a monetary policy contraction. However, it is important to note that the basic results do not change if we impose this restriction. Tax shocks are assumed to increase tax revenues for 2 quarters following the shock. Similarly, government spending shocks increase government consumption and investment for 2 quarters following the shocks.[15] Following previous results from the fiscal

[13] Recall that "slack" is defined as the difference between some measure of economic activity and its long run trend. Large negative values imply there is a lot of slack in the economy. Positive business cycle shocks would therefore, increase capacity utilization, or decrease the difference between output and trend output (when using the CBO gap or the MAOG), thus increasing the output gap. In the cases where the unemployment rate is used as a measure of slack, the signs of the responses are reversed (unemployment decreases in response to business cycle shocks).

[14] While we do not impose any restrictions on the responses of interest rates to business cycle shocks, the posterior responses indicate that interest rates increase in response to business cycle shocks.

[15] We also consider an alternative identification scheme where the restrictions are imposed for 4 quarters, and a restriction scheme where transfer payment are countercyclical. The responses look very similar to the responses presented here, and are available upon request.

literature, we also impose the restriction that output increases on impact in response to a government spending shock (see, for example, FMP and Auerbach and Gorodnichenko, 2013 a,b) and that exogenous tax increases decrease output on impact (see, for example, Romer and Romer, 2010). Even studies that find no evidence of state dependence or find that spending multiplier decline sharply after the first quarter find positive multipliers on impact (for example, Ramey and Zubairy, 2016). The responses of output are left unrestricted after quarter 1. These slightly tighter priors that restrict the responses of output on impact are based on consensus from the previous literature, and they are used merely for convenience to speed up the Bayesian estimation. However, the results presented in the next subsections do not hinge on these two restrictions. If the response of output to government spending and taxes were left unrestricted, almost the entire posterior distribution of the response of output to positive spending shocks and negative tax shocks was above zero at horizon zero, and almost the entire posterior distribution of the response of output to negative spending shocks and tax increases was below zero. Therefore, the slightly more restrictive prior is supported by the data.[16]

The responses to negative shocks have the opposite signs from the signs shown in Table 5. The responses of consumption and investment to fiscal shocks are left unrestricted. In the case when we consider the evolving-state impulse responses, the responses are constructed assuming that the economy endogenously evolves from one regime to another, and an orthogononalization is accepted if the sign restrictions hold for 2 quarters even if the economy evolves from one regime to another. The technical details of the impulse response construction are discussed in detail in Appendix A.

[16] The full set of responses for different identification schemes is available from the authors.

4 Empirical Results for the Benchmark Model

4.1 Evidence of State Dependence

Figure 6 plots the responses of output to a structural shock in government spending and to a structural shock in taxes in the fixed high state and in the fixed low state, and the posterior differences between the high and the low state. All responses are converted to dollar-to-dollar responses using the average G_t/Y_t and T_t/Y_t ratios for the corresponding low and high state for each draw of the threshold parameter. In the low state, an increase in government spending increases output by \$1.5 after 5 quarters, and the response is persistent. In the high state, an increase in government spending temporarily increases output on impact, but the response dies out and becomes negative after 2 years. A tax cut increases output both in the low state and in the high state, and the responses exhibit similar state dependence: tax cuts in the low state increase output by \$2 (dollar for dollar), whereas in the high state the response is smaller, peaking at \$1.3, and becomes insignificant after 7 quarters (vs 13 quarters in the low state). The responses in Figure 6 embed three different sources of uncertainty: uncertainty about the threshold estimate, uncertainty about the VAR parameters, and uncertainty about the orthogonalization matrix Q that is used to to identify the shocks. Even when accounting for the different sources of variability, there is still evidence of state-dependence in the response of output to government spending. Furthermore, in the high state, the responses of output to tax shocks are larger than the responses of output to spending shocks. To ensure that our results are not driven by the zero lower bound period, we also estimate the model for the period that excludes the zero lower bound (1967Q1-2006Q4). Because the results are very similar to the benchmark results, they are presented in Appendix B and Appendix C. The model comparison for the sample that excludes the zero lower bound period is presented in Table B2 in Appendix B, and the impulse responses are presented in Figure C2 in Appendix C.[17] Again, the nonlinear model is strongly preferred by the data, and the impulse responses are very similar to the impulse responses shown in Figure 6. For brevity, we report the results when the switching variable is restricted to lagged values of the measure of slack included in the VAR, but a full set for each cell in the 6-by-6 matrix for all combinations of measures of slack and switching variables is available from the authors upon request. The pattern is very similar for all measures of slack considered: strong and persistent increases in the low state (peaking above 1, with the median estimates ranging between 1.3 and 1.8, depending on the measure of slack, but the confidence intervals for all measures of slack overlap), short-lived positive multipliers that become insignificant after 6-8 quarters in the high state for all measures of slack. The results therefore conclusively support the conjecture that there is state dependence.

[17]Given the consistent results in Tables 1, 3, and 4, for brevity, Table B2 only presents the results when the measure of slack from FMP and the MAOG are used as the switching variables, and Table B3 only reports the expected posterior likelihood and the estimated thresholds. The full set of all estimates for all combinations and permutations of the measures of slack and switching variables is very similar to the results from Table 4 and is available upon request.

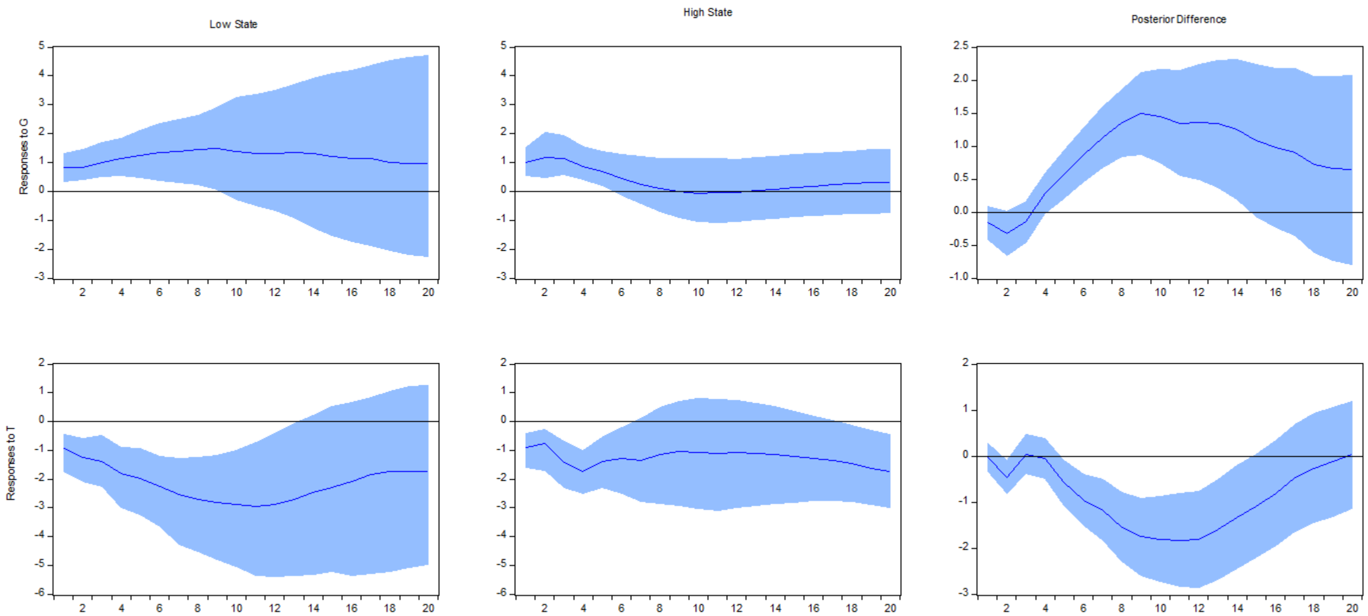


Figure 6: State-Dependence Responses of Y Fixed Low (Left) and Fixed High State (M), and posterior differences (R) with 90% CIs.

4.2 Evolving Responses

While we find conclusive evidence in favor of state dependence when comparing the fixed-state responses, from a policy perspective, policy makers are usually more concerned with the response of output (or any variable of interest) conditional on current economic conditions, rather than with the responses averaged over different historical conditions, or with the responses where the history is assumed to be fixed forever. In addition, while the fixed responses are quite useful for testing state dependence across regimes, the responses within a fixed regime are linear. If the economy is assumed to stay in one state forever, positive shocks will have exactly the opposite effect of negative shocks. However, if the economy is allowed to evolve, threshold models build in (but do not impose) the possibility that negative shocks can have different effects from positive shocks. Building on the evidence of state dependence, in order to evaluate whether there is significant evidence of sign dependence within a state, we abstract from the parameter uncertainty and we allow the economy to evolve from one state to another. We focus on three recent histories of interest that are relevant from a policy perspective: a deep recession, a sluggish recovery that would not be classified as an NBER recession but where the switching variable is close to the threshold, and a strong expansion, and we fix the parameters at the HDP values. The economy is then allowed to evolve endogenously from one state to another. We consider 3 histories that are relevant for policy:

- 1996Q1: a robust expansion, classified as being in the high state according to most studies that find evidence of state-dependence, and according to all switching variables we considered. For brevity, this will be referred to as the high state.

- 2008Q3, which is a deep recession, and would be classified as being in the low state according to our threshold and according to the vast majority of studies that find evidence in favor of state-dependence (see, inter alia, Auerbach and Gorodnichenko, 2012, 2013a,b, Caggiano et al., 2016, and FMP. For brevity, this will be referred to as the low state.

