

# State-Dependent Effects of Fiscal Policy

Steven M. Fazzari<sup>†</sup>    James Morley<sup>††</sup>    Irina Panovska<sup>†\*</sup>

May 15, 2012

## Abstract

We investigate the effects of government spending on U.S. economic activity using a threshold version of a structural vector autoregressive model. Our empirical findings support state-dependent effects of fiscal policy. In particular, the effects of a government spending shock on output are significantly larger and more persistent when the economy has a high degree of underutilized resources than when the economy is close to capacity. This evidence is consistent with an underlying structure of the economy in which insufficient aggregate demand often constrains the level of economic activity.

JEL Codes: C32, E32, E62

Keywords: Government Spending, Threshold Model, Vector Autoregression, Nonlinear Dynamics

\*Corresponding author: [ipanovsk@wustl.edu](mailto:ipanovsk@wustl.edu). †Washington University in St. Louis; ††University of New South Wales. We thank Alan Auerbach, Rudi Bachmann, Lutz Kilian, Raul Santaaulalia-Llopis, Houston Stokes, John Taylor, Pao-Lin Tien, and the participants at the 2012 SNDE Meetings, the 2012 CAMA Macro Workshop, the 2011 EC2 Conference, the 2011 SCE Meetings, the 2011 Midwestern Econometrics Group Meetings, and the 2010 Washington University Graduate Student Conference for helpful comments. All errors are our own. We also thank Valerie Ramey for providing the narrative military spending data set. Fazzari and Panovska acknowledge the generous support of The Institute for New Economic Thinking. Panovska also gratefully acknowledges the support of the Graduate School of Arts and Sciences and the Olin Fellowship program.

# 1 Introduction

The Great Recession and the subsequent American Recovery and Reinvestment Act fiscal stimulus package of more than \$700 billion dollars has reignited debate, academic and otherwise, about the stabilization role of discretionary fiscal policy. More broadly, these developments have raised questions about the relevance of aggregate demand and government spending as possible engines of economic activity. As a way to answer these questions, many recent academic studies have sought to determine whether government spending has significant effects on aggregate output and components of output, such as consumption and investment.

This debate is of central importance not only for economic policy, but also for the insights it provides into the underlying structure of modern developed economies. Theoretical models in which resources are fully employed predict that the direct effect of a positive shock to government spending, given preferences and technology, should completely crowd out private economic activity. The direct government spending multiplier arising from such models is zero, at least as a first approximation.<sup>1</sup> In contrast, traditional Keynesian models predict that the economy will not always fully employ available resources, possibly for extended periods of time, because of insufficient demand. If output is below its potential level due to insufficient aggregate demand, an increase in government spending can directly motivate the employment of idle resources and raise output.

Much of the recent empirical research on fiscal policy considers the effects of spending shocks on different components of output. On the one hand, a baseline neoclassical model predicts crowding out of both consumption and investment, and therefore implies negative responses of these variable to a positive shock to government spending. On the other hand, if government spending raises resource use through traditional Keynesian channels, consumption and investment should respond positively to spending shocks. Indeed, these spillovers create the possibility that the government spending multiplier exceeds one because higher government spending induces increases in other components of demand.

Most of the existing empirical research on fiscal policy employs linear time series models in which the size of the response of output or other variables to government spending is independent of the state of the economy (important recent exceptions include Mittnik and Semmler, 2011, and Auerbach and Gorodnichenko, 2012a; see the following section). The results from such models are useful, especially if the maintained null hypothesis is the neoclassical baseline of zero effects on output and crowding out of consumption and investment. But the Keynesian alternative suggests an important state dependence (and therefore a nonlinearity) in the effect of any demand shock on output, including a

---

<sup>1</sup>These models can generate indirect allocational effects of government spending on output, but of ambiguous sign. For example, the higher interest rate or negative wealth effect (see, for example, Parker, 2011) induced by a rise in government spending could encourage higher labor supply that raises output, but higher interest rates also reduce capital accumulation that lowers output.

government spending shock. Higher demand cannot raise output indefinitely. Eventually, resource constraints bind: even a Keynesian economy behaves like a neoclassical system if demand is sufficiently high. Threshold models provide a natural econometric framework for exploring this basic state dependence. If government spending shocks affect output through Keynesian demand channels, we expect such effects to be larger when the economy has significant resource slack than when it is operating at or near full capacity. The purpose of this paper, then, is to test this simple, but fundamentally important hypothesis.

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

Our empirical results provide strong evidence in favor of state-dependent nonlinearity; specifically, government spending shocks have larger effects on output when they occur with relatively low resource utilization than when they appear at times of high resource use. Furthermore, threshold estimates for capacity utilization place half or more of its historical observations in the low-utilization regime. This evidence implies, therefore, that the “normal” state of the U.S. economy is one in which positive demand shocks have large positive and persistent effects on output and its components. We also employ generalized impulse response functions to isolate the different effects of fiscal policy under particular economic conditions. We find that the responses of output and output components depend crucially on the state of the economy when a policy shock occurs.

The rest of the paper is organized as follows. Section 2 reviews previous research that has estimated the aggregate effects of fiscal policy in a time-series context. Section 3 introduces the baseline empirical model and the estimation method. Section 4 presents the empirical results and extends the baseline model to models that include consumption, investment, and other variables of interest. Section 5 concludes.

