Business School / School of Economics

UNSW Business School Working Paper

UNSW Business School Research Paper No. 2017 ECON 04
First version February 2017
UNSW Business School Research Paper No. 2017 ECON 04A
This version September 2017

When Is Discretionary Fiscal Policy Effective?

Steven M. Fazzari
James Morley
Irina B. Panovska

When Is Discretionary Fiscal Policy Effective?

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 the effects of discretionary fiscal policy using a medium-scale nonlinear vector autoregressive model with government spending and tax shocks identified via sign restrictions. Our estimates suggest a strong state dependence in the relationship between fiscal policy and the macroeconomy, with the nonlinearity closely related to the phase of the business cycle. This state dependence has important implications for the timing of stimulus or austerity measures. Tax cuts and government spending increases have similarly large stimulative effects when there is excessive slack in the economy around recessions and in early recoveries, but they are much less effective, especially in the case of government spending increases, as the expansion becomes more established and the economy is closer to potential. Tax increases and government spending cuts are contractionary and largely self-defeating in reducing the debt-to-GDP ratio during periods of excess slack. The effectiveness of discretionary government 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 tax changes are state dependent, but investment plays the larger quantitative role in terms of their effectiveness.

JEL Classification: E32, E62, C32. Key words: Government spending, austerity, nonlinear dynamics, Bayesian, sign restrictions, vector autoregression

*Corresponding author. Email irp213@lehigh.edu. This draft: September 6th, 2017. We thank the conference and seminar participants at the 2017 Conference of the Society for Computational Economics and Finance, 2017 NBER-NSF Seminar on Bayesian Inference in Economics and Statistics, 2016 Symposium of the Society of Nonlinear Dynamics and Econometrics, the Sydney Macroeconomics Readings Group, the Australian Treasury, the Bank for International Settlements, the Federal Reserve Bank of Philadelphia, Monash University, and the Reserve Bank of New Zealand. 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 the macroeconomy; that is, the effects of a fiscal shock may depend on economic conditions at the time when the policy is undertaken. Because "discretionary" implies choice, including choice about timing, state dependence opens up important questions about *when* to implement different policies, questions that would simply be irrelevant under linearity. In this paper, we focus on the following two questions about the timing of policy that become important given state dependence:

- When do discretionary government spending increases or tax cuts provide effective stimulus to the macroeconomy?
- When are discretionary government spending cuts or tax increases effective in reducing the debt-to-GDP ratio?

We make two contributions to the growing empirical literature on state-dependent effects of fiscal policy. First, we examine the nature of possible state dependence by considering a medium-scale threshold vector autoregression (TVAR) with fiscal shocks identified via sign restrictions and by comparing different possible threshold variables. Using U.S. data, we find strong support for nonlinearity in terms of recurrent changes in the relationship between fiscal policy and the macroeconomy that corresponds closely to the onset and initial recovery from deep recessions, as captured by the level of a particular measure of economic slack that displays strong asymmetry across the business cycle. However, our findings are highly robust to other measures of slack. Second, using generalized impulse response analysis for different initial conditions, we answer the questions raised above about how the timing of discretionary policy changes its effectiveness.

The measure of economic slack that we find most closely relates to the nonlinear relationship between fiscal policy and aggregate output is the model-averaged output gap (MAOG) developed by Morley and Panovska (2017) and based on earlier research by Morley and Piger (2012).[1] In contrast to more traditional measures of slack such as the CBO output gap, the MAOG turns out to display much larger negative movements during recessions than positive movements during expansions. For the TVAR model with this threshold variable, we find very different dynamic cross correlations between fiscal policy and aggregate output during periods of excessive economic slack than when the economy is close to or above potential. Notably, various Bayesian model selection criteria rank 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, especially when accounting for structural breaks in other measures of slack to ensure estimates for the threshold model are not distorted by possible permanent level shifts in the threshold variable, we find generally robust results in terms of the timing and implications of the nonlinearity for the other measures of slack. 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 the empirical literature following Auerbach and Gorodnichenko (2012, 2013a,b) about 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 abrupt transitions between regimes that are tied to the level of slack rather than the direction of movements in economic activity or in fiscal policy. We also directly explore whether state-dependence leads to different dollar-for-dollar effects of austerity and stimulus measures.

For our impulse response analysis, we find that tax cuts and government spending increases have similarly large stimulative effects during periods of excessive slack, but they

[1] The MAOG approach addresses uncertainty about the appropriate forecasting model for aggregate output by averaging implied estimates of the output gap from forecast-based trend-cycle decompositions for a large set of similarly-fitting reduced-form time series models. The model set includes linear and Markov-switching autoregressive and unobserved components models. Trend-cycle decomposition is based on the Beveridge and Nelson (1982) decomposition or the Kalman filter for the linear models and the Morley and Piger (2008) or Kim (1994) filter generalizations to Markov-switching processes for the nonlinear models.

are much less effective, especially in the case of government spending increases, when the economy is close to or above potential. Measures designed to reduce the debt-to-GDP ratio (i.e., tax increases and government spending cuts) are most contractionary and largely self-defeating in periods of excessive slack. The effectiveness of discretionary government 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 the larger quantitative role in the overall effectiveness of tax changes.

Our analysis builds on and merges different strands of the large empirical literature on fiscal policy multipliers. 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, Poplawski-Ribeiro, and Weber (2012), Cagianno, Castelnuovo, Colombo, and Nodari (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 medium-scale TVAR model has more fiscal variables and uses sign restrictions for identification, we are able to address potential problems in identifying discretionary government spending shocks separately from endogenous responses to economic conditions, which helps to reconcile conflicting results in the previous literature.

Including a larger number of variables and our 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 considered by, for example, 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, while tax increases have no substantial effect on US output, but they have contractionary effects on UK output. Barnichon and Matthes (2017) apply a small-scale fiscal vector autoregressive model and find that government spending cuts have larger effects than increases, with the results driven primarily by very strong negative responses of output to government spending decreases implemented during recessions. Guajardo, Leigh, and Pescatori (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, Leigh, and Pescatori (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 all of these and other possible asymmetries.

The rest of our paper is organized as follows. Section 2 presents the empirical model used in our analysis. Section 3 examines the existence and nature of nonlinearity in the relationship between government spending and aggregate output, as well as considering its robustness to using different measures of slack. Section 4 reports generalized impulse response analysis to investigate when discretionary changes in government spending or taxes are effective and explore the roles of aggregate consumption and investment in driving the possible state-dependent effects of fiscal policy on aggregate output. Section 5 concludes.

2 A Medium-Scale TVAR Model

To investigate the evidence for nonlinearity, we consider a threshold vector autoregressive (TVAR) model and different possible threshold variables. Let Y_t denote the vector containing the endogenous variables in the TVAR model. The TVAR model splits the stochastic process for Y_t into two different regimes. Within each regime, the process for Y_t is linear, but Y_t can evolve endogenously between regimes. Let q_{t-d} denote the threshold 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 TVAR change at time t. Defining 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

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

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. In our benchmark specification of the TVAR, the vector Y_t includes nine variables: log real federal consumption and investment spending, log real federal transfer payments to persons, log real federal interest payments on debt, log real transfer taxes, log of other tax revenues in real terms, log real GDP, a measure of slack, interest rates (measured using the Federal Funds Rate and the Wu-Xia, 2016, shadow rate during the zero-lower bound period), and inflation (calculated using the GDP deflator). The sample period for the benchmark model is 1967Q1-2015Q4. All fiscal variables are converted to real terms using the GDP deflator, and all nominal series were obtained from NIPA-BEA.

By focusing on federal variables only, we are able to trace out the impact on public debt, a variable of obvious interest in debates about fiscal policy. 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.[2] Most fiscal stimulus or austerity measures that involve discretionary changes in government spending are usually implemented by changes in government consumption and/or investment. Nonetheless, transfer and interest payments play a significant role in determining the government fiscal stance and have sizable effects on the debt-to-output ratio. Furthermore, government transfer payments to persons are strongly affected by the state of the business cycle and respond to movements in output. Changes in transfer payments are occasionally used as a fiscal policy tool (for example, the unemployment benefit extensions during the Great Recession), but most movements in transfer payments are likely to be endogenous.[3] Similarly, government interest payments are determined by the historical path of government spending and taxes.

We also split the tax series into two sub-components. The first one corresponds to transfer taxes, which depend on the state of the business cycle and are rarely used as a discretionary fiscal policy tool. The second component corresponds to federal tax receipts net of transfer taxes (the federal equivalent of Blanchard and Perotti's, 2002, tax series). Meanwhile, 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 estimate the parameters Φ_i^j, the threshold c, the delay lag d, and the number of lags included in the TVAR using Bayesian methods (technical details are provided in Appendix

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

[3]When we perform a variance decomposition for our benchmark model, we find that the bulk of the variation in transfer payments is explained by business cycle shocks (61% on impact, 91% after 8 quarters).

A). A Bayesian approach to inference has two advantages in highly parametrized models such as the TVAR model. 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 TVAR models are highly parametrized, conventional tests can be severely underpowered. Using a Bayesian approach allows us to circumvent this problem by accommodating informative but reasonable priors on parameters.[4] We can then directly compare the linear to the nonlinear model with marginal likelihoods. The marginal likelihoods are calculated using Chib and Jeliazkov's (2001) algorithm and we compare models based on implied Bayes factors. In addition, motivated by the concerns described by Campolieti, Gefang, and Koop (2014), we also report the expected posterior likelihoods and the highest posterior density ordinates for all models, which lead to very similar inference as the Bayes factors. Second, the impulse responses for the endogenously evolving system have nonstandard distributions that are highly non-Gaussian and may depend on the history and the size or sign of the shocks, even when the true values of parameters are known. The Bayesian sampler produces the entire posterior distribution for c, Φ_i^j and Σ conditional on the data and we are able to directly account for all sources of dispersion in the posterior distribution of the parameters by simulating the impulse responses for each iteration of the Bayesian sampler.

The crucial empirical questions we consider with our model are whether the effects of government spending really do differ across regimes defined by economic slack and whether, conditional on any state dependence, there is any evidence that an austerity measure will have effects that are significantly different from a mirrored effect of stimulus of the same size. 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. However, the nature and degree of this asymmetry can be evaluated only 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

[4]Note that our priors, while informative, are symmetric across regimes, so findings in favor of nonlinearity are not driven by the priors. Details are given in Appendix A.

strands of the fiscal literature have taken different approaches. The most popular are the timing approach, the narrative approach, and the sign restriction approach. The timing approach is used by, for example, Blanchard and Perotti (2002), Auerbach and Gorodnichenko (2012), FMP, and Baum and Koester (2011) and entails imposing the restriction that government spending does not respond to business cycle shocks within a quarter.[5]

While the timing approach can be justified using institutional knowledge in small VARs, it is much more challenging to justify it with 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 shocks in other taxes within a period). Ramey (2011), Owyang, Ramey, and Zubairy (2013), Ramey and Zubairy (2015, 2016), Cloyne (2013), Romer and Romer (2010), Caggiano, Castelnuovo, Colombo, and Nodari (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 government spending shocks or tax shocks that are constructed by examining historical announcements about changes in government spending and taxes. Because our goal is to trace the effects of government 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 consider only military spending shocks, this means that many of the observations for the narrative shock series are equal to zero, which makes exploring state-dependence challenging from an econometric point of view. Therefore, we use the sign-restriction approach to identify structural shocks in our medium-scale TVAR model that controls for what could be possible omitted variables in a smaller VAR identified with timing restrictions.

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 restricting the sign of the response of variables over particular horizons. This approach is usually considered more

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

general and agnostic than the timing restrictions approach because it effectively nests the zero restrictions imposed by the timing restrictions. Using sign restrictions and a penalty function approach in a linear fiscal VAR, Mountford and Uhlig (2009) study the effects of fiscal policy and taxes, finding 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) and 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 one. However, recent developments in the time series literature have cast doubt on some of the results obtained using penalty functions. Arias, Rubio-Ramirez, and Waggoner (2016) 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 four structural shocks that are identified using sign restrictions on the impulse responses, with the impulse responses constructed using the QR decomposition approach and sampler proposed by Arias, Rubio-Ramirez, and Waggoner (2016).[6]

The four shocks we identify are a government spending shock, a tax shock, a "business cycle" shock, and a monetary policy shock. Table 1 summarizes the sign restrictions used to identify these shocks. The first sign in each cell shows the assumed direction of the effect of a shock on the response variable on impact; the second and third signs in each cell are the assumed signs in the first and second quarter following the shock. Question marks indicate that the sign is left unrestricted. All shocks are orthogonal to one another, which differs

[6] Pagan and Fry (2011) also suggest using the QR decomposition in larger VARs where the shocks are identified using sign restrictions. It is important to note that the our model also incorporates uncertainty about the threshold estimate and the priors and posteriors are not conjugate. In the linear model, we use the Arias, Rubio-Ramirez and Waggoner (2016) algorithm directly. For the nonlinear model, as noted in footnote 4, the priors are symmetric across regimes. In particular, the prior distributions are centered around the posterior means for the linear model. Figure C1 in Appendix C reports the impulse responses for the linear model, which correspond to the impulse responses implied by the priors for the nonlinear model.

from Mountford and Uhlig (2009) who do not impose the restriction that tax shocks are orthogonal to government spending shocks.