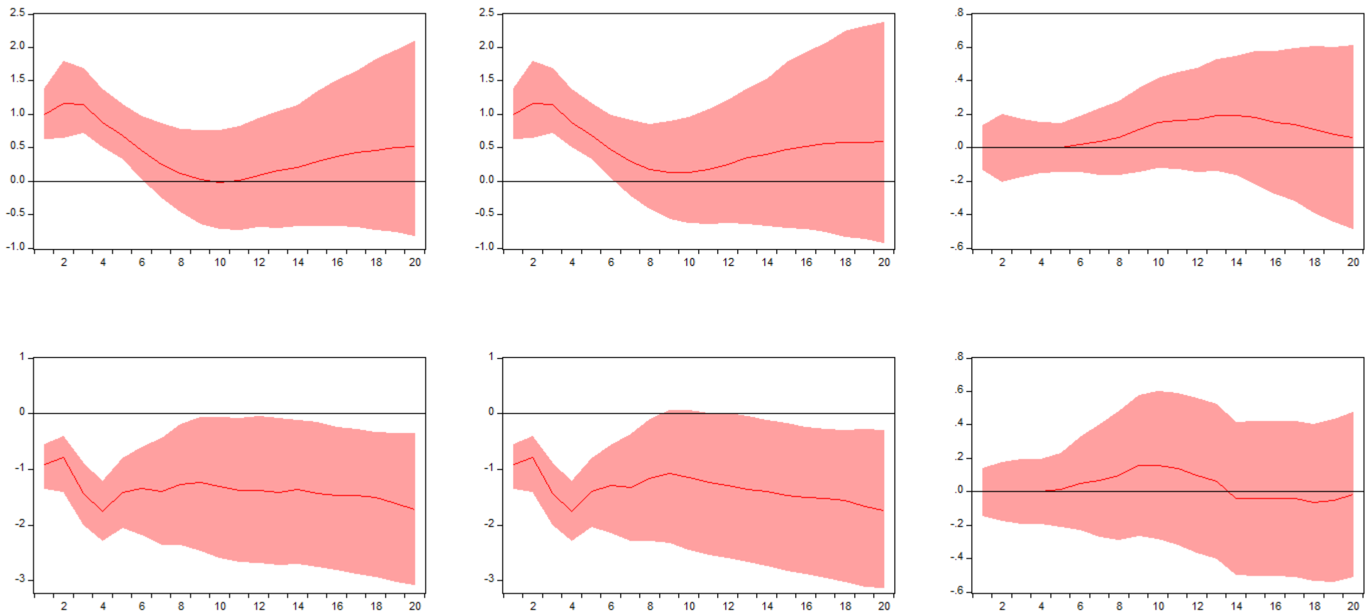
- 2012-2014: "intermediate state" not an NBER recession, would not be classified as being in the low state according to studies that use NBER recessions as the switching variable, but the switching variable is close to the threshold.[18]

For a fixed history (or points within the set of histories in the case of the "intermediate" state) , we construct the responses to an increase in government spending and taxes and to decreases in government spending and taxes (scaled to 1% of GDP). The sign restrictions are reversed for negative shocks.

Figure 7 plots the responses of output to changes in government spending and taxes in the high state, Figure 8 plots the responses in the low state, and Figure 9 plots the responses in the intermediate state. The top panels of the Figures plot the responses to government spending, the bottom panels plots the responses to tax changes. The left columns plot the responses to a positive shock, the middle panels plots the response to a negative shock (scaled by -1 for ease of comparison), and the right panel plots the difference between the scaled response to a contractionary shock and the response to an expansionary shock.

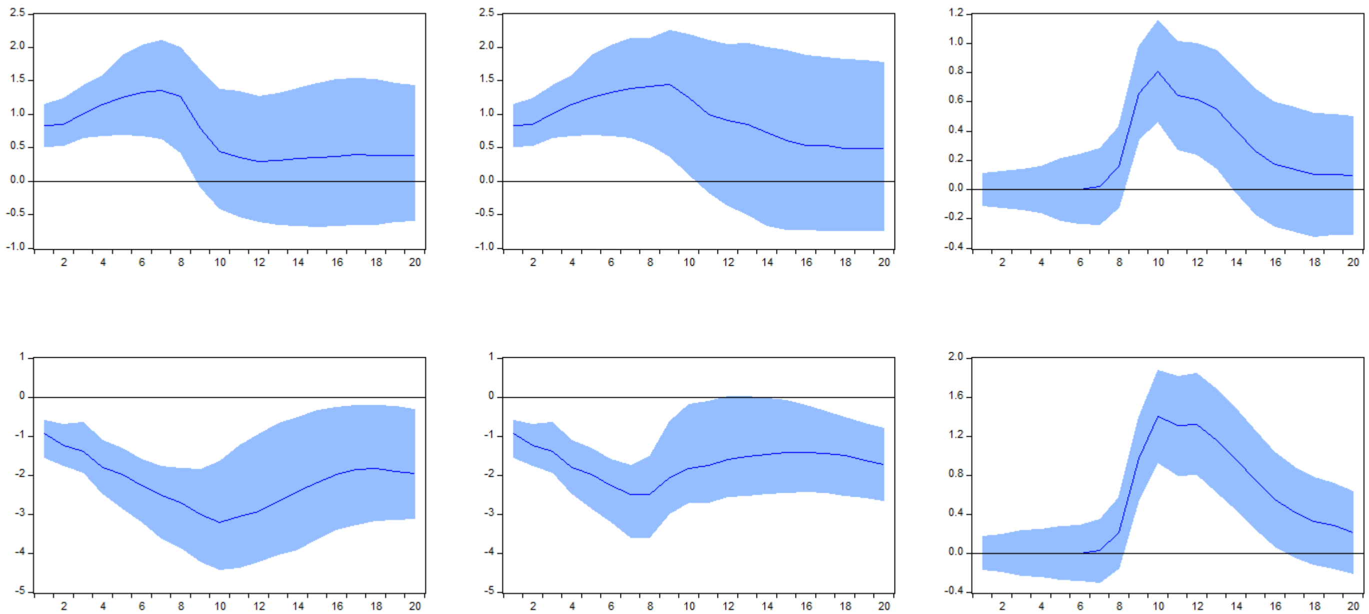
[18]It is important to note that averaging over all similar points (deep recessions, robust expansions, and sluggish recoveries) gives very similar results.

Figure 7: Sign Asymmetry in the Responses of Output High State



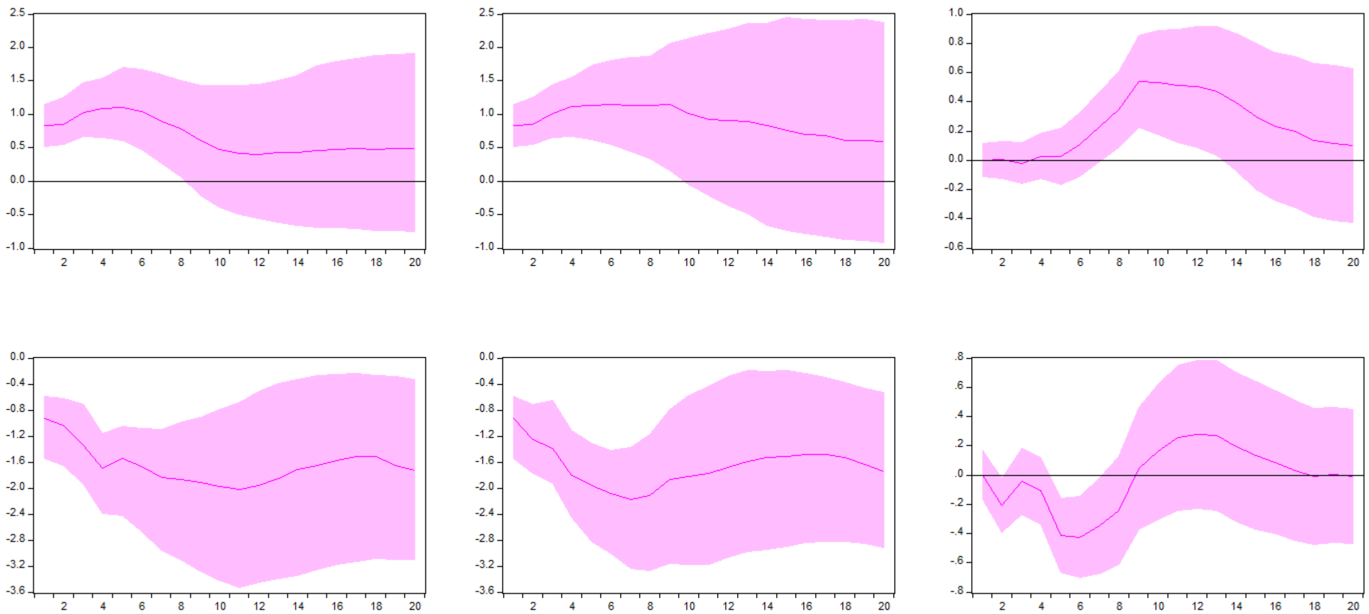
The left columns plot the responses to a positive shock, the middle panels plots the response to a negative shock (scaled by -1 for ease of comparison), and the right panel plots the difference between the scaled response to a contractionary shock and the response to an expansionary shock with 90% CIs.

Figure 8: Sign Asymmetry in the Responses of Output Low State



The left columns plot the responses to a positive shock, the middle panels plots the response to a negative shock (scaled by -1 for ease of comparison), and the right panel plots the difference between the scaled response to a contractionary shock and the response to an expansionary shock.

Figure 9: Sign Asymmetry in the Responses of Output Medium State



The left columns plot the responses to a positive shock, the middle panels plots the response to a negative shock (scaled by -1 for ease of comparison), and the right panel plots the difference between the scaled response to a contractionary shock and the response to an expansionary shock with 90% CIs.

In the high state, negative demand shocks have larger effects, on average, than positive demand shocks, but this difference is not significant. In the high state, tax cuts are more efficient at stimulating output than increases in government spending (1.7 vs 0.6 after a year), which is consistent with Mountford and Uhlig (2009). The magnitude of the peak responses to tax shocks is also in line with, for example, the responses obtained by Romer and Romer (2010). On the other hand, our results for tax increases in the high state stand in sharp contrast to the findings of Jones, Olson, and Wohar (2015), who find that tax increases do not affect output, but decreases have a strong positive effect. We find that tax increases have a strong contractionary effect on output.

The effects of negative demand shocks are much more persistent and larger than the effects of positive demand shocks in the low state. Cuts in government spending decrease output by $1.7. Tax increases decrease output by almost $3 (which is comparable in magnitude to the responses from studies that use narrative tax shocks, thus lending further credence to the identification strategy we use). The responses to tax cuts and increases in spending are smaller (peaking at approximately $2 and $1.5, respectively). Both increases in taxes and and decreases in government spending significantly decrease output (the response is different from zero at all horizons for tax increases, and for 2 years for spending cuts).

Similarly, in the intermediate state, negative spending shocks also have stronger effects than positive spending shocks. Recall that we defined the "intermediate" state as states when the economy is close to the threshold. Spending cuts decrease output by $1.3 and this decrease is quite persistent, whereas spending increases increase output, but only temporarily. Our results indicate that if the goal of fiscal policy is to stimulate the economy, government spending and tax cuts could be used in periods of slack, and tax cuts should be used in expansions.

While negative austerity measures have stronger effects than stimulus measures, the difference for government spending shocks is much smaller in expansions. This implies both that austerity measures could be self-defeating under some circumstances, and that different stimulative policies can have different effects on deficits and the debt to output ratio.

Figure 10 plots the effect of stimulative policies on the government deficit, and Figure 11 plots the effects on the debt to output ratio. Similarly, Figure 12 plots the effect of austerity policies on the government deficit, and Figure 13 plots the effects on the debt to output ratio. The left panel plots the responses to an increase in G, and the right panel plots the responses to tax cuts. In the high state, tax cuts have stimulative effects on output, and lead to increases in output that offset the increase in deficits.[19] The full set of impulse responses that includes credibility intervals is presented in Appendix C (Figures C3 through C6).

While increases in spending increase government deficits in all states, they increase output by a larger amount in the low state.[20] Therefore, increases in government spending increase the debt ratio more in the high state than in the low state (with the long-horizon responses being similar, because we consider evolving responses where the economy eventually return to the high state, which is the "normal" more common state). By contrast, decreases in taxes increase the debt ratio, but only temporarily.

Austerity measures in the low state decrease the debt ratio, but this decrease is temporary, and insignificant at all horizons. Conversely, spending cuts

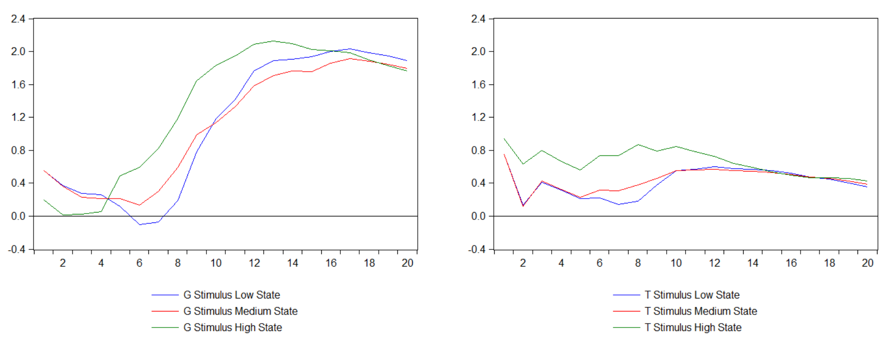
[19]The full set of responses for all variables in the VAR, and the confidence intervals for all responses are available in a supplemental appendix that is available upon request. When output increases, even though we did not impose any additional restrictions, as expected, transfer payments decrease after a short lag, and tax revenues increase. Similarly, when output falls, transfer payments increase.

[20]In addition, they also increase tax revenues more strongly and decrease transfer payments, but because these results were not statistically significantly different across states, they are relegated to a supplemental appendix that is available upon request.

implemented in the high state decrease the debt ratio significantly by 2% after 2 years. Increases in taxes reduce the debt ratio significantly only on impact, and the response is smaller than the response to spending shocks.[21]

Our results indicate that there is strong evidence of state dependence. Furthermore, in the low state, both government spending and taxes can be used as tools to stimulate the economy. In the high state and the intermediate, tax cuts lead to larger increases in output. Contractionary fiscal shocks have, on average, larger effects than expansionary fiscal shocks of the same magnitude. Austerity measures increase deficits and the debt to output ratio in the medium horizon if implemented in periods of slack due to a sharp decrease in output. Spending cuts can reduce the debt ratio without decreasing output when the economy is close to trend, but tax increases tend to retard output growth across the business cycle.

Figure 10: Effect of Stimulative Policies on Deficits



[21] The results for tax increases are consistent with Fotiou's (2016) findings, who uses a panel of OECD countries and a narrative measure of austerity measures, and finds that tax increases are self-defeating.

Figure 11: Effect of Stimulative Policies on the Debt to GDP Ratio

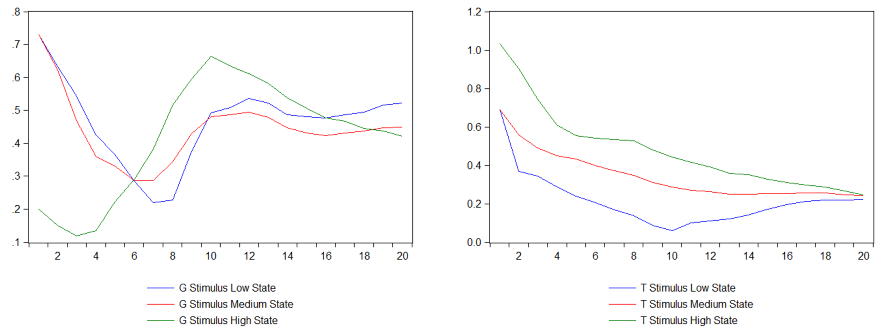


Figure 12: Effect of Austerity Policies on Deficits

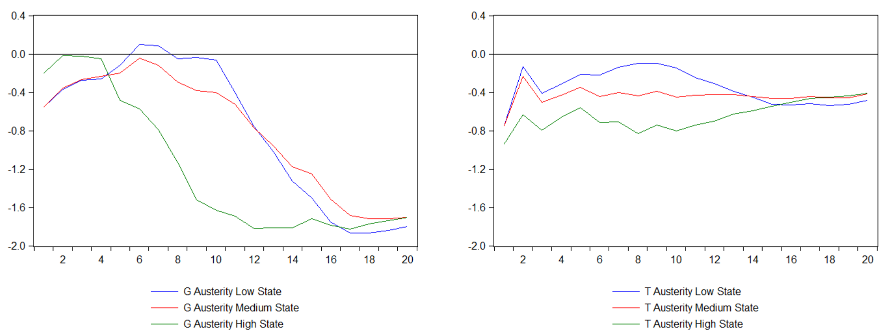
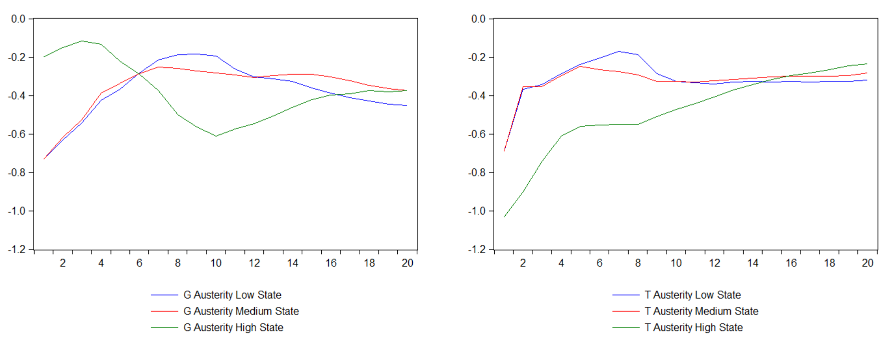


Figure 13: Effect of Austerity Policies on the Debt to GDP Ratio



5 Responses of Consumption and Investment

In this section, we disaggregate the response of output and consider the responses of consumption and investment to spending and tax shocks. The responses were constructed following the same approach used to construct the responses of output, but the responses at horizon zero of consumption and investment to fiscal shocks were left unrestricted. Figure 14 plots the fixed-state responses of consumption, Figures 15 through 17 plot the responses of consumption to positive and negative shocks in government spending and taxes, and Figures 18 through 20 plot the responses of investment to positive and negative shocks in government spending and taxes.

There is very strong evidence in favor of state dependence in the response of consumption to spending shocks, and this state dependence drives a lot of the state dependence in the response of output. In particular, consumption increases very strongly in the low state. While it also increases in the high state, the response is much smaller. Similarly, the response of consumption to tax shocks is stronger in the low state. The response pattern for spending shocks is consistent with the findings of FMP, and it is also consistent, with, for example, models that incorporate a time-varying share of rule-of-thumb consumers, or with the empirical findings of Klein (2016), where household debt and the consumption channel are the most important transmission channel in the response of output to austerity measures.

There is no evidence of state dependence in the response of investment to spending shocks, even when we restrict our attention to the HPD estimates and eliminate parameter uncertainty (top panels of Figures 18-20).[22] However, while there is no state dependence or sign asymmetry, there is evidence that in-

[22]The fixed-state responses for investment look virtually identical to the responses from the left-hand-side panels of Figures 18 through 20.

vestment responds very differently to spending and tax shocks. Investment does not respond significantly to spending increases, but it responds significantly to tax cuts. The response of output to spending shocks in the low state is therefore primarily driven by the response of consumption. By contrast, investment increases in response to tax cuts in both states, and the response of output to tax cuts is driven both by the responses of consumption and investment.

Therefore, our model indicates that if the goal of stimulative policy is to increase consumption, spending increases and tax cuts have multipliers that are not significantly different in the low state, but spending increases led to smaller medium-term increases in the debt ratio. Meanwhile, if the goal is to increase investment, tax cuts work uniformly better than spending increases, and they work uniformly better in the high state.