Table 1: Sign Identification

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

As specified in Table 1, a positive business cycle shock increases output, tax revenues, and the measure of slack on impact and for 2 quarters following the shock.[7] Although we do not impose any restrictions on the responses of the interest rate to a business cycle shock, we note that the posterior responses indicate that the interest rate also increases in response to a positive business cycle shock. Meanwhile, a positive monetary policy shock is specified to increase the interest rate contemporaneously and for the subsequent two quarters, while decreasing inflation on impact and for the subsequent two quarters. That is, a "positive" shock is contractionary in the sense of having a negative liquidity and disinflationary effect. However, because there is conflicting evidence from the monetary policy literature (see, for example, Lo and Piger, 2005, and Tanreyro and Thwaites, 2016) about whether the responses of output to monetary policy can vary and possibly be insignificant at some points of the business cycle, we do not impose the restriction that output falls in response to a contractionary monetary policy shock, although our main results did not change when we imposed this restriction. A positive tax shock is assumed to increase tax revenues contemporaneously and for two quarters following the shock. Similarly, a positive government spending shock increases government consumption and investment contemporaneously and for two quarters

[7] Recall that "slack" is conventionally 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. A positive business cycle shock 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 (making the gap less negative or more positive). However, in the cases where the unemployment rate is used as a measure of slack, we reverse the sign of the restricted responses (i.e., unemployment decreases in response to a positive business cycle shock). Also, note that when we consider the disaggregated responses of consumption and investment in Section 4.3, we assume that a positive business cycle shock increases consumption and investment. As in Mountford and Uhlig (2009), positive business cycle shock in this context could be consistent with either demand of supply shocks, and we remain agnostic on the source of business cycle fluctuations.

following the shock.[8] Following previous results from the fiscal literature, we also impose the restriction that output increases on impact in response to a positive 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 that find government spending multipliers decline sharply after the first quarter find positive multipliers on impact (for example, Ramey and Zubairy, 2016).[9] Meanwhile, the responses of output are left unrestricted after impact.

The responses to negative shocks have the opposite signs from the signs shown in Table 1. In the case where we consider the evolving-state impulse responses, the responses are constructed assuming that the economy evolves endogenously from one regime to another, and an orthogonalization is accepted if the sign restrictions hold for two quarters even if the economy evolves from one regime to another. The technical details of the impulse response calculation are discussed in detail in Appendix A.

3 Revisiting the Evidence for Nonlinearity and State Dependence

In FMP, we found strong state-dependent effects of government spending shocks on output based on generalized impulse response analysis for a four-variable TVAR model. We identified shocks with timing restrictions and set the threshold variable as capacity utilization adjusted for a structural break in 1973. However, a recent study by Ramey and Zubairy (2016) using local projection methods and narrative shocks instead raises the concern that our results may not be robust to considering alternative threshold variables. In addition, they point out that there is lower correlation between the CBO output gap and capacity

[8]We also consider an alternative identification scheme where the restrictions are imposed for four 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.

[9]Assuming the responses of output on impact based on the consensus from the previous empirical literature is done to speed up the Bayesian estimation, including the calculation of marginal likelihoods. However, the results presented in the next sections do not hinge on these assumptions. If the response of output to government spending and taxes were left unrestricted, almost the entire posterior distribution of the response of output to a positive government spending shock or a negative tax shock would be above zero at horizon zero, while almost the entire posterior distribution of the response of output to a negative government spending shocks or a tax increase would be below zero. Therefore, the slightly more restrictive prior is supported by the data. The full set of responses for different identification schemes is available from the authors.

utilization adjusted for an imposed structural break compared to the correlation between the CBO output gap and the raw capacity utilization series. They also raise similar concerns about the sensitivity of the state-dependent results reported in Auerbach and Gorodnichenko (2013 a,b).

Because a major goal of this paper is to explore when different types of discretionary fiscal policy are effective in the presence of state dependence, we revisit the evidence for nonlinearity to address Ramey and Zubairy's (2016) concerns about the existence and nature of state dependence in the first place. In particular, we discuss potential issues when selecting threshold variables in a threshold model and then we evaluate the evidence of nonlinearity and state dependence for different threshold variables, different subsamples, and modifications of the identification scheme from Table 1.

From a macroeconometric point of view, there are three main issues that arise in this framework that could complicate determining the best choice of threshold variable, as well as evaluating whether there is nonlinearity in the first place. First, any measure of slack should accurately capture the true degree of under (or over) utilization of resources in the economy. Second, the threshold 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, and Perron and Wada, 2016). The nonlinear fiscal spending multiplier literature has used different observed variables as potential proxies of slack. In FMP, we used capacity utilization adjusted for an imposed structural break in 1973Q4. The capacity utilization series is survey-based and 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 slack obtained from a wide range of linear and nonlinear time series models and find that, as an observable data series, capacity utilization is particularly highly correlated with their composite measure of slack calculated by averaging across many different estimates 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 adjusted for the imposed break date as the threshold variable. Meanwhile, a large number of studies use the CBO output gap (see, for example, Baum and Koester, 2011, and Baum, Poplawski-Ribeiro, and Weber, 2012) and find evidence of state dependence similar to the evidence of state dependence found in FMP. Auerbach and Gorodnichenko (2012, 2013a,b) consider moving averages of output growth and find state dependence, while 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 based on a TVAR model in particular, it is important to ensure that the threshold variable is stationary and not subject to 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 threshold variable is subject to structural breaks, conventional model comparisons can erroneously identify the structural break as evidence in favor of nonlinearity. In FMP, we imposed a structural break in the level of capacity utilization that matches the productivity slowdown identified by Perron and Wada (2009).[10] However, when applying tests for multiple breaks to all of the commonly used measures of slack, we find evidence of multiple structural breaks in capacity utilization, in the unemployment rate (consistent with the findings of Ghiblavi, Murray, and Papell, 2000), and in the CBO output gap. Table 2a

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

summarizes the results of conventional tests for structural breaks in the levels of capacity utilization, the CBO output gap, and the unemployment rate. Meanwhile, to account for the possibility that nonlinearity could be driven not by the degree of slack, but by asymmetries in terms of the direction of changes in economic activity or in fiscal policy, we also consider the following possible threshold variables: a 4-quarter moving average change in log output, a 4-quarter moving average change in the log of government spending and consumption, and a 4-quarter moving average change in log tax revenues (net of transfer taxes).[11] For completeness, we also consider the level of the ex ante real interest rate (based on static expectations) as a threshold variable to allow for the possibility that any asymmetry is related to the stance of monetary policy rather than fiscal policy. Table 2b reports the results of the structural break tests for these additional variables in the cases where there is evidence of structural breaks.[12]

[11] We also considered longer moving averages of output growth, as in Auerbach and Gorodnichenko (2012). Results were similar to those for the 4-quarter moving average.

[12] There was no evidence of a structural break in the moving averages of the changes in government spending and tax revenues. Likewise, the real interest rate was not affected by the productivity slowdown, although there is a large literature that suggests the behavior of real interest rates changed in 1984. Therefore, for the imposed break in this case, we use 1984Q1.

Table 2: Structural Breaks

(a) Measures of Slack

		Break Dates and Test Statistics		
Maximum Number of Breaks	Imposed or Estimated	Capacity Utilization	CBO Output Gap	Unemployment Rate
1	Imposed	1973$Q4$ $F = 43.22$ (< 0.001)	1973$Q4$ $F = 43.85$ (< 0.001)	1973$Q4$ $F = 35.78$ (< 0.001)
1	Estimated (LR)	2001$Q1$ 102.37 (< 0.001)	1974$Q3$ 74.11 (< 0.001)	1974$Q3$ 38.85 (< 0.001)
5	Estimated (BP)	2001$Q1$ 27.75 (< 0.001) 1974$Q4$ 9.50 (0.07)	1974$Q1$ 23.46 (< 0.001) 2008$Q4$ 24.48 (< 0.001)	1974$Q3$ 17.43 (< 0.001) 1987$Q1$ 10.41 (< 0.001) 2008$Q4$ 14.05 (< 0.001) 1996$Q1$ 13.71 (< 0.001)

(b) Growth Rates and Policy

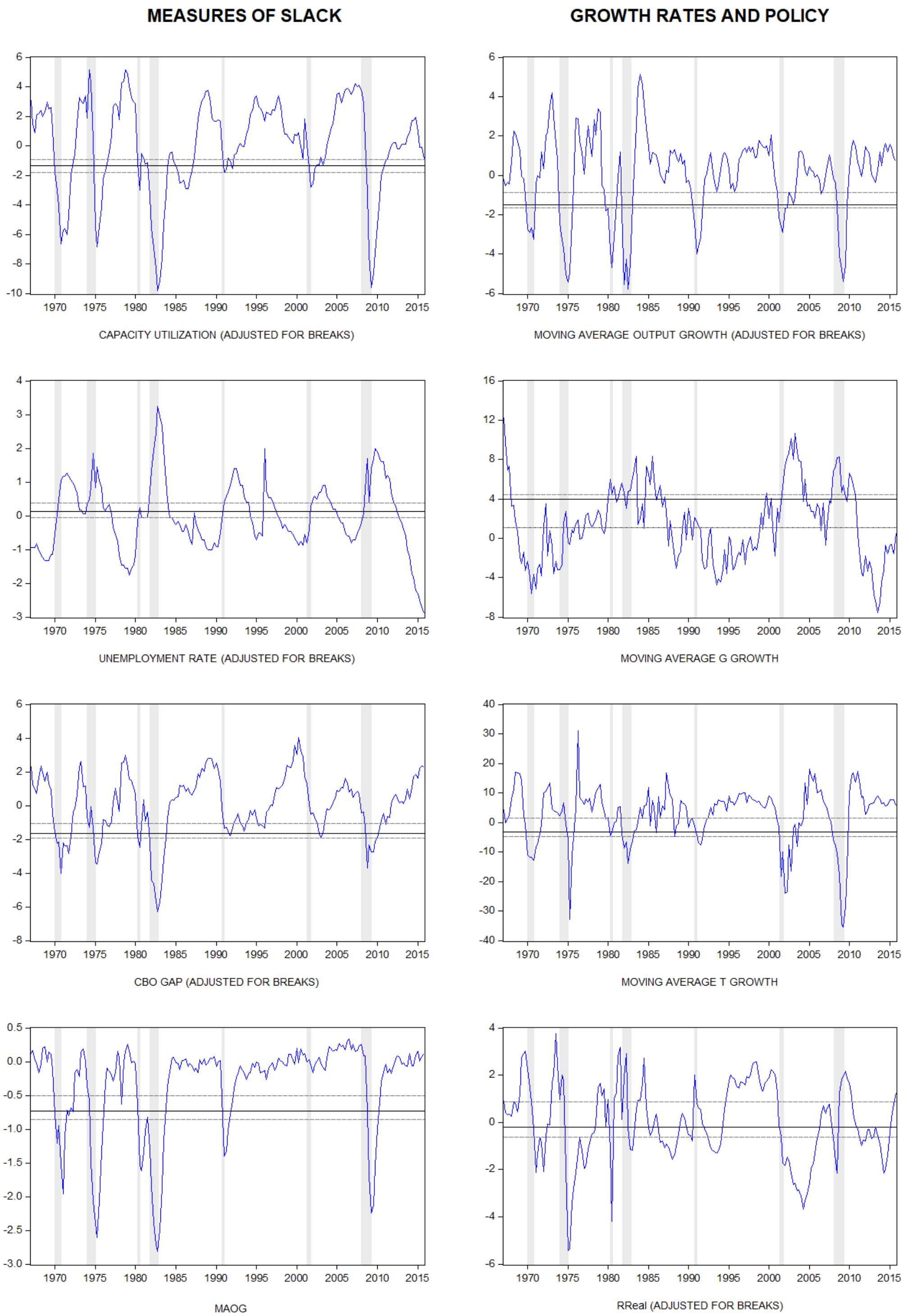
		Break Dates and Test Statistics	
Maximum Number of Breaks	Imposed or Estimated	Moving Average of Output Growth	Real Interest Rate
1	Imposed	1973$Q4$ $F = 3.86$ (0.06)	1984$Q1$ $F = 1.37$ (0.24)
1	Estimated (LR)	2007$Q1$ 24.11 (< 0.001)	2001$Q4$ 126.11 (< 0.001)
5	Estimated (BP)	2007$Q1$ 8.07 (0.03)	1979$Q4$ 33.58 (< 0.001) 1990$Q4$ 27.08 (< 0.001) 2008$Q4$ 23.65 (< 0.001)

The break dates for the multiple break test are estimated using a Bai-Perron test of a break in mean with Newey-West HAC standard errors and a 10% significance level. Results are only reported for variables in the cases where there is evidence of structural breaks.

Looking at Table 2, the 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 break-adjusted capacity utilization and the CBO output gap is lower than the correlation between unadjusted capacity utilization and the CBO output gap. Meanwhile, given the evidence of multiple additional structural breaks in the various series, we re-examine the evidence in favor of nonlinearity for a TVAR model considering different combinations and permutations of variables as the measure of slack and as the threshold variable, always adjusting for structural breaks.

Again, for the measure of slack in the TVAR model, we consider the following: capacity utilization, the unemployment rate, the CBO output gap, and the MAOG from Morley and Panovska (2017). All measures of slack are adjusted for structural breaks, with the break dates estimated from the data. 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 univariate time series models that are commonly considered in the empirical macroeconomics literature to model quarterly real GDP. We estimate a total of 29 different time series models, both linear and nonlinear, that imply different estimates of the long term trend in output and then we average the estimates for the output gap across the different models. The MAOG is calculated using the full available data sample for US real GDP (1947Q1-2016Q1) and is treated as data in the TVAR model. 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.

Figure 1: Threshold Variables and Estimated Thresholds



The threshold variables are adjusted for structural breaks in mean according to the estimated break dates reported in Table 2. The estimated threshold (median) and 90% credibility intervals from Tables 3 and 4 are also displayed.

Figure 1 plots all of the possible threshold variables considered in our analysis. The left panels plot the measures of slack, and the right panel plots the additional threshold variables. For ease of direct comparison with FMP, the sample period for all models starts in 1967, although we consider robustness of our results to different sample periods later. Table 3 reports the threshold estimates, the highest posterior log likelihood for each model, and the log marginal likelihoods. It can be seen that the implied Bayes factors strongly favor the TVAR model over the linear counterpart in all cases.[13] As can be seen from Figure 1, the threshold estimates split up the sample into periods of excess slack (recessions and the recovery periods immediately following the recessions) and "normal" times when the economy is closer to potential. Notably, from Table 3, the MAOG is selected as the preferred threshold variable in all cases and the estimates of the threshold for it are quite precise. Furthermore, as can be seen by comparing Tables 3 and 4, any of the slack variables is strongly preferred as a threshold variable compared to the policy variables, while the MAOG is preferred over the moving average of output growth, even though both identify similar dates for the regimes, as can be seen in Figure 1. Therefore, for most of our analysis, we focus on the model with the MAOG as a measure of slack and as a threshold variable as our benchmark case. However, to ensure that our results are not seen to be driven entirely by the choice of the measure of slack or the threshold variable, we also consider the responses using the other measures of slack.