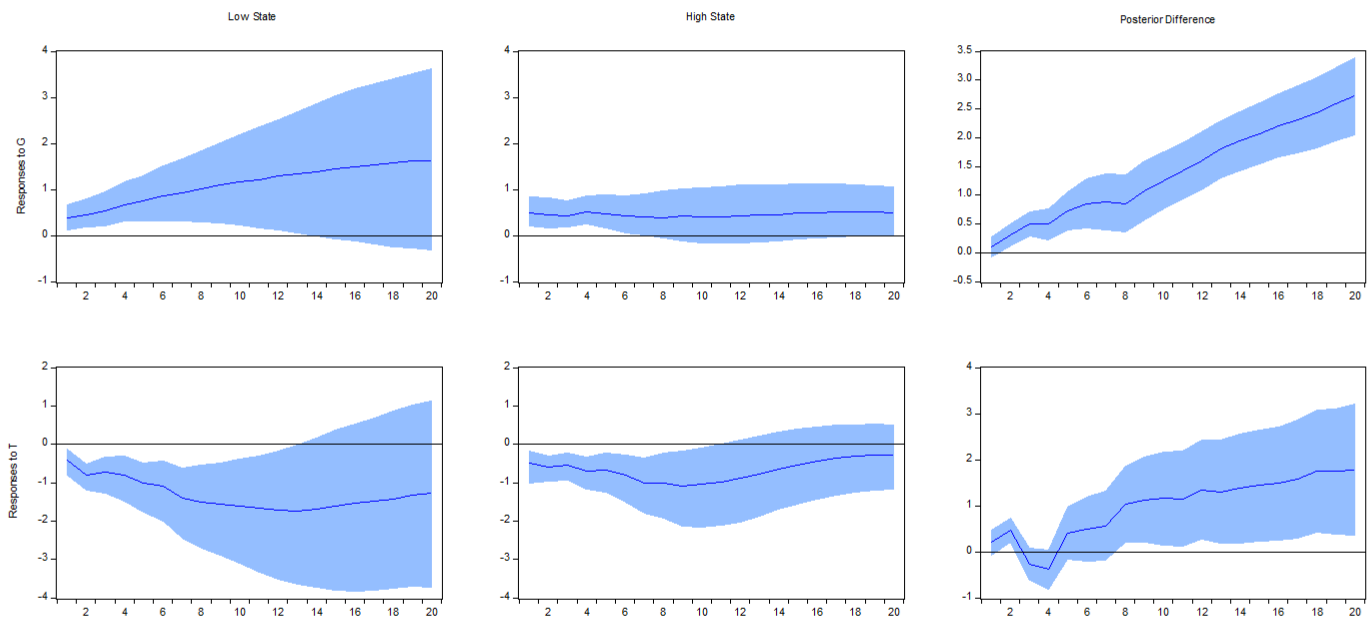
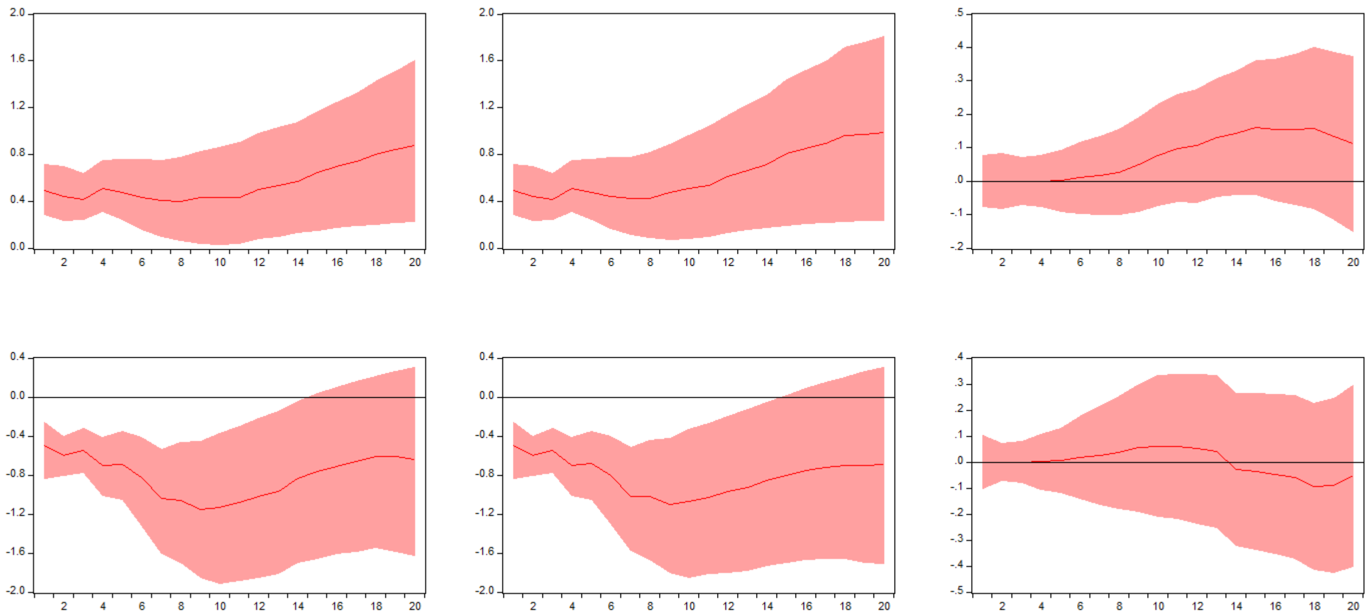


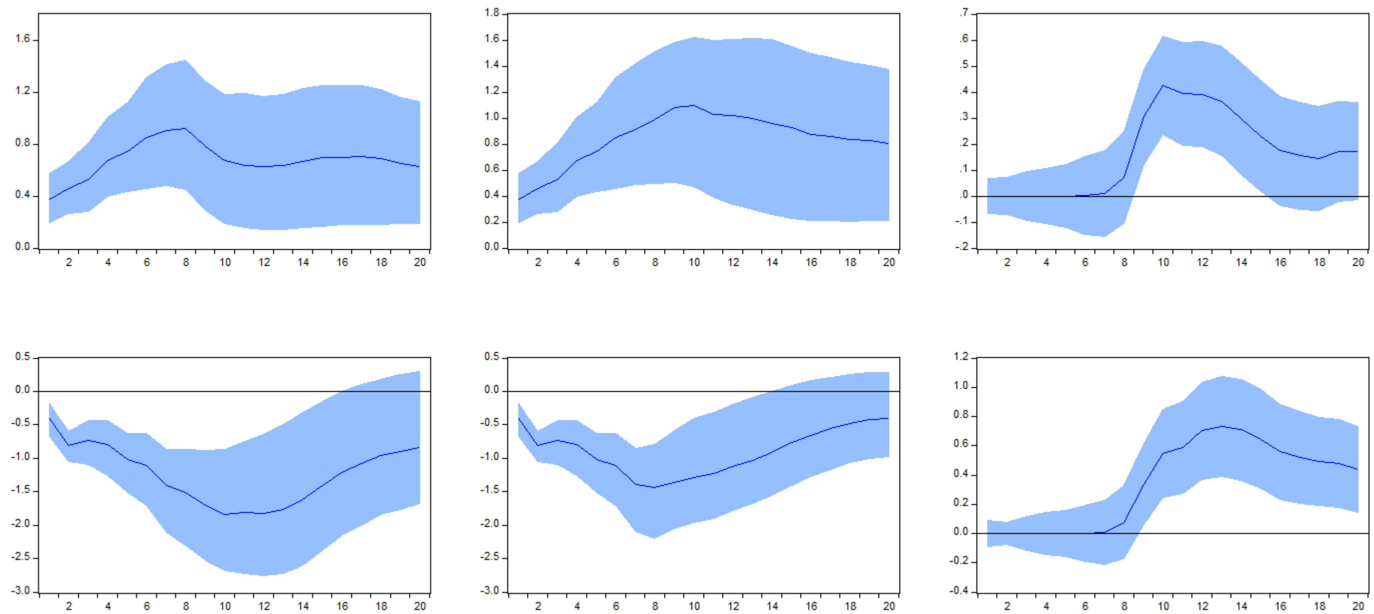
Figure 14: State-Dependence Responses of C Fixed Low (Left) and Fixed High State (M), and posterior differences (R) with 90% CIs.

Figure 15: Sign Asymmetry in the Responses of Consumption High State



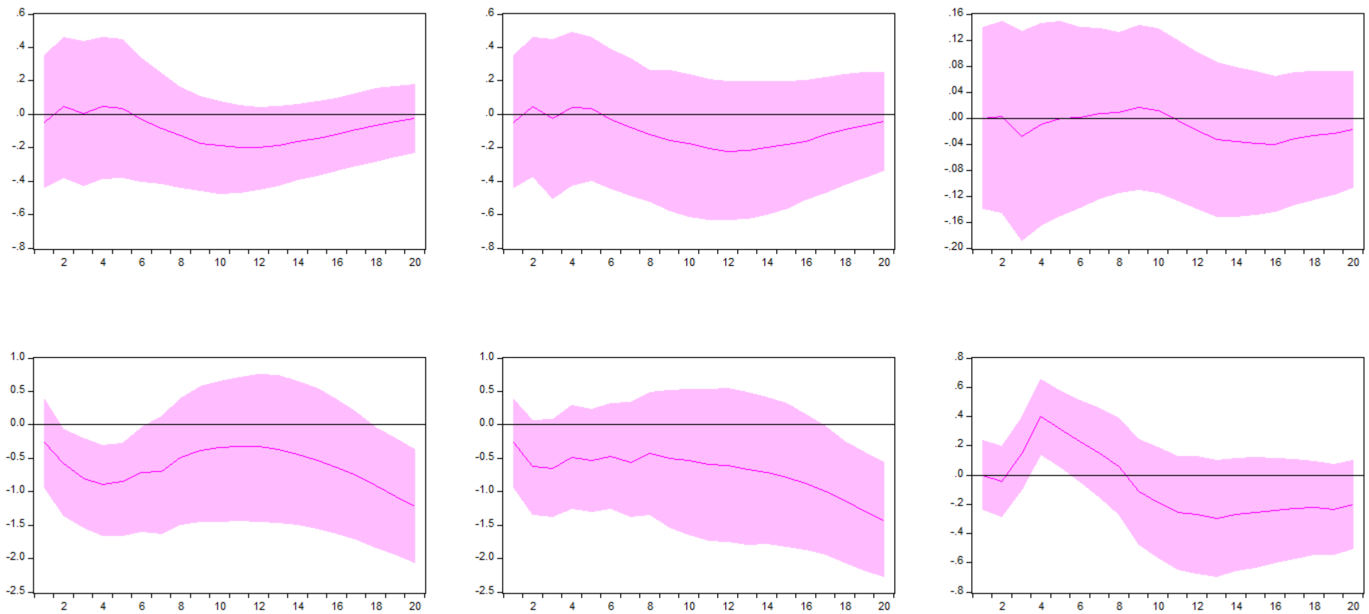
The left columns plot the responses to a positive shock, the middle panels plots the response to a negative shock (scaled by -1 for ease of comparison), and the right panel plots the difference between the scaled response to a contractionary shock and the response to an expansionary shock with 90% CIs.