[13]The Bayes factor is the ratio of marginal likelihoods and is 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.

Table 3: Model Selection Criteria for Competing Measures of Slack and Threshold Variables

	Measure of Slack in the VAR							
Threshold Variable	*capu*		*unrate*		*CBOgap*		*MAOG*	
Linear model (none)	−2670.03 −2676.38 −822.73		−2525.25 −2523.81 −507.72		−2430.87 −2429.42 −506.87		−2394.05 −2389.91 −833.73	
capu	−2238.54 −2239.00 −339.26	−1.35 (−1.66, −1.04)	−2137.32 −2029.01 −339.26	−2.00 (−2.63, −0.28)	−2028.08 −2029.47 −344.21	−0.30 (−1.79, 0.23)	−1938.70 −1938.02 −568.77	−1.33 (−1.82, −0.91)
unrate	−2294.43 −2293.77 −339.26	0.59 (−0.06, 0.73)	−2152.21 −2151.00 −362.84	0.06 (−0.29, 0.38)	−1764.52 −1761.30 −341.39	0.88 (0.06, 1.04)	−1958.42 −1958.63 −542.63	0.13 (−0.05, 0.38)
CBOgap	−2231.54 −2231.70 −303.37	−1.64 (−1.95, −0.84)	−2089.74 −2091.77 −329.92	−1.66 (−1.95, −0.66)	−1597.20 −1598.00 −312.61	−2.00 (−2.16, −1.30)	−1916.56 −1918.22 −487.05	−1.64 (−1.94, −1.06)
MAOG	**−1937.81** **−1937.00** **−269.66**	−0.69 (−0.74, −0.52)	**−2085.69** **−2084.27** **−329.92**	−0.74 (−0.80, −0.39)	**−1579.30** **−1578.33** **−309.65**	−0.71 (−0.82, −0.39)	**−1873.11** **−1870.00** **−450.37**	−0.74 (−0.86, −0.51)

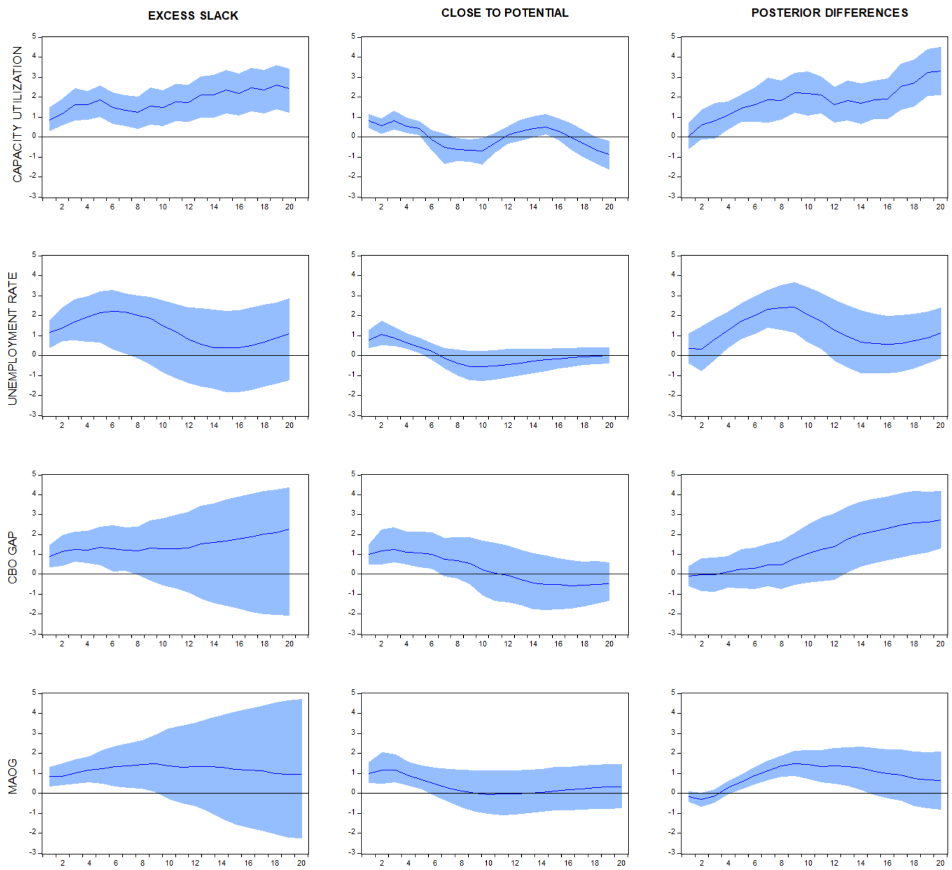
Each cell reports the log likelihood obtained using maximum likelihood estimation, the expected posterior log likelihood obtained Bayesian estimation, and the log marginal likelihood (Top, Middle, Bottom). The second entry is the threshold estimate, including 90% credibility intervals, obtained from the posterior Bayesian distribution. The best model fit for each measure of slack is reported in bold.

Table 4: Model Selection Criteria for Other Threshold Variables

	Measure of Slack in the VAR	
Threshold Variable	*MAOG*	
Linear model (none)	−2394.05 −2389.91 −833.73	
ΔY	−1879.45 −1878.00 −454.37	−1.51 (−1.67, −0.89)
ΔG	−1966.62 −1962.79 −503.65	3.93 (1.06, 4.45)
ΔT	−1924.22 −1923.16 −490.12	−3.31 (−4.79, 1.34)
r	−1953.92 −1953.00 −515.22	−0.21 (−0.62, 0.87)
MAOG	**−1873.11** **−1870.00** **−450.37**	−0.74 (−0.86, −0.51)

Each cell reports the log likelihood obtained using maximum likelihood estimation, the expected posterior log likelihood obtained Bayesian estimation, and the log marginal likelihood (Top, Middle, Bottom). The second entry is the threshold estimate, including 90% credibility intervals, obtained from the posterior Bayesian distribution. The growth rate threshold variables are 4-quarter moving averages of log differences. The real interest rate is an ex ante measure given static expectations. The best model fit is reported in bold.

Figure 2: State-Dependent Effects of Government Spending on Output Using Different Measures of Slack



The figure displays the fixed-regime responses of output to a government spending shock in the excess slack regime (left column) and close to potential regime (middle column) when using different measures of slack in the TVAR. The right column plots the posterior differences in responses across the two regimes. The rows correspond to results for different measures of slack from top to bottom of capacity utilization, the unemployment rate, the CBO output gap, Morley and Panovska's (2017) MAOG, all adjusted for structural breaks. Posterior medians and 90% credibility intervals are reported.

Figure 2 plots the impulse responses to a government spending shock for different measures of slack when only that measure of slack is considered as a threshold variable (i.e., capacity utilization as the measure of slack and as the threshold variable, the unemployment rate as the measure of slack and as the threshold variable, etc...). The left panels plot the fixed-regime responses for the "excess slack" regime (defined by the estimated threshold), the middle panels plot the fixed-regime responses when the economy is in the other regime, and the right panels plot the posterior differences for the responses between the two regimes. All responses are dollar-for-dollar, converted from log-to-log responses using the average G_t/Y_t

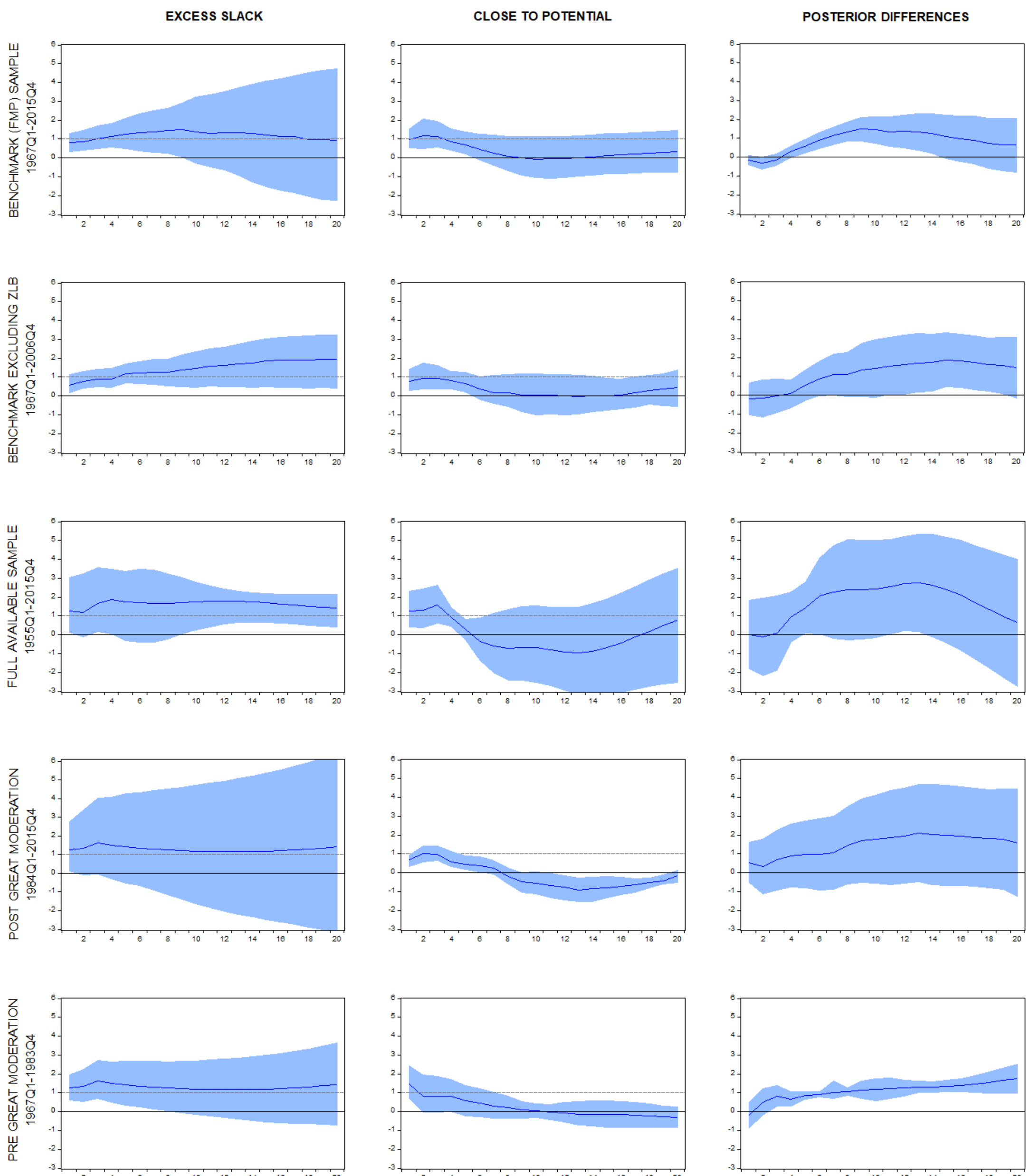
ratios for the corresponding regimes for each draw of the threshold parameter.[14] Median responses and 90% credibility intervals are reported. The evidence in favor of state dependence in response to a government spending shock is similar to the results presented in FMP based on a smaller TVAR and timing restrictions: in the excess slack regime, output responds with a large and persistent increase to a positive shock to government spending, while the response is smaller and less persistent when the economy is close to potential. The pattern is consistent across the different measures of slack, indicating that the results in FMP were not driven by the use of capacity utilization with an imposed break.[15]

Similarly, to ensure that the evidence for state dependence is not driven by the sample period based on the particular availability of capacity utilization, we re-estimate the model with the MAOG as the threshold variable for alternative sample periods, including subsamples to address the role of possible structural change in the TVAR and interest rates hitting the zero lower bound (ZLB) in recent years. Figure 3 plots the responses for the benchmark sample, for the benchmark sample excluding the Great Recession/ZLB period (1967Q1-2006Q4), for the sample extended back to 1955Q1, for the post-Great Moderation period (1984Q1-2015Q4), and for the pre-Great Moderation period (1967Q1-1983Q4). Although the credibility intervals for some of the shorter subsamples are wide, the pattern across the different samples is quite similar to the pattern in Figure 2, indicating that the evidence for state dependence is robust across sample periods. In particular, even if there are changes in the volatility of shocks with the Great Moderation or in some of the parameters in the TVAR over time, the pattern of state dependence remains fairly stable.

[14] The history-dependent evolving-regime responses presented in the next section are converted using the ratios at each corresponding history, with responses to tax shocks converted using T_t/Y_t ratios.

[15] Figure C1 in Appendix C plots the impulse responses for the linear VAR model for the benchmark case when the MAOG is used as a measure of slack. These responses correspond to the responses for the prior for the nonlinear model. As pointed out by Arias, Rubio-Ramirez and Waggoner (2016), the impulse responses obtained using their approach have much wider credibility bands than the responses reported by Mountford and Uhlig (2009). However, when we allow for nonlinearity, the impulse responses are more precisely estimated, implying that some of the uncertainty in the linear responses could also be driven by state dependence that is not accounted for in the linear model. Furthermore, the posterior distribution for the excess slack regime looks substantially different from the posterior distribution for the linear model, thus indicating that the findings in favor of state dependence are not driven by our choice of priors.

Figure 3: State-Dependent Effects of Government Spending on Output for Different Sample Periods



The figure displays the fixed-regime responses of output to a government spending shock in the excess slack regime (left column) and close to potential regime (middle column) for the benchmark model with the MAOG as the measure of slack and threshold variable estimated for different sample periods. The right column plots the posterior differences in responses across the two regimes. The rows correspond to the results for different sample periods. Posterior medians and 90% credibility intervals are reported.

Figure 4: Government Spending Shocks Identified from the Benchmark TVAR and Ramey's (2011) Narrative Approach

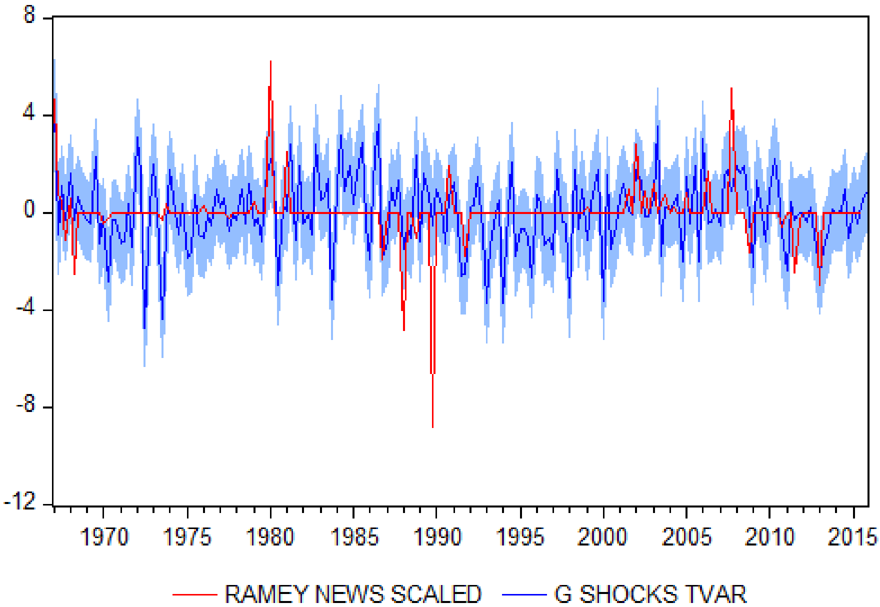
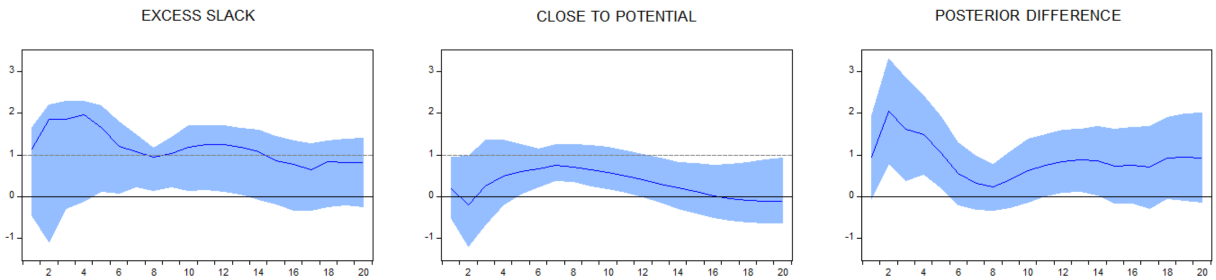


Figure 4 plots the government spending shocks identified from the benchmark TVAR model against Ramey's (2011) military spending news shocks.[16] Our identified government spending shocks pick up most military spending news shocks reasonably closely. However, in few cases, military spending announcements lead the shocks identified from the TVAR. To ensure that the evidence for state dependence is not driven by the particular identification scheme in Table 1, we also consider an alternative identification scheme whereby we consider responses to announcement shocks, as in Mountford and Uhlig (2009). For this alternative scheme, government spending and taxes do not increase for the first 4 quarters after a shock, but they then increase in quarters 5 through 8. All other shocks are defined as in Table 1, although the sign of the response of output is left unrestricted, as there is no consensus in the literature on the effects of announcement shocks. Figure 5 plots the responses to the shocks obtained using this alternative identification scheme. The pattern is quite similar to the pattern for the benchmark model, suggesting that our main findings are not due to any

[16]In terms of tax shocks that are considered along with government spending shocks in the next section, Figure C2 in Appendix C plots the tax shocks identified from the TVAR against Romer and Romer's (2010) narrative unanticipated tax shocks. Our tax shocks match the Romer and Romer shocks quite closely.

issues with identifying announcement effects.[17]

Figure 5: State-Dependent Effects of Government Spending Announcements on Output Using a Different Identification Scheme



The figure displays the fixed-regime responses of output to a government spending announcement shock in the excess slack regime (left column) and close to potential regime (middle column) for a TVAR with government spending and taxes restricted to increase only a year after the respective announcement shocks. The right column plots the posterior differences in responses across the two regimes.

The similarity of our government spending shocks to the narrative shocks suggests that some of the past reported differences in impulse responses are related to other factors, including the fact that most studies with narrative shocks use local projection methods with a linear approximation, while a TVAR data generating process could require higher-order (e.g., quadratic, cubic, etc...) terms for the projection to accurately approximate the true impulse responses (which are consistently estimated by the simulation-based approach used in, inter alia, FMP). This issue was first raised in the original study by Jorda (2005), who cautioned that incorporating higher-order terms might be necessary in nonlinear models when 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–e.g., see Teräsvirta and Yang (2016), who 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, while even the linear approximation could

[17] We should note that it is unclear whether Arias, Rubio-Ramiraz, and Waggoner (2016) approach corrects the credibility intervals in this setting, although it would given a fixed and known threshold. Regardless, the robustness of the pattern of impulse responses is reassuring.

be acceptable if the speed of transition between regimes is relatively low. However, given the fact that the transition between regimes for our TVAR model is abrupt, an accurate approximation likely requires many more higher order terms. We therefore focus on using impulse responses obtained directly from the TVAR rather than on projection-based impulse responses.[18]

The model selection criteria, the threshold estimates, and the impulse responses for different threshold variables and different samples yield similar results in terms of the timing and implications of nonlinearity, thus assuaging the concerns that previous findings in favor of nonlinearity were driven by idiosyncratic behavior in a specific threshold variable, by the choice of the threshold variable, or by the choice of the sample period. Furthermore, in the next section, when the log-to-log responses are converted to dollar-to-dollar responses using the ratios at each history of interest instead of the average ratios, as suggested by Ramey and Zubairy (2016), the results look quite similar to the results obtained using the the average ratios.

Given the strong support in favor of nonlinearity and state dependence both from the previous literature and from our results, we turn next to exploring more realistic scenarios for fiscal policy shocks in which the economy is allowed to evolve endogenously from one regime to another. This also allows us to consider possible sign and size asymmetries and to determine when discretionary fiscal policy is effective.

4 Impulse Response Analysis of Discretionary Fiscal Policy

4.1 Evidence of State Dependence for a Discretionary Change in Taxes

Before we consider evolving-regime responses to fiscal shocks, we first establish that tax shocks have similar state-dependent effects as do government spending shocks. Figure 6 plots the fixed-regime responses of output to a shock in government spending and to a shock in

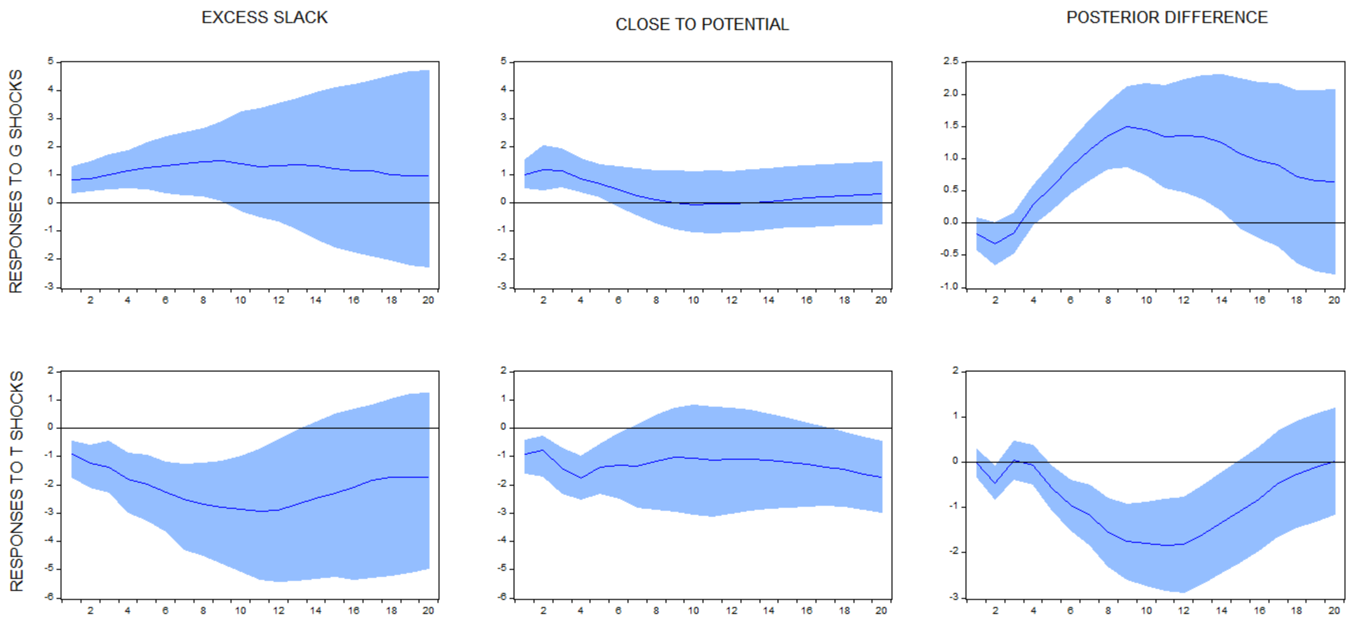
[18] We also considered a smooth transition specification, similar to Auerbarch and Gorodnichenko (2012), as an alternative to our benchmark TVAR benchmark. When the transition speed parameter is calibrated and equal to 2, as in Auerbach and Gorodnichenko (2012), the log marginal likelihood is -463.6, indicating a substantially worse fit than our benchmark TVAR model presented in Table 3. When the transition speed is estimated using a rough grid for the speed of transition, the estimated transition speed is 4.5 and the transition function is abrupt and, therefore, very close to that for a TVAR model.

taxes, along with the posterior differences between the states (the responses to a government spending shock reproduces the bottom row of Figure 2 for purposes of comparison with the response to a tax shock). All fixed-regime responses are converted to dollar-to-dollar responses using the average G_t/Y_t and T_t/Y_t ratios for the corresponding regime for each draw of the threshold parameter. By contrast, the history-dependent evolving responses presented in the next subsections are converted using the ratios at the corresponding history.

As can be seen in Figure 6, an increase in government spending increases output by \$1.5 after 5 quarters in the excess slack regime, with the response being quite persistent. By contrast, when the economy is close to potential, an increase in government spending temporarily increases output on impact, but the response dies out and becomes negative after two years. Tax cuts have similar state dependence. Flipping the sign of the displayed response to a tax hike, the fixed-regime impulse responses suggest that a tax cut would increase output in both regimes. A tax cut in the excess slack regime increases output by \$2 (dollar-for-dollar) and is significant for 13 quarters, whereas the response is smaller in the other regime, peaking at \$1.3 and becoming insignificant after 7 quarters.

The responses in Figure 6 (and all of the responses presented in Figures 2, 3, and 5) 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. Notably, even when accounting for these different sources of uncertainty, there is still evidence of state dependence in the responses of output to fiscal policy. In particular, the posterior differences in the right column of Figure 6 are large in magnitude and different from zero. The response to government spending shocks in the excess slack regime is approximately \$1.4 dollars higher after 8 quarters than in the regime when the economy is close to potential, while the response to a tax cut in the excess slack regime is approximately \$1.5 dollars higher. Therefore, the fixed-regime responses show that there is clear state dependence in the responses of output both to government spending and to taxes.

Figure 6: State-Dependent Effects of Government Spending and Taxes on Output



The figure displays the fixed-regime responses of output to a government spending shock (first row) and a tax shock (second row) in the excess slack regime (left column) and close to potential regime (middle column) for the benchmark model. The right column plots the posterior differences in responses across the two regimes. Posterior medians and 90% credibility intervals are reported.

4.2 Evolving-Regime Responses of Output

Although we find conclusive evidence in favor of state dependence when comparing the fixed-regime responses, policymakers are more concerned with the responses 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 with the regime assumed to be fixed forever. In addition, while the fixed-regime responses are quite useful for testing state dependence across regimes, the responses within a regime are linear by construction–i.e., they are proportional to the size and sign of a shock. However, if the economy is allowed to evolve across regimes, threshold models admit (but do not impose) the possibility that negative shocks can have different effects from positive shocks or that shocks of different magnitudes have non-proportional effects. Thus, building on the evidence of state dependence, we employ generalized impulse response analysis to evaluate whether there is significant evidence of sign or size dependence allowing the economy to evolve from one regime to another.[19]

We focus on three particular 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 threshold variable is close to the threshold level, and a strong expansion. In particular, the three histories that we consider are the following:

- 1996Q1: a robust expansion, when the economy is usually classified as being above or close to potential according to our various threshold variables and threshold estimates;
- 2008Q3: a deep recession, when the economy is clearly classified as the excess slack regime according to our various threshold variables and threshold estimates;
- 2012-2014: a sluggish recovery, when the economy is not in an NBER recession, but is close to the estimated threshold for at least some of our threshold variables and threshold estimates.[20]

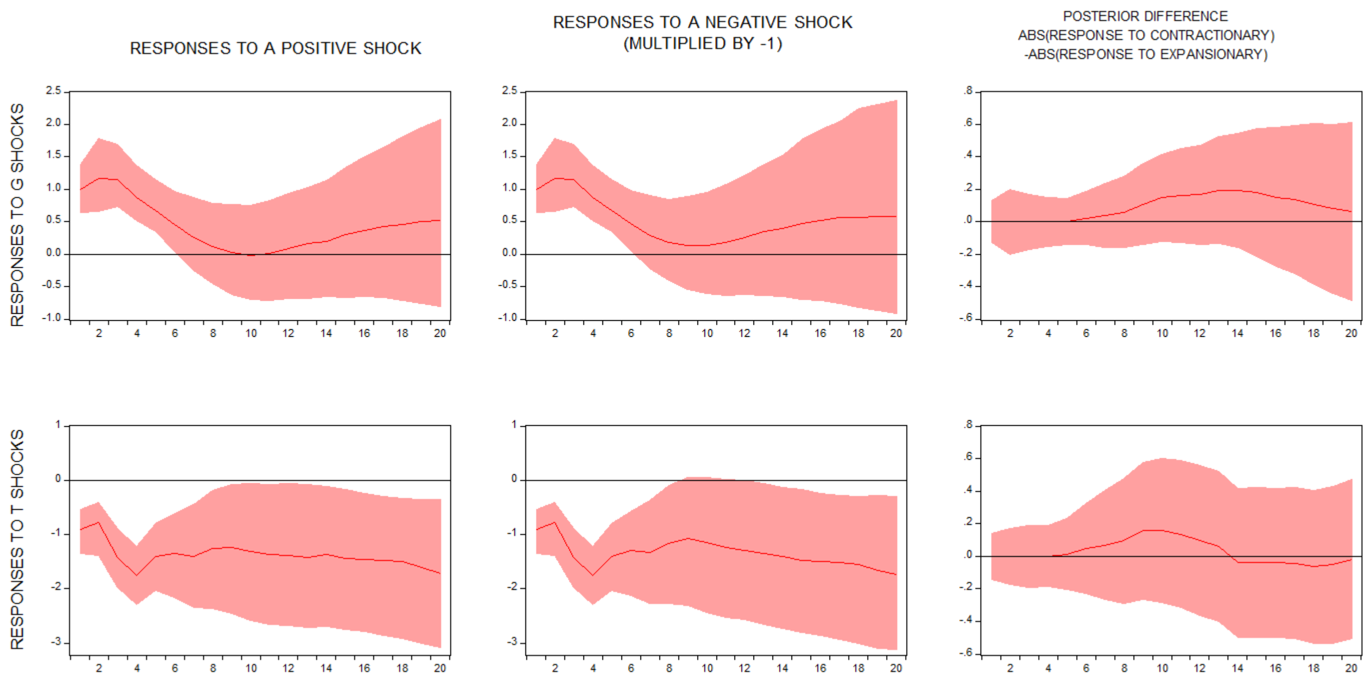
[19]For computational simplicity when calculating the generalized impulse response functions, we abstract from the parameter uncertainty and fix the parameters at their highest posterior density values.