Figure 16: Sign Asymmetry in the Responses of Consumption Low State



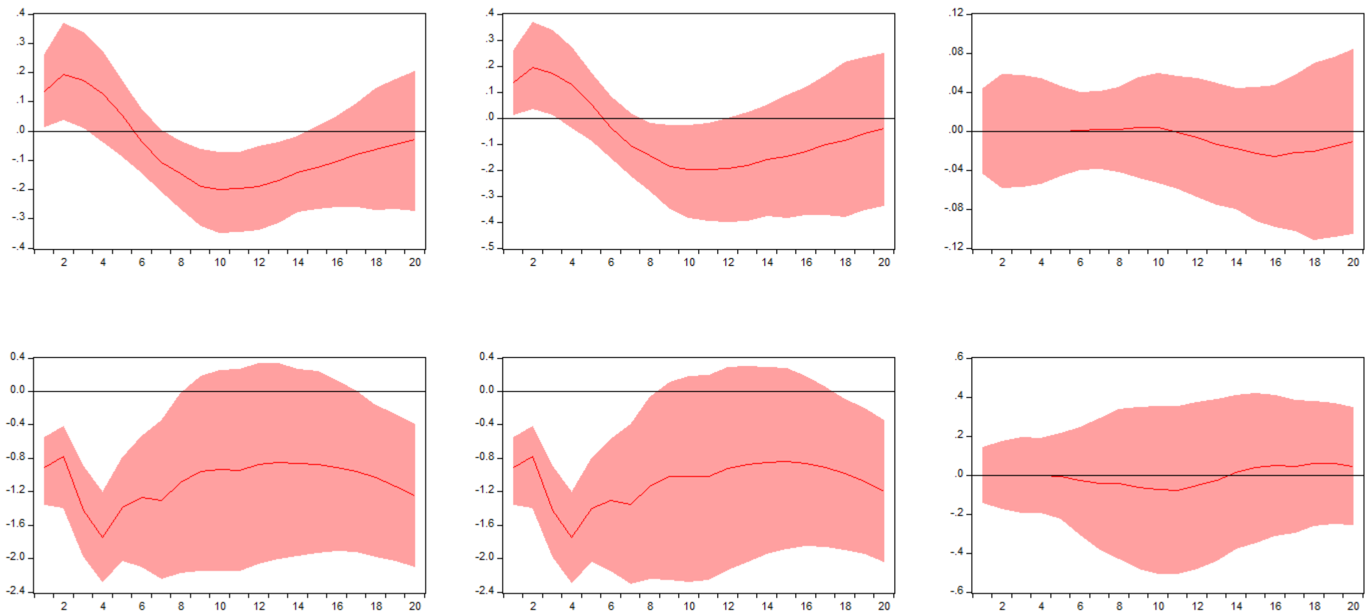
The left columns plot the responses to a positive shock, the middle panels plots the response to a negative shock (scaled by -1 for ease of comparison), and the right panel plots the difference between the scaled response to a contractionary shock and the response to an expansionary shock with 90% CIs.

Figure 17: Sign Asymmetry in the Responses of Consumption Medium State



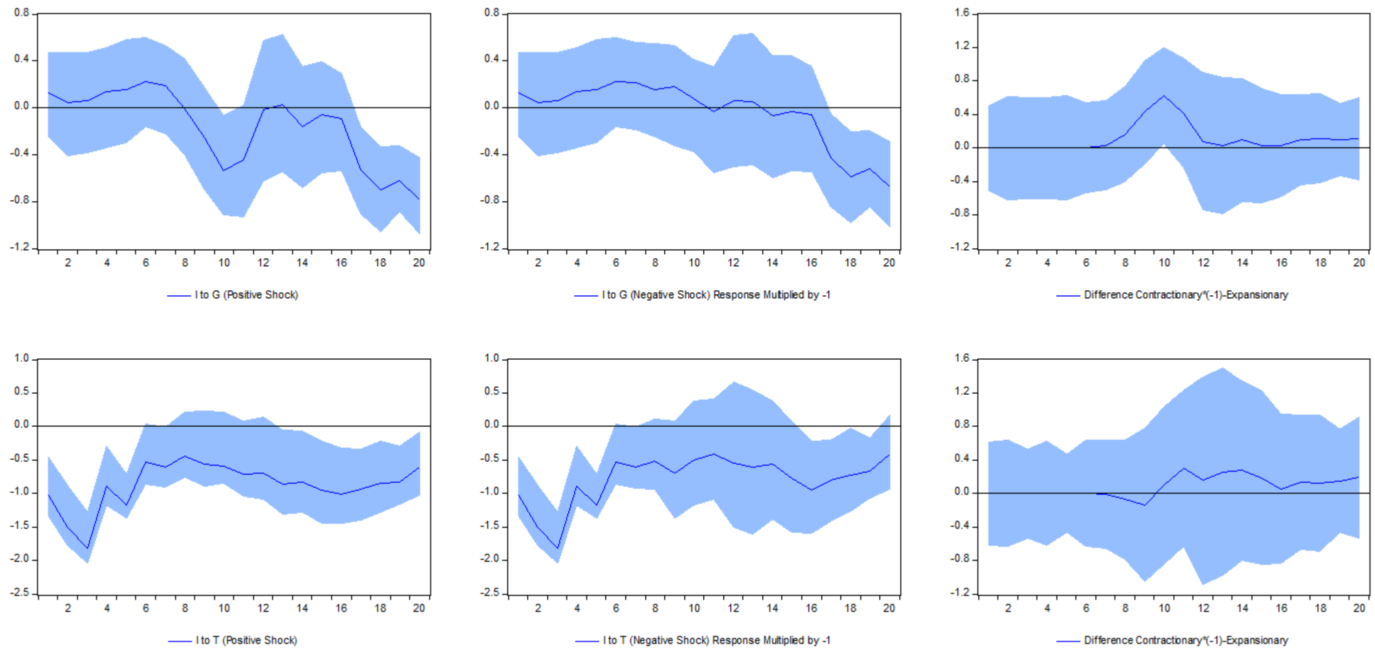
The left columns plot the responses to a positive shock, the middle panels plots the response to a negative shock (scaled by -1 for ease of comparison), and the right panel plots the difference between the scaled response to a contractionary shock and the response to an expansionary shock with 90% CIs.

Figure 18: Sign Asymmetry in the Responses of Investment High State



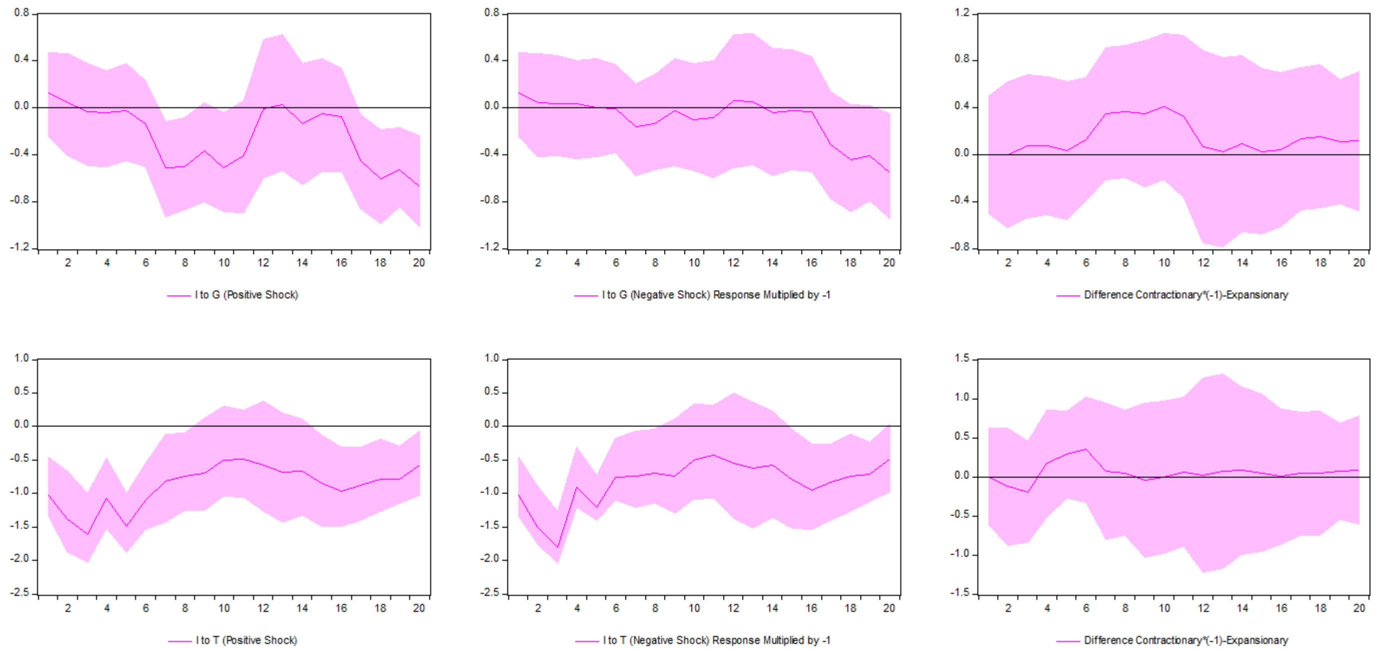
The left columns plot the responses to a positive shock, the middle panels plots the response to a negative shock (scaled by -1 for ease of comparison), and the right panel plots the difference between the scaled response to a contractionary shock and the response to an expansionary shock with 90% CIs.

Figure 19: Sign Asymmetry in the Responses of Investment Low State



The left columns plot the responses to a positive shock, the middle panels plots the response to a negative shock (scaled by -1 for ease of comparison), and the right panel plots the difference between the scaled response to a contractionary shock and the response to an expansionary shock with 90% CIs.

Figure 20: Sign Asymmetry in the Responses of Investment Medium State



The left columns plot the responses to a positive shock, the middle panels plots the response to a negative shock (scaled by -1 for ease of comparison), and the right panel plots the difference between the scaled response to a contractionary shock and the response to an expansionary shock with 90% CIs.

6 Conclusion

In this paper, we have examined important time dependencies in the responses of output to discretionary changes in government spending and taxes. We have presented strong empirical evidence in favor of nonlinear, state-dependent effects of fiscal policy. In particular, the estimates from a threshold structural vector autoregression identify different responses of the economy both to government spending and to tax shocks that depend on the degree of economic slack. Positive spending shock and tax cuts have larger effects on output in the low state than in the high state. Spending shocks have short-lived effects on output in the high state. Meanwhile, tax cuts have persistent stimulative effects in both the high and the low state.

Contractionary fiscal shocks have large negative effects on output when there is a lot of slack in the economy. In particular, dollar for dollar, a decrease in government spending decreases output by \$1.7. In the high state, tax increases have significant contractionary effects, decreasing output, consumption, and investment, while tax cuts have stimulative effects. Our results indicate that if the goal of policy is to stimulate the economy during periods of high slack, spending multipliers are higher than tax multipliers, and they work primarily through the consumption channel. In expansions, tax cuts have larger effects than spending increases. Austerity measures are self-defeating in periods of slack. If the goal of austerity measures is to reduce the debt to output ratio, our results suggest that austerity measures should be implemented in expansions through fiscal cuts, but not through tax increases.

References

[1] Arias, J. E., J. F. Rubio-Ramirez, and D. F. Waggoner (2015). "Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications." Federal Reserve Bank of Atlanta Working Paper.

[2] Auerbach, A. J. and Y. Gorodnichenko (2012). "Measuring the Output Responses to Fiscal Policy." American Economic Journal: Economic Policy 4(2), pp. 1-27.

[3] Auerbach, A. J. and Y. Gorodnichenko (2013a). "Fiscal Policy, Profits, and Investment." In: Alesina, A., and Giavazzi, F. (Eds), Fiscal Policy after the Financial Crisis, Chicago, University of Chicago Press.

[4] Auerbach, A. J. and Y. Gorodnichenko (2013b). "Output Spillovers from Fiscal Policy." American Economic Review 103(3), pp. 11-146.

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

[6] Bachmann, R. and E. Sims (2012):" Confidence and the Transmission of Government Spending Shocks." Journal of Monetary Economics 59, pp. 235-249.

[7] Barnichon, R. and C. Matthes (2015). "Stimulus versus Austerity: The Asymmetric Government Spending Multiplier." CEPR Discussion Papers 10584.

[8] 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.

[9] Baum, A., M. Poplawski-Ribeiro, and A. Weber (2012). "Fiscal Multipliers and the State of the Economy." International Monetary Fund Working Paper No. 12/286.

[10] 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), pp. 1329-1368.

[11] Caggiano, G., E. Castelnuovo, V. Colombo, and G. Nodari (2016). "Estimating Fiscal Multipliers: News From a Nonlinear World." Economic Journal, forthcoming.

[12] Campolieti, M. D. Gefang, and G. Koop (2014). "Time Variation in the Dynamics of Worker Flows: Evidence from North America and Europe." Journal of Applied Econometrics 29, 265-290.

[13] Candelon, B. and L. Lieb (2013). "Fiscal Policy in Good and Bad Times." The Journal of Economic Dynamics and Control 37(12), pp. 2679-2694.

[14] Chib, S. and I. Jeliazkov (2001). "Marginal Likelihood from the Metropolis-Hastings Output." Journal of the American Statistical Association 96, pp. 270-281.

[15] Cloyne, J. (2013). "Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom." American Economic Review 103(4), pp. 1507-1528.

[16] Favero, C. and F. Giavazzi (2012). "Measuring Tax Multipliers: The Narrative Method in Fiscal VARs." American Economic Journal: Economic Policy 4(2), pp. 69-94.

[17] Fazzari, S., J. Morley, and I. Panovska (2015). "State-dependent Effects of Fiscal Policy." Studies in Nonlinear Dynamics and Econometrics 19(3), pp. 285-315.

[18] Fotiou, A. (2016) "Nonlinearities and Fiscal Policy." Bocconi Univeristy working paper.

[19] Fry, R. and A. Pagan (2011): "Sign Restrictions in Structural Vector Autoregressions: A Critical Review." Journal of Economic Literature 49(4),pp 938-960.

[20] Ghiblavi, H., C. J. Murray, and D. H. Papell (2000): "The Structure of Unemployment." Review of Economics and Statistics 82(2), pp 309-315.

[21] Guajardo, J. D. Leigh and A. Pescatori (2014). "Expansionary Austerity? International Evidence." Journal of the European Economic Association 12 (4), pp 949-968.

[22] Inoue, A. and L. Kilian (2013). "Inference on Impulse Response Functions in Structural VAR Models." Journal of Econometrics 31(1), pp. 78-93.

[23] Hansen, B. (1997). "Inference in TAR Models." Studies in Nonlinear Dynamics and Econometrics 2(1), pp. 1-14.

[24] Jones, P. M., E. Olson, and M. E. Wohar. (2015) "Asymmetric Tax Multipliers." Journal of Macroeconomics 43, pp. 38-48.

[25] Jorda, O. (2005). "Estimation and Inference of Impulse Responses by Local Projections." American Economic Review 95(1), 161-182.

[26] Jorda, O. and A. M. Taylor (2016). "The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy." The Economic Journal 126, pp 219-255.

[27] Klein, M. (2016). "Austerity and Private Debt." Working Paper.

[28] Koop, G. and S. Potter. (1996) "Impulse-Response Analysis in Nonlinear Multivariate Models.", Journal of Econometrics, 74, 119-147.

[29] Koop, G. and S. Potter. (2001) "Are Apparent Findings of Nonlinearity Due to Structural Instability in Economic Time Series?" Econometrics Journal 4, 37-55.

[30] Lo, M. C. and J. Piger (2005). "Is the Response of Output to Monetary Policy Asymmetric? Evidence from a Regime-Switching Coefficients Model." Journal of Money, Credit, and Banking 37(5), pp. 865-886.

[31] Morita, H. (2015). "State-dependent Effects of Fiscal Policy in Japan: Do Rule-of-thumb Households Increase the Effects of Fiscal Policy?" Journal of Macroeconomics 43, pp. 49-61.

[32] Morley, J. and J. Piger (2012). "The Asymmetric Business Cycle." The Review of Economics and Statistics 91(1), pp. 208-221.

[33] Morley, J. and I. Panovska (2017). "Is Business Cycle Asymmetry Intrinsic in Industrialized Economies?" Working Paper.

[34] Mountford, A. and H. Uhlig (2009): "What are the Effects of Fiscal Policy Shocks?" Journal of Applied Econometrics 24(6), pp. 960-992.

[35] 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), pp. 583-601.

[36] Owyang, M., 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), pp. 129-134.

[37] Perron, P. and T. Wada. (2009) “Let’s Take a Break: Trends and Cycles in U.S. Real GDP.” Journal of Monetary Economics 56, 749-765.

[38] Perron, P. and T. Wada. (2016) “Measuring Business Cycles with Structural Breaks and Outliers: Applications to International Data.” Research in Economics 70(2), pp. 281-303.

[39] Ramey, V. A. (2011). “Identifying Government Spending Shocks: It’s All in the Timing.” Quarterly Journal of Economics 126(1), pp. 1-50.

[40] Ramey, V. A. and S. Zubairy (2015). “Are Government Spending Multipliers State Dependent? Evidence from Canadian Historic Data.” Texas A&M University Working Paper.

[41] Ramey, V. A. and S. Zubairy (2016). “Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data.” (forthcoming) Journal of Political Economy.

[42] Romer, C. and D. Romer (2010). “The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks.” American Economic Review 100(3), pp. 763-801.

[43] Tenreyro, S. and G. Thwaites (2016). “Pushing on a String: US Monetary Policy is Less Powerful in Recessions.” American Economic Journal: Macroeconomics 8(4):43-74.

[44] Terasvirta, T. and Y. Yang (2016). “Linearity and Misspecification Tests for Vector Smooth Transition Regression Models.” Macroeconomic Dynamics (forthcoming).

[45] Wu, J. C. and F. D. Xia (2016) “Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound.” Journal of Money, Credit, and Banking 48(2-3): 253-291.

Appendix A: Bayesian Estimation and Impulse Responses

For the small linear VAR model, we assume that the prior for the conditional mean parameters and the autoregressive lag polynomial parameters is multivariate normal with mean zero and variance $100 * I_n$ and the prior for the variance matrix Σ is an inverse Wishart distribution with mean I_4 and scale parameter 25. For the large linear VAR model, the priors for the variance is $100 * I_n$ and the prior for the variance matrix Σ is an inverse Wishart distribution with mean I_9 and scale parameter 25. Because these priors are conjugate, the posterior for the VAR parameters is normal and the posterior for the variance-covariance matrix Σ is an inverse Wishart distribution. Let $\Phi_{lin}^{(mh)}$ and $\Sigma_{lin}^{(mh)}$ denote the mh^{th} draw from the posterior distributions. In the small model, the shocks are identified using a Cholesky decomposition.

In the large model, this draw from the posterior is a draw from the unrestricted posterior that does not take into account any sign restrictions that are imposed. To ensure that the sign restrictions do not artificially bias the impulse responses and to ensure that we are sampling from the correct posterior distribution, we follow the algorithm from Arias et al. (2015). For a draw $\Phi_{lin}^{(mh)}$ and $\Sigma_{lin}^{(mh)}$, we generate an orthonormal matrix Q by using the QR decomposition of the matrix $X'X$ where X is a (9×1) draw from a standard normal distribution. The draw is kept if the impulse response that use $Qchol(\Sigma)$ as the impact response satisfy the sign restrictions, and the sampler moves on to $mh + 1$. If the sign restrictions do not hold, the sampler immediately moves to the $mh+1$ iteration, and the impulse responses from the mh^{th} draw are discarded. The acceptance rate for the linear model was 12%.