[20]It is important to note that averaging over similar histories of robust expansions, deep recessions, or sluggish recoveries produces very similar results to considering these specific histories.

For a given history (or points within the set of histories in the sluggish recovery scenario), we calculate the responses to an increase in government spending and taxes and to decreases in government spending and taxes scaled to either 1% of GDP or 3% of GDP to consider the impact of the size of a shock. Sign restrictions are simply reversed for negative shocks.

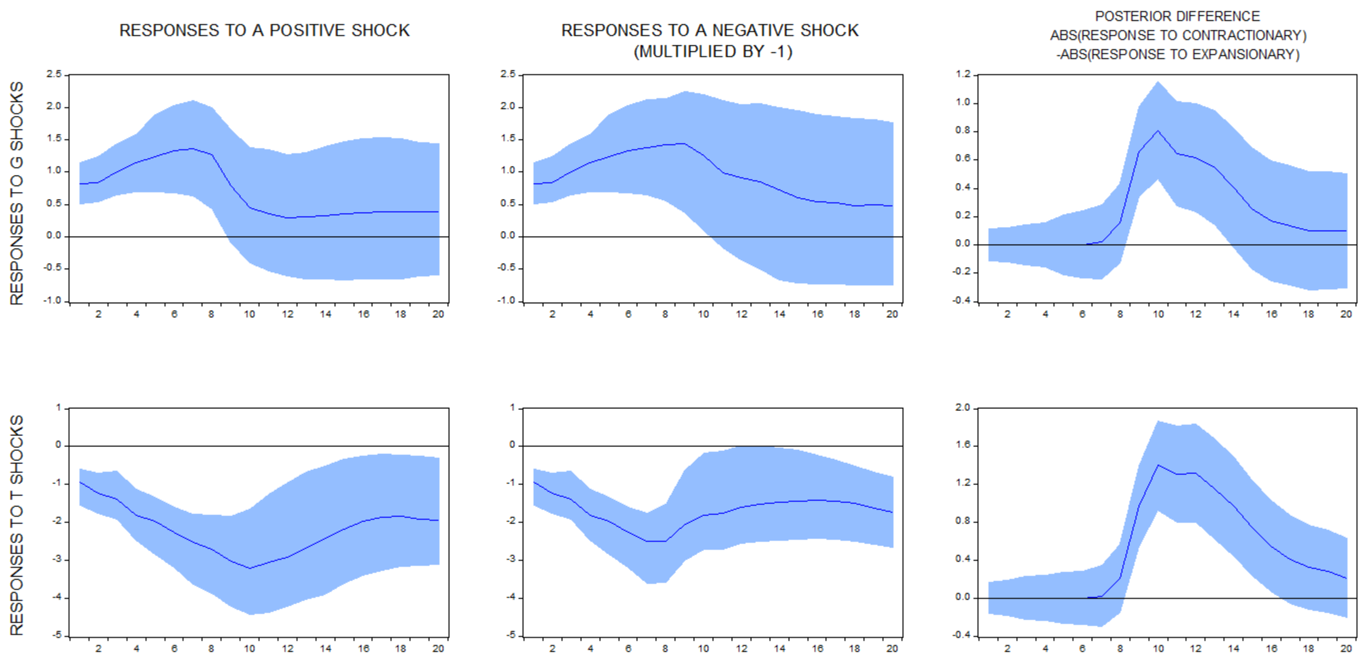
Figure 7 plots the responses of output to changes in government spending and taxes when the economy is in robust expansion and Figure 8 plots the responses of output when the economy is in a deep recession. In both cases, a shock scaled to 1% of GDP is considered. The top panels of the figures plot the responses to government spending, the bottom panels plot the responses to tax changes. The left column plots the responses to positive shocks (higher government spending or higher taxes), the middle panel 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 7: Sign-Dependent Effects of Government Spending and Taxes on Output in a Robust Expansion



The figure displays the evolving-regime responses of output to a government spending shock (first row) and a tax shock (second row) for the benchmark model in a robust expansion. The left columns plot the responses to a positive shock, the middle column plots the response to a negative shock (scaled by -1 for ease of comparison), and the right column plots the difference in magnitude of (scaled) responses for positive and negative shocks. The shocks are equal to 1% of GDP. Posterior medians and 90% credibility intervals are reported.

Figure 8: Sign-Dependent Effects of Government Spending and Taxes on Output in a Deep Recession



The figure displays the evolving-regime responses of output to a government spending shock (first row) and a tax shock (second row) for the benchmark model in a deep recession. The left columns plot the responses to a positive shock, the middle column plots the scaled (by -1 for ease of comparison) response to a negative shock, and the right column plots the difference in magnitude of the scaled responses for positive and negative shocks. The shocks are equal to 1% of GDP. Posterior medians and 90% credibility intervals are reported.

The results in Figure 7 show that contractionary shocks (i.e., cuts in government spending or tax increases) have somewhat larger effects, on average, than expansionary shocks when the economy starts in a robust expansion. However, the difference is economically small and not significant. In this case, tax cuts are more efficient at stimulating output than increases in government spending ($1.7 vs. $0.6 after one 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). However, our results for tax increases when the economy is close to potential stand in 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. Our results indicate tax increases have a strong contractionary effect on output.

According to Figure 8, the effects of contractionary shocks are much more persistent and larger than the effects of expansionary shocks when the economy starts in a deep recession. Cuts in government spending decrease output by $1.7 after 9 quarters. Tax increases decrease output by almost $3 after 10 quarters. The responses to tax cuts and increases in government spending are smaller than the responses to tax increases and cuts in government spending (peaking at approximately $2 and $1.5, respectively). Both increases in taxes and decreases in government spending significantly decrease output (the response is different from zero at all horizons for tax increases, and for two years for government spending cuts).

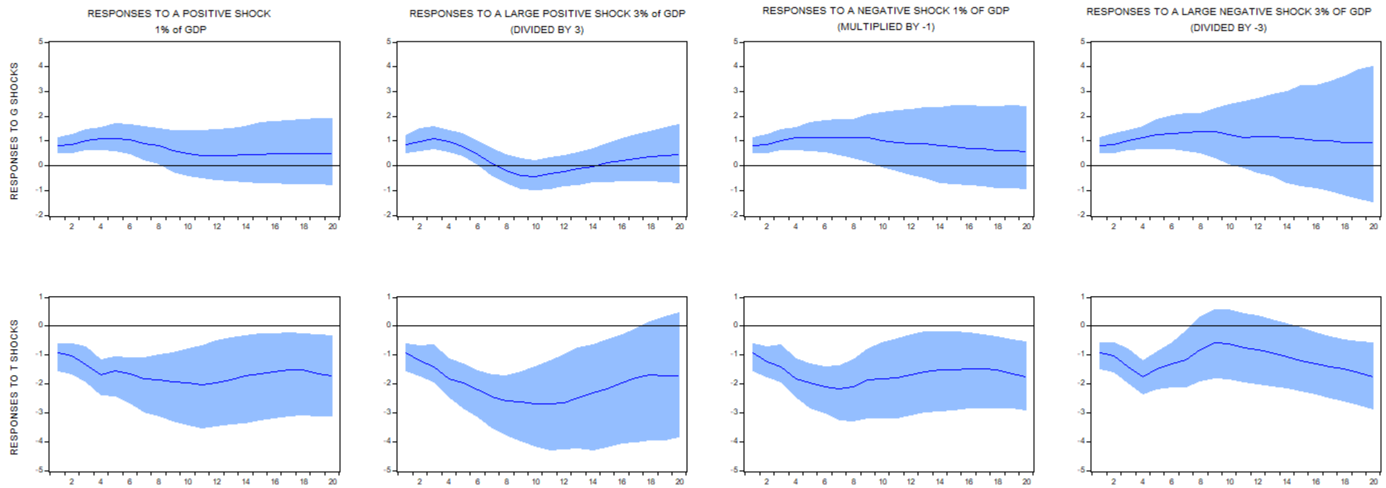
Our results indicate that, if the aim of discretionary fiscal policy is to stimulate the economy, either government spending or tax cuts could be used in periods of excess slack, but tax cuts should be used when the economy is at or above potential. Meanwhile, an advantage of the generalized impulse response approach used here is that, in addition to considering sign asymmetry, it also allows us to evaluate whether there is evidence in favor of size asymmetry at any given history of interest. In particular, size and sign asymmetries might be particularly relevant in a sluggish recovery, when the economy is close to the threshold, given that different shocks can influence the probability of crossing the threshold.

Figure 9 plots the dollar-for-dollar responses of output to "small" (1% of GDP) and "large" (3% of GDP) shocks, both positive and negative, when the economy starts in a sluggish recovery.[21] Figure 10 then plots the posterior differences between the responses to positive and negative shocks, the responses to large positive and large negative shocks, the responses to small and large negative shocks, and the responses to small and large positive shocks.

From Figures 9 and 10, it is clear that the responses to contractionary shocks are larger than the responses to expansionary shocks. This is particularly pronounced when we consider the differences between large shocks. Large contractionary shocks (large negative government spending shocks or large positive tax shocks) have very persistent effects, as they move the the economy below the threshold for an extended period of time. By contrast, large expansionary shocks have positive effects in the short run that quickly die out as the economy gets closer to potential. The responses to large expansionary shocks (positive government spending shocks or negative tax shocks) are proportionally smaller than the responses to smaller expansionary shocks. In particular, when the economy is above the threshold, the dollar-for-dollar responses become more muted. By contrast, the responses to small and large cuts in government spending and to small and large tax increases are more proportionally similar. Contractionary shocks move the economy into the excess slack regime for an extended period of time, making the dollar-for-dollar responses very similar to the fixed-regime responses.

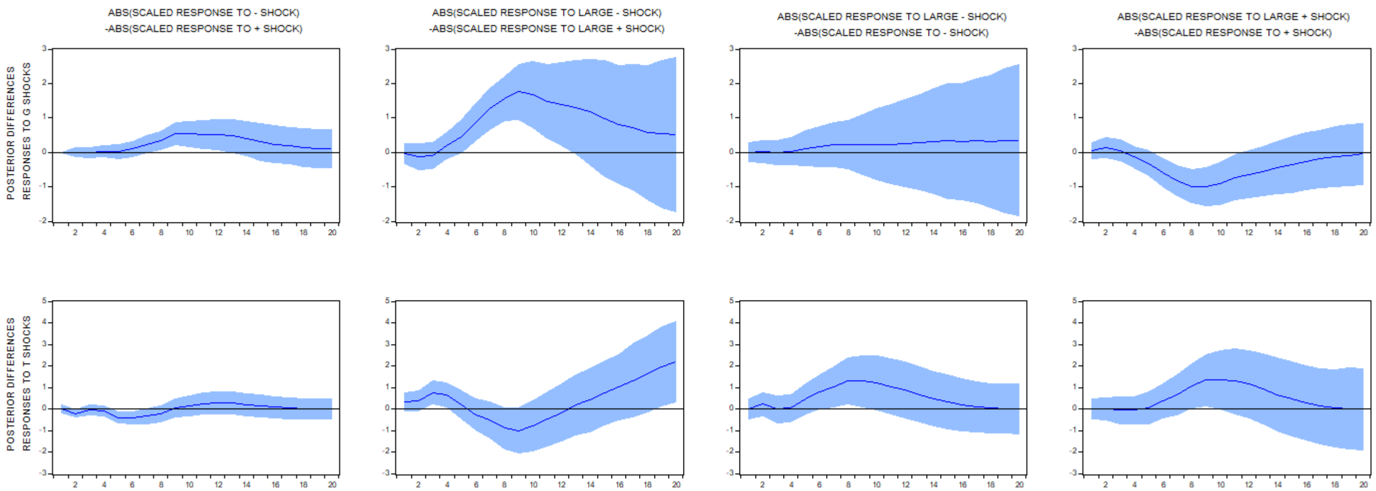
[21] Looking back at Figure 4, it is clear that a shock equal to 3% of GDP is large, although not completely rare.