Let Φ denote $[vec(\Phi_0^1)'\ vec(\Phi_1^1)'\ vec(\Phi_0^2)'\ vec(\Phi_1^2)']'$. For the nonlinear model, we assume that the prior for the autoregressive and conditional mean parameters is multivariate normal with mean zero and variance $100 * I_{n_1}$ and the prior for the variance matrix Σ is an inverse Wishart distribution with mean I_4 and scale parameter 25. The prior for the threshold parameter c is uniform and covers the middle 80% of the observations for q_{t-d}. Conditional on c^{mh}, the model is linear in Φ and Σ. The posterior for Φ_i^j $(i = 0, 1, j = 1, 2)$ is Gaussian, and the posterior for Σ^{mh} is an inverse Wishart, and those parameters can be sampled directly. The posterior distribution for c is unknown, but can be sampled using a Metropolis-Hastings step. We use a student-t distribution with mean c^{mh-1} and variance equal to $std(q_{t-d})$ as the proposal density.

Conditional on c^{mh}, Φ^{mh}, and Σ^{mh}, we generate an orthonormal matrix Q by using the QR decomposition of the matrix $X'X$ where X is a (9×1) draw from a standard normal distribution. We then compute the generalized impulse responses as follows:

1. Check if the linear impulse responses for the fixed low state and for the fixed high state that use the orthogonalization $Qchol(\Sigma)$ satisfy the sign restrictions. If yes, move to step 2. If no, move to $mh + 1$.

2. Pick a history Ψ_{t-1}. This is the actual value of the lagged endogenous variables at time t.

   (a) Draw a sequence of forecast errors ϵ_{t+k} from $N(0, \Sigma)$ for $k = 0, 1, ..., 20$.

   (b) Using Ψ_{t-1} and ϵ_{t+k}, simulate the evolution of Y_{t+k} over 21 periods. Denote the resulting path $Y_{t+k}(\epsilon_{t+k}, \Psi_{t-1})$

   (c) Using $Qchol(\Sigma)$ to orthogonalize the shocks at time zero, construct the implied vector of forecast errors. At time 0, $\epsilon_t^{shock} = Q * chol(\Sigma)$, and $\epsilon_{t+k}^{shock} = \epsilon_{t+k}$ for $k \geq 1$. Denote the simulated evolution of Y_{t+k} as $Y_{t+k}(\epsilon_{t+k}^{shock}, \Psi_{t-1})$ for $k = 0, ..., 21$.

(d) Construct a draw of a sequence of impulse responses as $Y_{t+k}(\epsilon_{t+k}^{shock}, \Psi_{t-1}) - Y_{t+k}(\epsilon_{t+k}, \Psi_{t-1})$ for $k = 0, 1, .., 20$.

(e) Repeat steps 2.a through 2.d $B = 500$ times to obtain the average responses of Y_t conditional on c, Φ, Σ, Q.

3. To obtain the average response for a subset of histories, repeat step 2 for each history and report the distribution averaged across histories.

4. Repeat steps 1 through 2 (or through 3, if averaging over histories) for each draw of the sampler.

The acceptance rate for the nonlinear model was 6%. We used 200,000 MH iterations.

Appendix B: Supplemental Tables

Table B1: Model selection: Linear vs Nonlinear Models Large Model: Consumption and Investment

(a) Consumption

	Measure of Slack in the TVAR			
Switching Variable	cap_{1973b}		$MAOG$	
Linear model (none)	-2662.16 -2261.30 -670.85		-2395.05 -2393.30 -617.81	
cap_{1973b}	-2239.40 -2236.26 -442.91	-0.64 $(-0.94, 0.43)$	-1978.22 -1976.20 -348.53	-0.56 $(-0.94, -0.24)$
$MAOG$	-1932.24 -1930.87 -333.80	-0.55 $(-0.74, -0.36)$	-1932.24 -1930.87 -333.80	-0.54 $(-0.74, -0.36)$

(b) Investment

	Measure of Slack in the TVAR			
Switching Variable	cap_{1973b}		$MAOG$	
Linear model (none)	-2981.60 -2975.40 -957.80		-2718.36 -2715.42 -873.77	
cap_{1973b}	-2544.41 -2540.10 -548.77	-0.42 $(-2.56, 0.45)$	-2300.75 -2289.73 -625.24	-2.96 $(-3.04, 0.46)$
$MAOG$	-2457.98 -2450.14 -454.50	-0.69 $(-0.81, -0.52)$	-2221.51 -2214.82 -400.64	-0.71 $(-0.83, -0.54)$

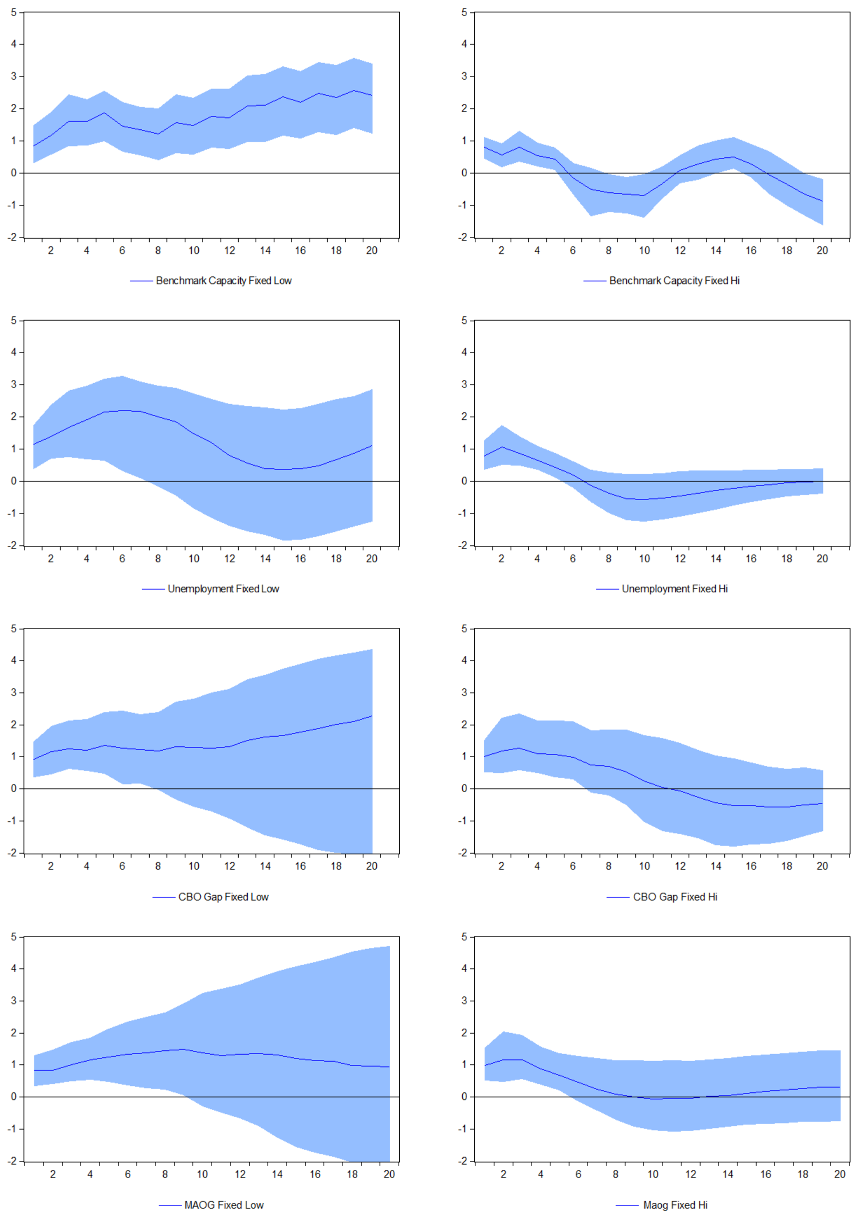
Each cell reports the likelihood obtained using a frequentist grid search procedure, the expected posterior likelihood obtained Bayesian estimation, and the marginal likelihood (T, M, B).
The second entry is the threshold estimate, including 90% credibility intervals, obtained from the posterior Bayesian distribution

Table B2: Model selection: Linear vs Nonlinear Models Large Model No Zero Lower Bound

	Switching Variable		
	Linear Model (None)	$MAOG$	
$Elik_{HPD}$ and Threshold Estimate	-2065.55	-1486.86	-0.54 $(-0.82, -0.36)$