Figure 9: Sign-Dependent and Size-Dependent Effects of Government Spending and Taxes on Output in a Sluggish Recovery



The figure displays the evolving-regime responses of output to a government spending shock (first row) and a tax shock (second row) for the benchmark model in a sluggish recovery. The first column plots the responses to a positive shock equal to 1% of GDP, the second column plots the scaled (by 1/3 for ease of comparison) responses to a positive shock equal to 3% of GDP, the third column plots the scaled (by -1 for ease of comparison) responses to a negative shock equal to 1% of GDP, and the fourth column plots the scaled (by -1/3 for ease of comparison) responses to a negative shock equal to 3% of GDP. Posterior medians and 90% credibility intervals are reported.

Figure 10: Differences in Sign and Size Effects of Government Spending and Taxes on Output in a Sluggish Recovery



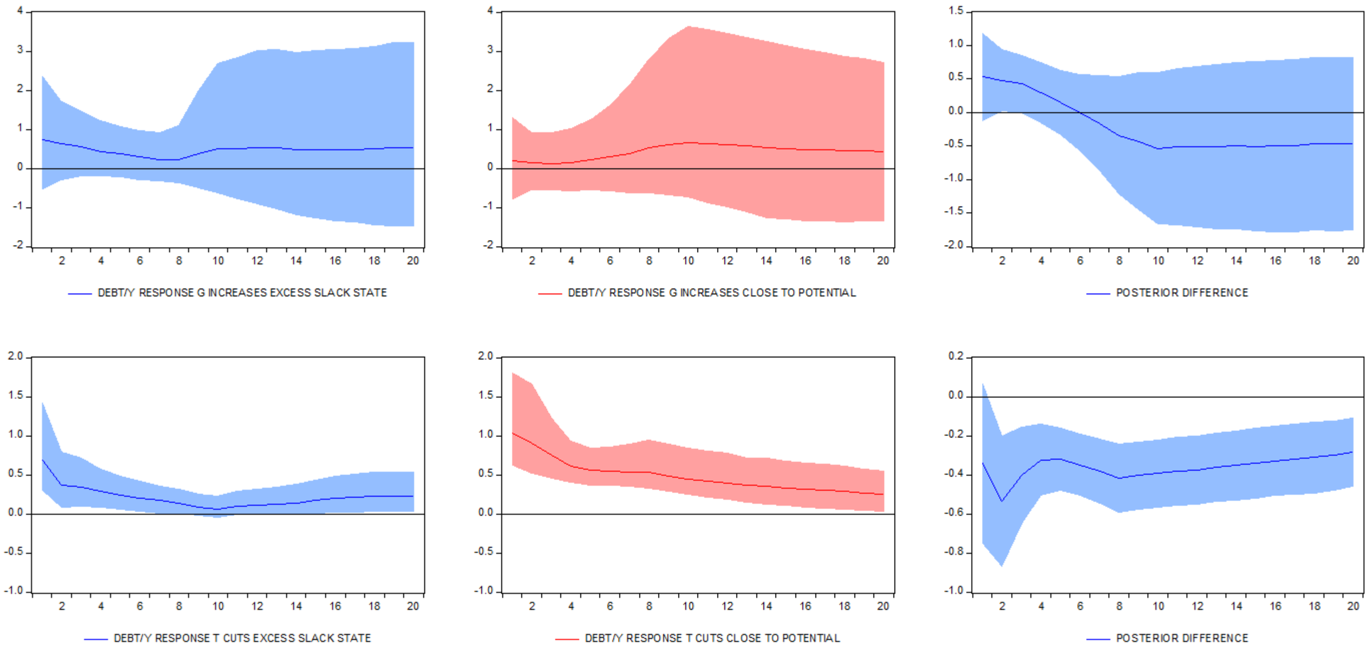
The figure displays the differences in evolving-regime responses of output to a government spending shock (first row) and a tax shock (second row) for the benchmark model in a sluggish recovery. The first column plots the difference in magnitude of (scaled) responses for positive and negative shocks equal to 1% of GDP, the second column plots the difference in magnitude of (scaled) responses for positive and negative shocks equal to 3% of GDP, the third column plots the difference in magnitude of (scaled) responses for a negative shock equal to 3% of GDP and a negative shock equal to 1% of GDP, and the fourth column plots the difference in magnitude of (scaled) responses for a positive shock equal to 3% of GDP and a positive shock equal to 1% of GDP. Posterior medians and 90% credibility intervals are reported.

Our model also allows us to trace the effects of different fiscal shocks on government debt. While austerity measures have stronger effects on output than stimulus, dollar-for-dollar, the difference for government spending shocks is much smaller in a robust expansion. This implies both that austerity measures could sometimes be largely self-defeating and that different fiscal policies of the same magnitude can have different effects on the debt-to-GDP ratio.[22] Figure 11 plots the effects of stimulative measures on the debt-to-GDP ratio. Figure 12 plots the effects of austerity measures on the debt-to-GDP ratio. The top rows plot the responses to a change in government spending and the bottom rows plot the responses to tax changes. The left columns plot the evolving responses when the economy starts in a deep recession, the middle columns plot the responses when the economy starts in a robust expansion, and the the right columns plot the posterior differences.

On average, increases in government spending increase the debt-to-GDP ratio more in a robust expansion than in a deep recession. However, the long-horizon responses are similar because we consider evolving responses where the economy tends to revert to the regime in which the economy is close to potential. By contrast, decreases in taxes raise the debt-to-GDP ratio and this increase is significant, but temporary. In both scenarios, tax cuts have stimulative effects on output and lead to increases in output that mostly offset the increase in deficits. Because the increase in output is larger when taxes are cut in a deep recession, the increase in the debt-to-GDP ratio from a tax cut is smaller given excess slack than when tax cuts are implemented in a robust expansion.

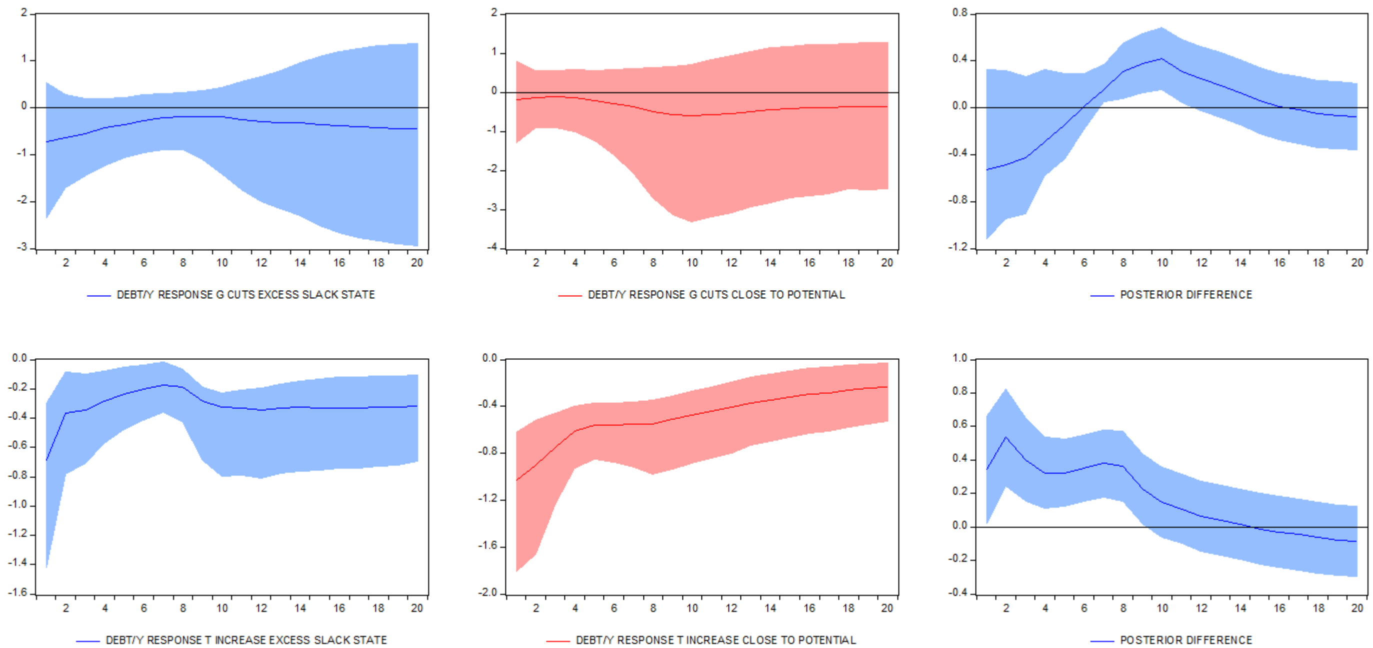
[22] The underlying effects on the government budget deficit are reported in Appendix C.

Figure 11: State-Dependent Effects of Fiscal Stimulus on the Debt-to-GDP Ratio with Evolving Regimes



The figure displays the evolving-regime responses of the debt-to-GDP ratio to a positive government spending shock (first row) and a negative tax shock (second row) for the benchmark model in a deep recession (left column) and in a robust expansion (middle column). The right column plots the difference between the responses in a deep recession and a robust expansion. The shocks are equal to 1% of GDP. Posterior medians and 90% credibility intervals are reported.

Figure 12: State-Dependent Effects of Austerity Measures on the Debt-to-GDP Ratio with Evolving Regimes



The figure displays the evolving-regime responses of the debt-to-GDP ratio to a negative government spending shock (first row) and a positive tax shock (second row) for the benchmark model in a deep recession (left column) and in a robust expansion (middle column). The right column plots the difference between the responses in a deep recession and a robust expansion. The shocks are equal to 1% of GDP. Posterior medians and 90% credibility intervals are reported.

Even though there is a substantial amount of uncertainty associated with the responses of the debt-to-GDP ratio to cuts in government spending, the posterior differences indicate that government spending cuts implemented in a robust expansion decrease the debt-to-GDP ratio significantly more than government spending cuts implemented in a deep recession. The posterior difference peaks at 0.5% of GDP at medium horizons. Similarly, increases in taxes reduce the debt-to-GDP ratio more if implemented in a robust expansion.

Our results indicate that there is strong state dependence in the behaviour of the debt-to-GDP ratio in response to discretionary fiscal policy. Furthermore, in deep recessions, both government spending and taxes can be used as tools to stimulate the economy. When the economy is close to potential, 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 reduce the debt-to-GDP ratio significantly less over the medium horizon if implemented in periods of excess slack due to a sharp decrease in output. The implication, then, is that austerity measures are more effective in reducing the debt-to-GDP ratio when the economy is close to potential.

4.3 Disaggregated Responses of Consumption and Investment

In this subsection, we disaggregate the responses of output by considering the responses of consumption and investment to government spending and tax shocks separately. The responses are calculated following the same approach used to calculate the responses of output, but the impact responses of consumption and investment to fiscal shocks are left unrestricted, as there is a less clear consensus on the underlying source of the overall response of output than about the response itself. Figure 13 plots the fixed-regime responses of consumption, Figures 14 and 15 plot the responses of consumption to positive and negative shocks in government spending and taxes, and Figure 16 plots the fixed-regime responses of investment.

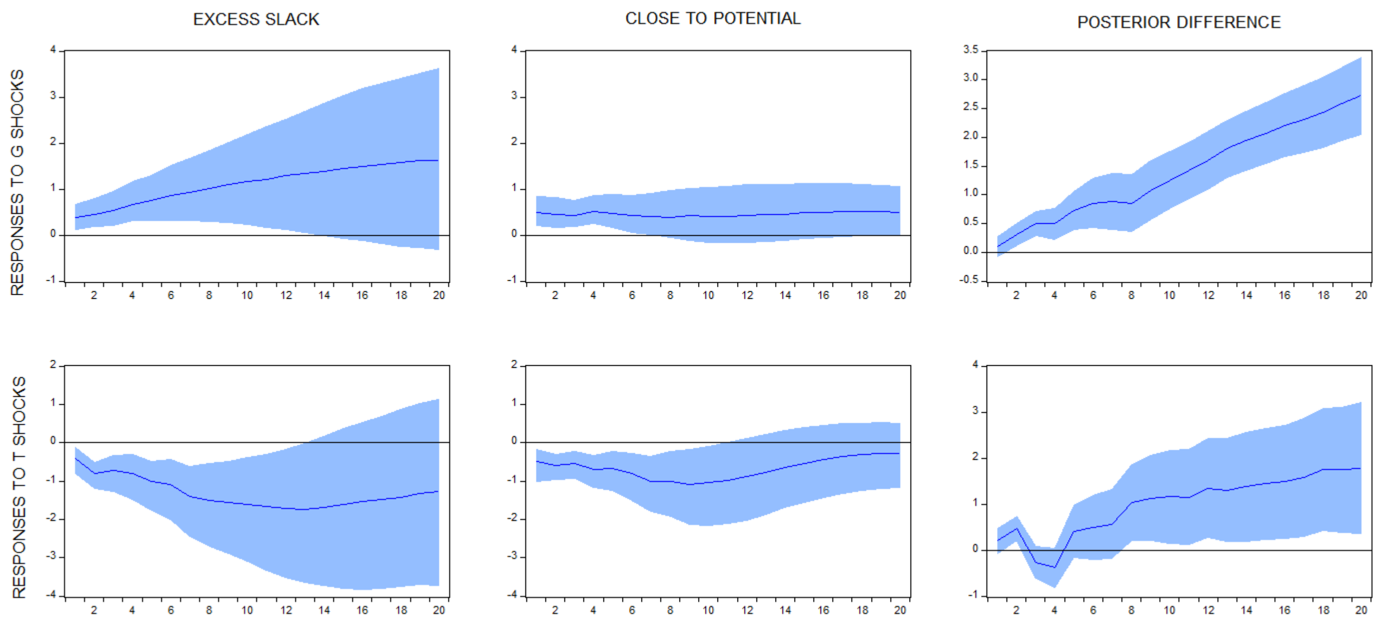
From Figure 13, there is strong evidence in favor of state dependence in the response of consumption to government spending shocks, with implication that this state dependence drives much of the state dependence in the response of output documented earlier. Consumption responds strongly in the excess slack regime. Although it also responds significantly in the regime when the economy is closer to potential, the response is much smaller in magnitude, dollar for dollar. Similarly, the response of consumption to tax shocks is stronger in the excess slack regime. The response pattern for government spending shocks is consistent with the findings in FMP for a smaller model with timing restrictions and it is also consistent with, for example, theoretical models that incorporate a time-varying share of rule-of-thumb consumers, models featuring habit formation in which government spending and consumption are complements, as in Leeper, Traum, and Walker (2017), and the empirical findings of Klein (2016) that imply 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 government spending shocks. However, there is evidence that investment responds very differently to government spending and tax shocks, dollar for dollar, and some evidence of state dependence in the responses of investment to tax shocks.[23] Investment does not increase significantly in response to government spending shocks, but it responds significantly to tax changes. The earlier reported response of output to government spending shocks in the excess slack regime is therefore apparently driven largely by the response of consumption. By contrast, investment responds to tax changes in both regimes, suggesting that the responses of output to discretionary changes in taxes is driven by both the responses of consumption and investment.

In summary, our results indicate that, if the aim of stimulative fiscal policy is to increase consumption, government spending increases and tax cuts have multipliers that are not significantly different in periods of excess slack, but government spending increases lead to smaller medium-term increases in the debt-to-GDP ratio. Meanwhile, if the aim is to increase investment, tax cuts work uniformly better.

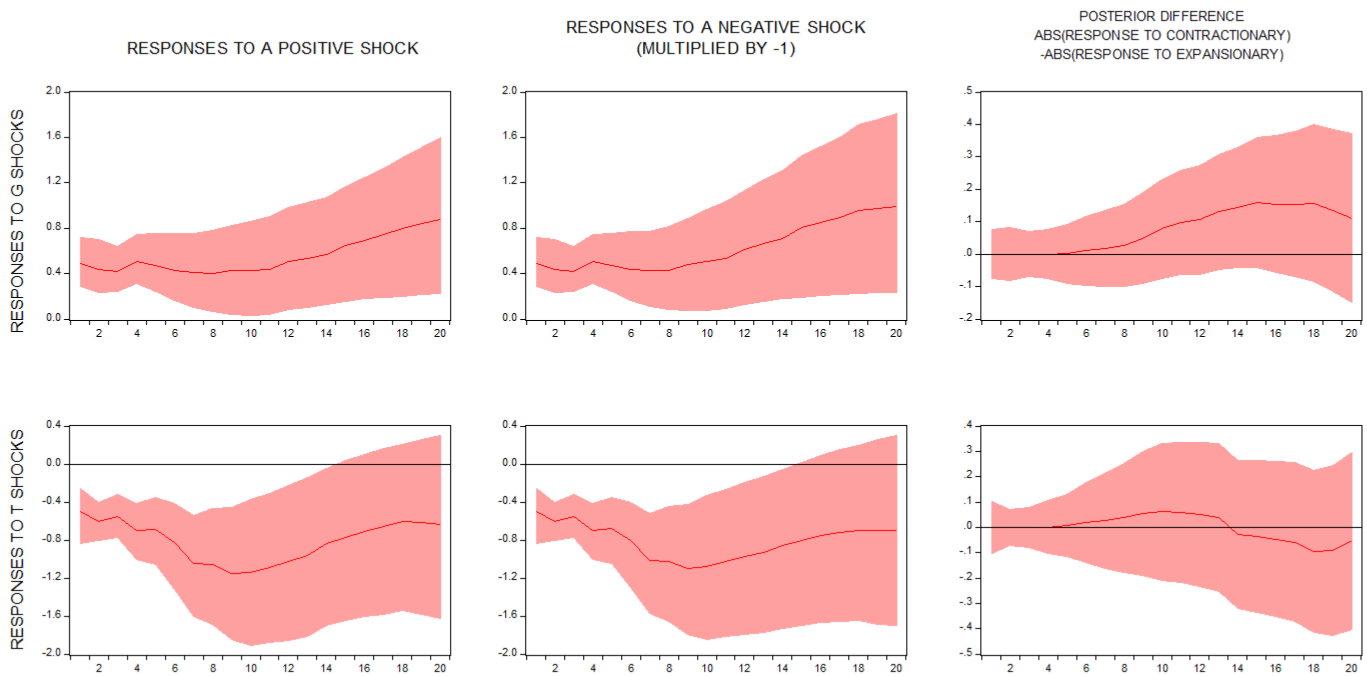
[23] Not surprisingly given the fixed-regime responses, the evolving-regime responses of investment do not exhibit much evidence in favor of sign or size dependence and are not reported to conserve space.

Figure 13: State-Dependent Effects of Government Spending and Taxes on Consumption



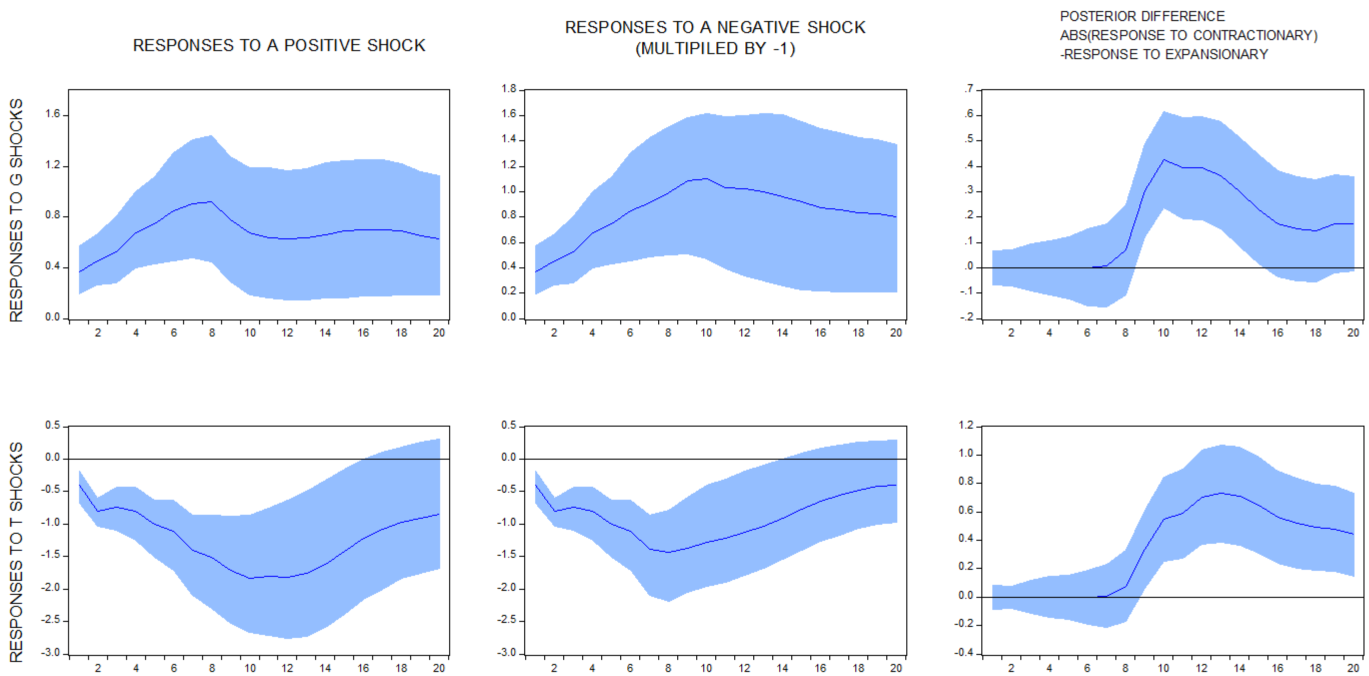
The figure displays the fixed-regime responses of consumption to a government spending shock (first row) and a tax shock (second row) in the excess slack regime (left column) and close to potential regime (middle column) for the benchmark model with consumption instead of output. The right column plots the posterior differences in responses across the two regimes. Posterior medians and 90% credibility intervals are reported.

Figure 14: Sign-Dependent Effects of Government Spending and Taxes on Consumption in a Robust Expansion



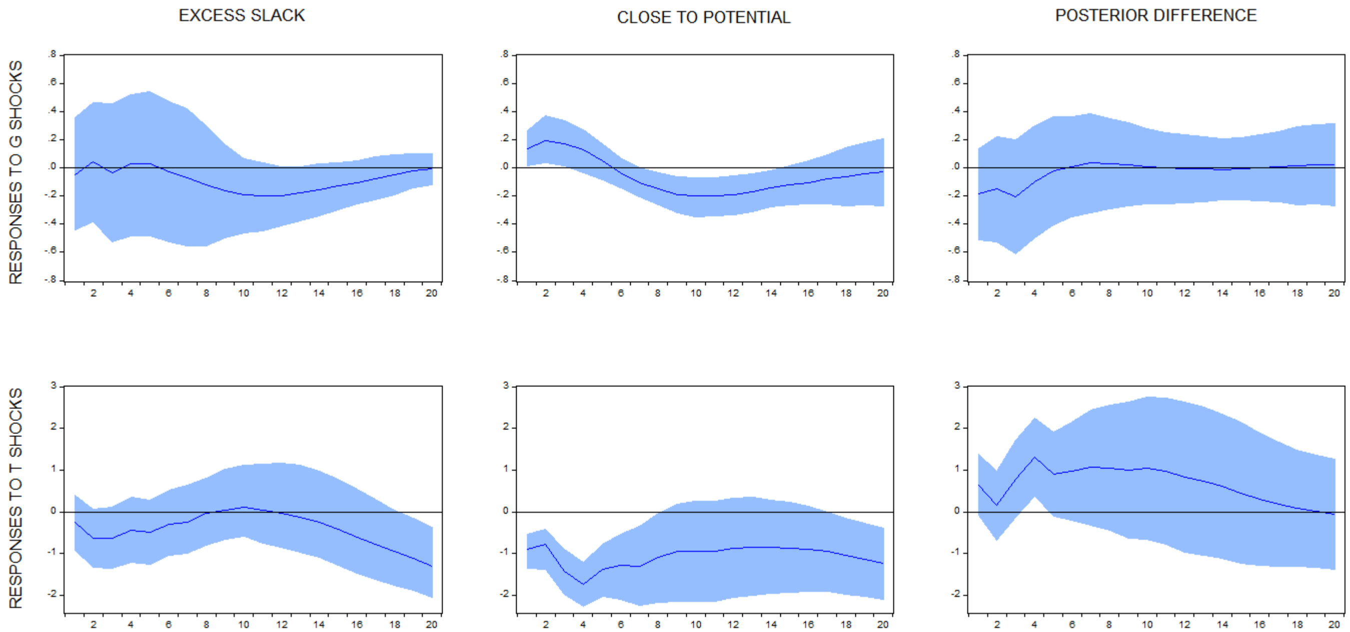
The figure displays the evolving-regime responses of consumption to a government spending shock (first row) and a tax shock (second row) for the benchmark model with consumption instead of output in a robust expansion. The left columns plot the responses to a positive shock, the middle column plots the response to a negative shock (scaled by -1 for ease of comparison), and the right column plots the difference in magnitude of (scaled) responses for positive and negative shocks. The shocks are equal to 1% of GDP. Posterior medians and 90% credibility intervals are reported.

Figure 15: Sign-Dependent Effects of Government Spending and Taxes on Consumption in a Deep Recession



The figure displays the evolving-regime responses of consumption to a government spending shock (first row) and a tax shock (second row) for the benchmark model with consumption instead of output in a deep recession. The left columns plot the responses to a positive shock, the middle column plots the scaled (by -1 for ease of comparison) response to a negative shock, and the right column plots the difference in magnitude of the scaled responses for positive and negative shocks. The shocks are equal to 1% of GDP. Posterior medians and 90% credibility intervals are reported.

Figure 16: State-Dependent Effects of Government Spending and Taxes on Investment



The figure displays the fixed-regime responses of investment to a government spending shock (first row) and a tax shock (second row) in the excess slack regime (left column) and close to potential regime (middle column) for the benchmark model with investment instead of output. The right column plots the posterior differences in responses across the two regimes. Posterior medians and 90% credibility intervals are reported.

5 Conclusion

In this paper, we have examined time dependencies in the responses of output to discretionary changes in government spending and taxes. We have presented strong empirical evidence in favor of nonlinearity and state-dependent effects of fiscal policy. In particular, the estimates from a threshold structural vector autoregressive model identify different responses of the economy both to government spending and to taxes that depend on the degree of economic slack. Government spending increases and tax cuts have larger and more persistent effects on output when there is excess slack in the economy than when the economy is close to potential. Meanwhile, tax cuts always have persistent stimulative effects.

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 a robust recovery, tax increases have significant contractionary effects, decreasing output, consumption, and investment, while tax cuts have stimulative effects. Our results indicate that, if the aim of discretionary fiscal policy is to stimulate the economy during periods of excess slack, both government spending multipliers and tax multipliers are high and work primarily through a consumption channel. When the economy is closer to potential, tax cuts have larger effects than government spending increases. Austerity measures are largely self-defeating in periods of slack. If the aim of austerity measures is to reduce the debt-to-GDP ratio, our results suggest that austerity measures have smaller negative effects on economic activity if pursued when the economy is close to potential or in a robust expansion.

References

[1] Arias, J. E., J. F. Rubio-Ramirez, and D. F. Waggoner (2016). "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 Papers and Proceedings 103(3), pp. 141-146.

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

[6] Barnichon, R. and C. Matthes (2017). "Stimulus versus Austerity: The Asymmetric Government Spending Multiplier." CEPR Discussion Papers 10584.

[7] 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.

[8] 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.

[9] Beveridge, S. and C. R. Nelson (1981). "A New Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle." Journal of Monetary Economics 7, pp. 151-174.

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

[11] 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.

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

[13] 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.

[14] 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.

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

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

[17] 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.

[18] 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.

[19] Fotiou, A. (2016) “Nonlinearities and Fiscal Policy.” Bocconi Univeristy working paper.

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

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

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

[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] Kim C. J. (1994). "Dynamic Linear Models with Markov Switching" Journal of Econometrics 60, pp. 1-22.