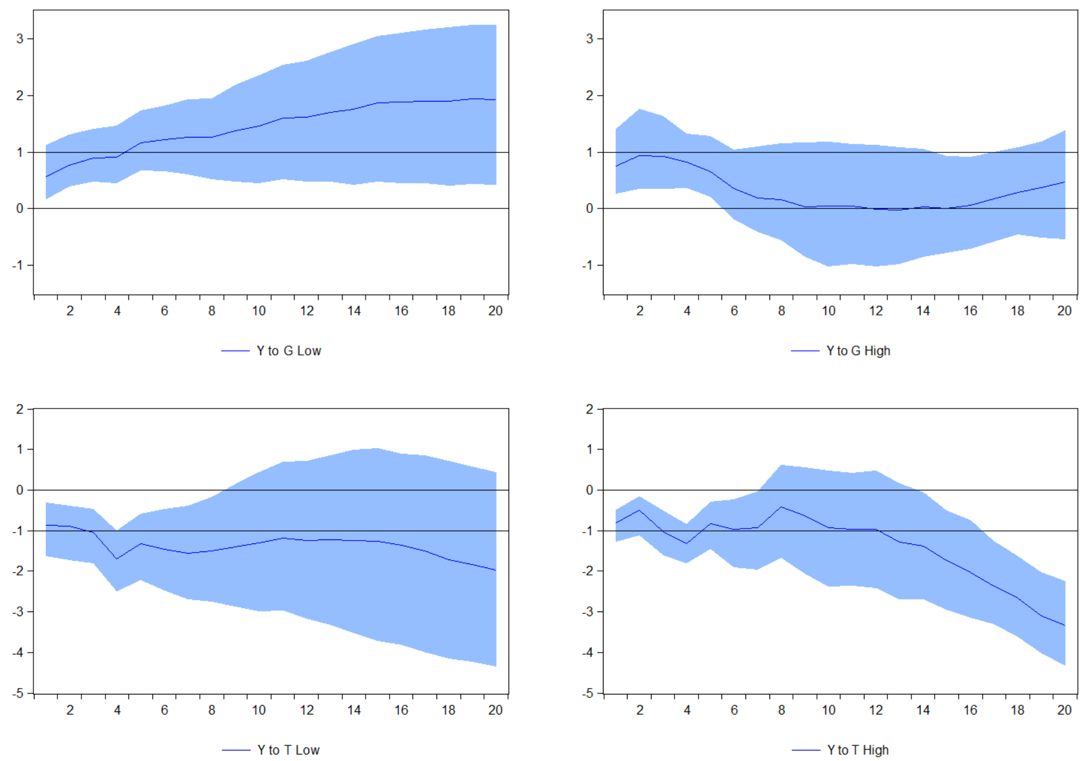
Appendix C: Supplemental Figures

Figure C1: State-Dependence Large Model 1967Q1-2015Q4



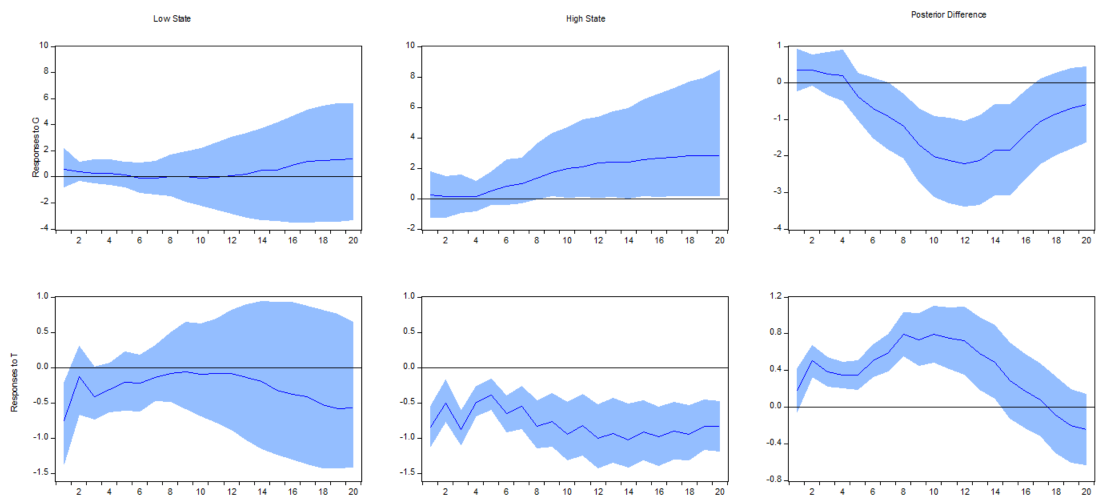
The responses of output to government spending in the low state (L) and high state (R) when using different measures of slack with 90% CIs. Measures of slack (top to bottom) capacity with an imposed break in 1973, capacity with an estimated LR break in 2001, capacity with 2 structural breaks (2001, 1973), the unemployment rate, the CBO output gap, Morley and Panovska's MAOG.

Figure C2: State-Dependence Large Model: Excluding the ZLB Period 1967Q1-2006Q4



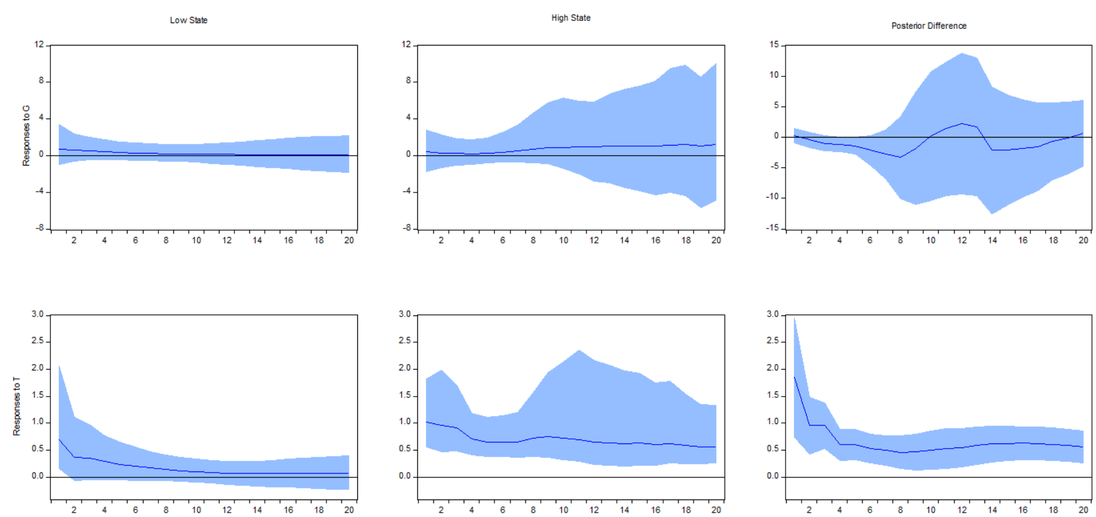
Responses of output to G (top row) and T (bottom row). MAOG as a measure of slack and switching variable. Fixed low (L) and high state (R) with 90% CIs. Horizontal lines are at 1 and -1.

Figure C3: State-dependence in the Response of Deficits



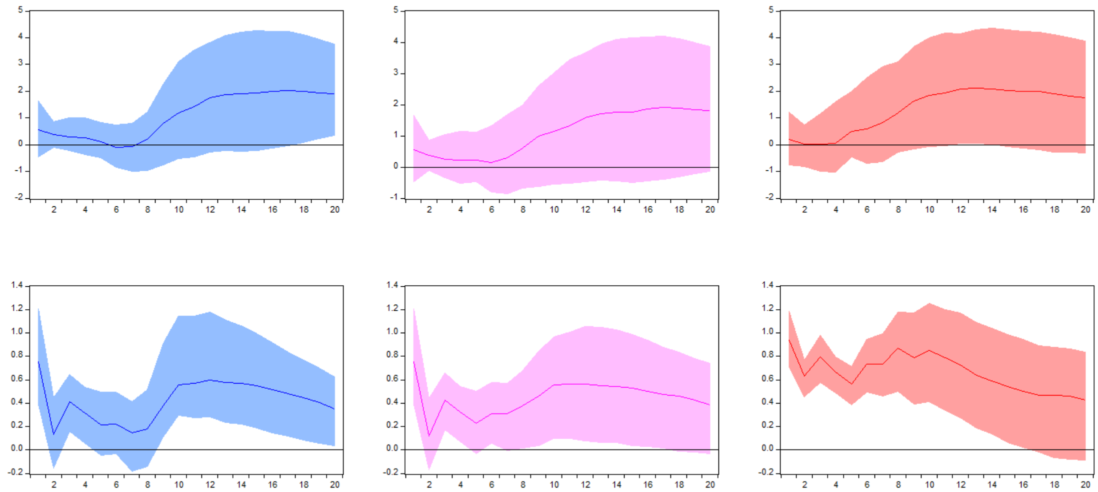
State-Dependence Responses of Deficits to G (top) and T (bottom). Fixed Low (Left) and Fixed High State (M), and posterior differences (R) with 90% CIs

Figure C4: State-dependence in the Response of the Debt Ratio



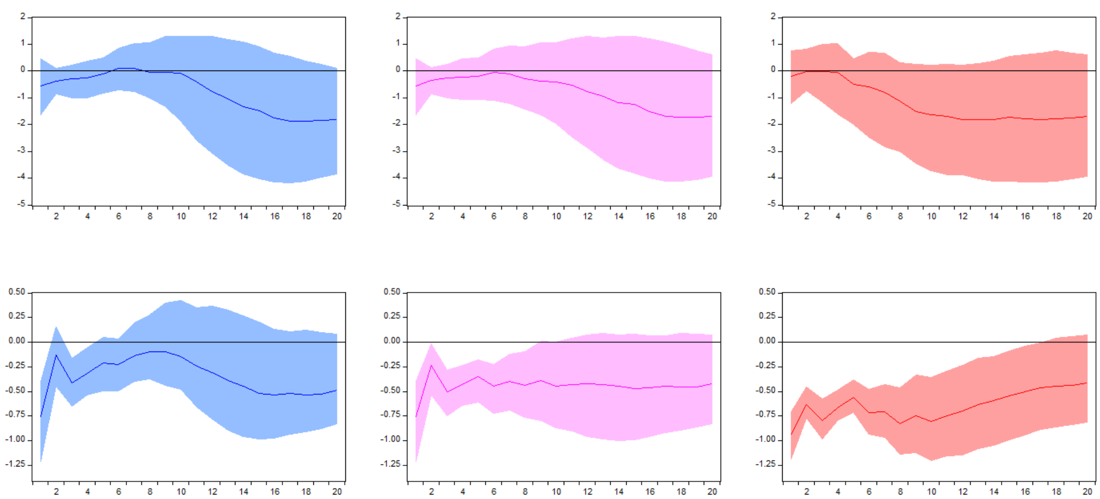
State-Dependence Responses of D/Y to G (top) and T (bottom). Fixed Low (Left) and Fixed High State (M), and posterior differences (R) with 90% CIs

Figure C5: Response of Deficits to Stimulus Measures



Responses of Deficits to increases in G (top) and decreases in T (bottom).
Evolving low (Left), medium (M), and high state (R), 90% CIs

Figure C6: Response of Deficits to Austerity Measures



Responses of Deficits to decreases in G (top) and increases in T (bottom).
Evolving low (Left), medium (M), and high state (R), 90% CIs