[28] Klein, M. (2016). "Austerity and Private Debt." Journal of Money, Credit and Banking (forthcoming).

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

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

[31] Leeper, E. M., N. Traum, and T. B. Walker (2017)." Clearing Up the Fiscal Multiplier Morass." (forthcoming) American Economic Review.

[32] 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.

[33] 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.

[34] Morley, J. and J. Piger (2008). "Trend/Cycle Decomposition of Regime-Switching Processes" Journal of Econometrics 146, pp. 220-226. 208-221.

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

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

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

[38] 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.

[39] 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.

[40] 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.

[41] 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.

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

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

[44] 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.

[45] 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.

[46] 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.

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

[48] 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 Response Calculation

For the 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_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.

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, Rubio-Ramirez, and Waggoner (2016). 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 $Q * chol(\Sigma)$ 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%. We used 200,000 MH iterations.

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 centered around the parameters for the linear model. In particular, $[vec(\Phi_0^2)'\ vec(\Phi_1^2)']'$ is centered around zero, implying no nonlinearity. The variance for Φ is $100 * I_{n_1}$. The prior for the variance matrix Σ is an inverse Wishart distribution centered around the variance-covariance matrix for the linear model 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.To obtain the average response for a subset of histories, repeat step 2 for each history and report the distribution averaged across histories;

3. Repeat Steps 1 and 2 for each draw of the sampler.

Appendix B: Supplemental Table

Table B1: Model Selection Criteria for Linear VAR and TVAR Models of Consumption and Investment

	MAOG as Measure of Slack in the TVAR			
	Consumption		*Investment*	
Linear model	−2395.05 −2393.30 −617.81		−2718.36 −2715.42 −873.77	
TVAR	**−1932.24** **−1930.87** **−333.80**	−0.54 (−0.74, −0.36)	**−2221.51** **−2214.82** **−400.64**	−0.71 (−0.83, −0.54)

Each cell reports the log likelihood obtained using maximum likelihood estimation, the expected posterior log likelihood obtained Bayesian estimation, and the log marginal likelihood (Top, Middle, Bottom). The second entry is the threshold estimate, including 90% credibility intervals, obtained from the posterior Bayesian distribution. The best model fit for each measure of slack is reported in bold.

Appendix C: Supplemental Figures

Figure C1: Responses of Output to Government Spending and Tax Shocks for the Linear VAR Model

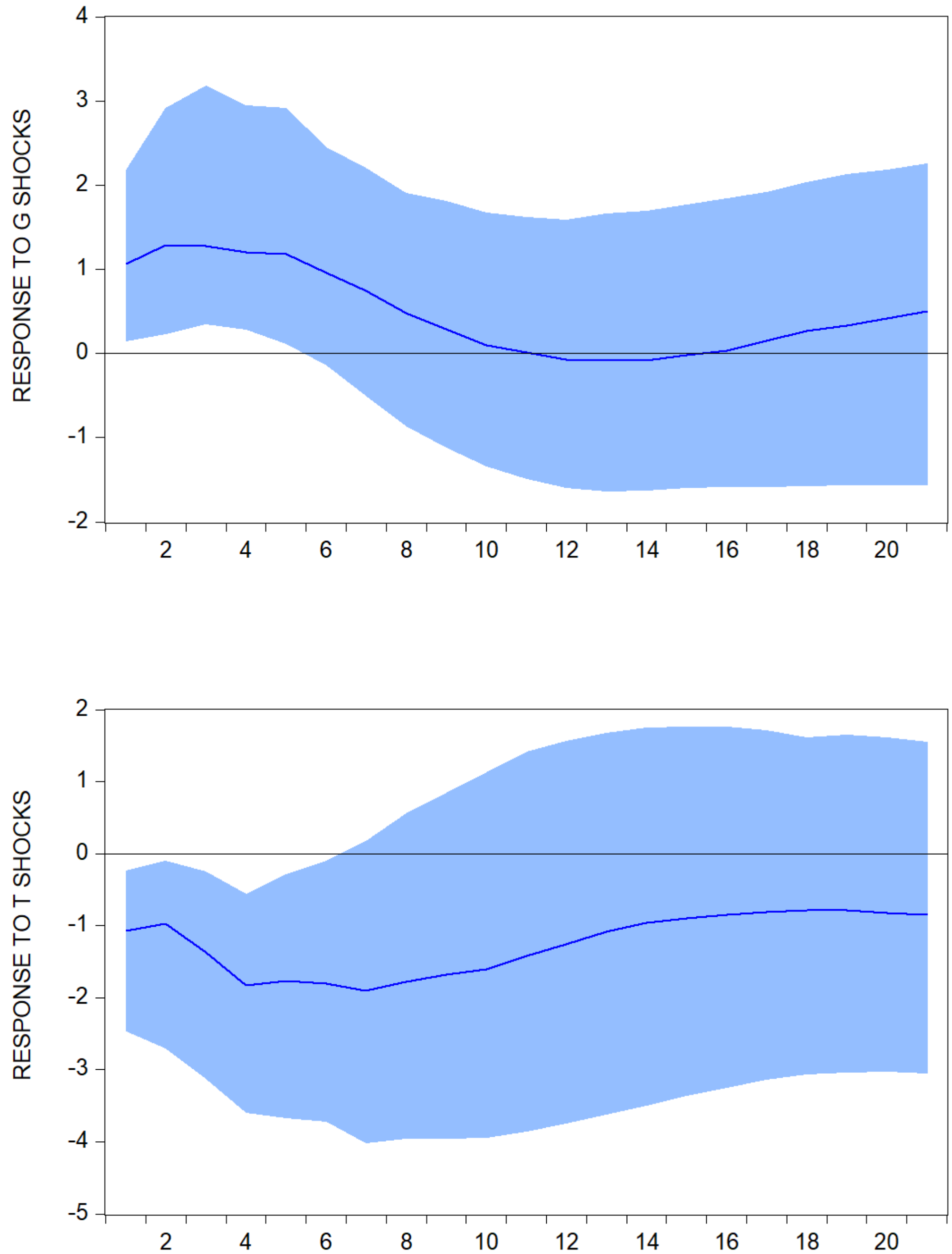


Figure C2: Tax Shocks Identified from the Benchmark TVAR and Romer and Romer's (2010) Narrative Approach

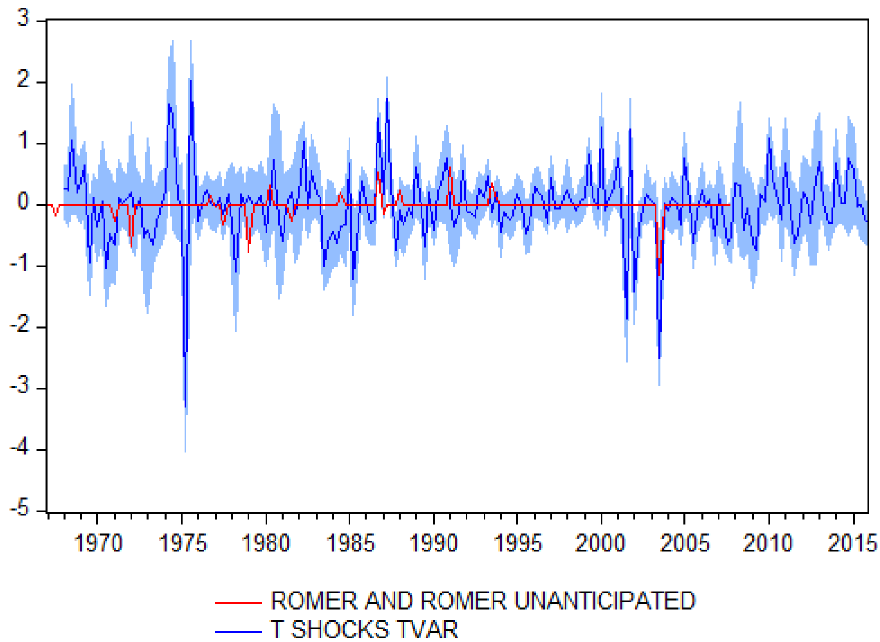
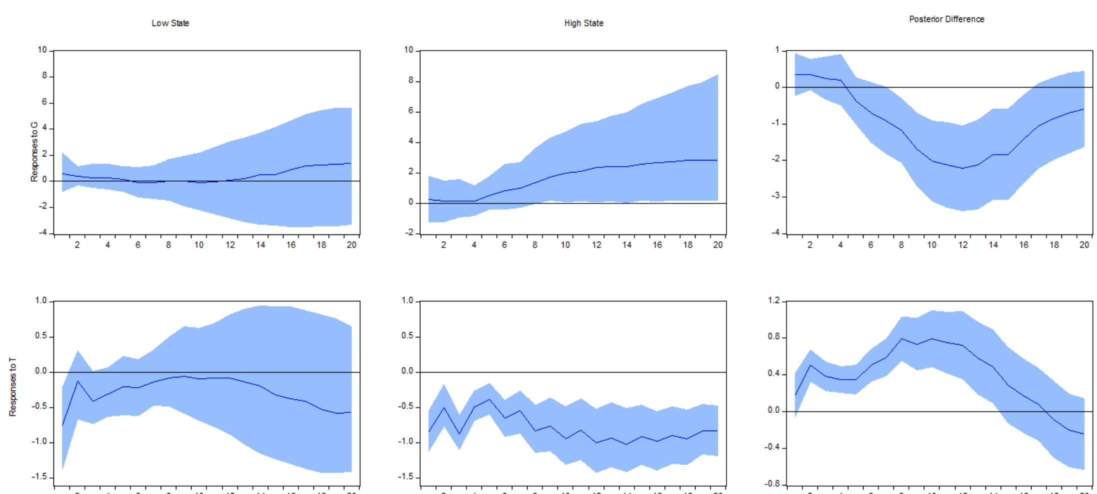
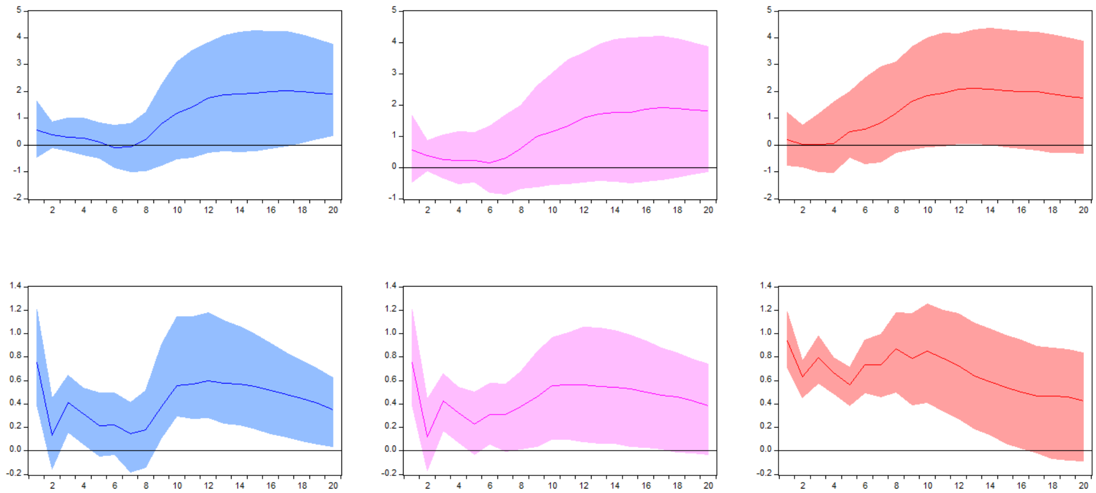


Figure C3: State-Dependent Effects of Government Spending and Taxes on the Budget Deficit



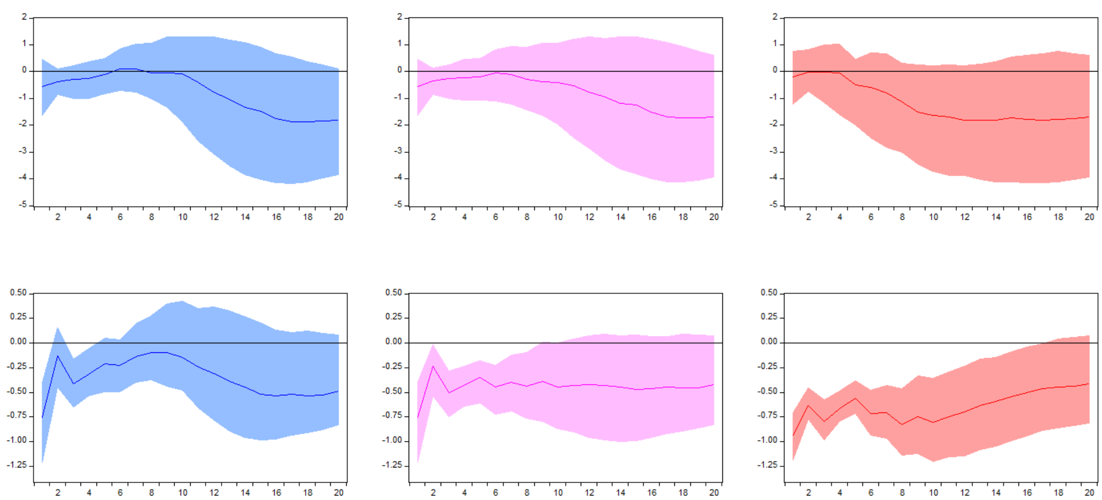
Responses of Deficits to G (top) and T (bottom). Fixed Excess Slack Regime (Left), Fixed Close to Potential (M) Regimes, and posterior differences (R) with 90% credibility intervals.

Figure C4: State-Dependent Effects of Fiscal Stimulus on the Budget Deficit with Evolving Regimes



Responses of Deficits to increases in G (top) and decreases in T (bottom). Deep recession (Left), sluggish recovery (M), and robust expansion (R) with 90% credibility intervals.

Figure C5: State-Dependent Effects of Austerity Measures on the Budget Deficit with Evolving Regimes



Responses of Deficits to decreases in G (top) and increases in T (bottom). Deep recession (Left), sluggish recovery (M), and robust expansion (R) with 90% credibility intervals.