## 2 Related Literature

The empirical literature that explores the effects of government spending on macroeconomic variables, both old and new, is divided in its findings. Most studies fall in one of four main strands: models based on traditional Keynesian theory, structural vector autoregressive (SVAR) models, dynamic stochastic general equilibrium (DSGE) models, and models based on the narrative approach introduced by Ramey and Shapiro (1998).

Traditional Keynesian models usually relate an outcome variable such as aggregate output to different components of spending or taxes, typically with

a reduced-form, linear specification. The interest rate is usually held fixed over the whole forecasting horizon, and the multipliers obtained from those models are often very large (always greater than 1, sometimes as big as 4). In particular, the American Recovery and Reinvestment Act (ARRA) fiscal stimulus package was designed based on a study of this kind by Romer and Bernstein (2009) that estimated a short-run tax multiplier for output of approximately 1.6.

Studies based on SVAR models in which government spending is assumed to be predetermined typically find that output, consumption, and real wages increase after a positive government spending shock. Blanchard and Perotti (2002) and Perotti (2008) find that the response of output and consumption to government spending is positive and persistent, although, perhaps surprisingly, they find a negative response of investment. This discrepancy between the positive response of consumption (implied by Keynesian models), and the negative response of investment (implied by neoclassical models) is commonly referred to as the “investment puzzle” in the fiscal policy literature. The magnitude of the estimated effects depends on the identification of the model. Blanchard and Perotti (2002) and Perotti (2008) use institutional information to identify the shocks, and they get government spending multipliers for output that are about 1.3. Mountford and Uhlig (2005) use an alternative approach based on sign restrictions, and they get a smaller government spending multiplier for output of 0.5 and a multiplier for consumption that is very close to zero. These SVAR studies are sometimes criticized because they do not allow for state-dependent responses (see Parker, 2011, in particular), an issue addressed here.

Most DSGE studies are based on a New Keynesian model with Calvo pricing frictions and make a variety of assumptions about whether interest rates adjust to an increase in government spending. Pappa (2009) uses an DSGE model with Calvo pricing and a fixed interest rate. She finds that output, real wages, and consumption rise, while investment falls, but the magnitude of the responses is very sensitive to the parameterization of the model. Some DSGE models allow for different responses to government spending shocks depending on the interest rate. For example, Cogan et al. (2010) consider the importance of interest rate responses to fiscal policy by holding the interest rate pegged at zero for four quarters, but afterwards allowing the interest rate to revert to the “natural level” determined by neoclassical first-order conditions. The estimated government spending multiplier for output from this model is only 0.4, and their simulations yield multipliers that are above one only when the interest rate is fixed exogenously.<sup>2</sup>

Van Brusselen (2009), Leigh et al. (2010), and Ramey (2011a) provide extensive surveys of the empirical literature on fiscal multipliers. Consistent with the summary above, they show that the estimated government spending multipliers are highly sensitive to the model and parameters. In particular, Van Brusselen (2009) compares a wide variety of empirical models and points out that, even within the same class of models (i.e., DSGE models with Calvo pric-

---

<sup>2</sup>Although Barro and Redlick (2011) use a simple linear regression model with annual data instead of a DSGE model, they find that the government spending multiplier for output is just 0.3 on impact, similar to the DSGE results.

ing), the government spending multiplier for output varies between -3.7 and 3.7, depending on how the increase in spending is financed, how long the interest rate is pegged, and whether the economy is closed or open. Ramey (2011a) also points out the sensitivity of estimates, but concludes that the multiplier for the U.S. economy likely lies within a range between 0.8 and 1.5. Leeper et al. (2012) examine the variability of multipliers for output and consumption in DSGE models in great detail, and they conclude that the estimates are very sensitive to the specification of the model. In particular, when the proportion of rule-of-thumb consumers is large, as in Galí et al. (2007), the multipliers are large and positive, but they are smaller when the proportion of rule-of-thumb consumers is closer to zero.

Another strand of the literature uses a narrative approach to identify exogenous government spending shocks. Ramey and Shapiro (1998), Eichenbaum et al. (1999), and Ramey (2011b) find that the response of output is small and short-lived, but they use military spending rather than total government spending. Ramey (2011b) argues that the SVAR-based multipliers are large only because they fail to capture the importance of timing of government shocks and because the combined narrative data Granger-causes the government spending shocks. However, Perotti (2007) shows that lagged government spending, tax, and GDP shocks also predict the Ramey-Shapiro narrative dates.

In this paper, we adopt a nonlinear SVAR-based approach that allows the impact of shocks to depend on the level of resource utilization at the time of the shock. We do not impose any assumptions about the response of interest rates or about the degree of price or wage stickiness, and we check the robustness of our results to including military spending separately in the model. Our approach nests the possibility of empirical results consistent with neoclassical models that predict small or even negative responses of output and other macro variables to positive government spending shocks. But to the extent that the data identify Keynesian effects, our model allows these effects to vary with the state of the economy that prevails at the time of the shock; that is, we test whether positive responses are stronger when there is substantial slack in the economy. The size and difference in the multipliers across regimes is of fundamental interest, and we also consider what the data tell us about the amount of time the economy spends in the different regimes.

This nonlinear approach is strongly suggested by Parker (2011) and it is similar in motivation to recent studies by Auerbach and Gorodnichenko (2012a) and Mittnik and Semmler (2011). Auerbach and Gorodnichenko (2012a) estimate a smooth transition threshold SVAR model for government spending, taxes, and output, in which they impose the restrictions that government spending has different effects during recessions and expansions, and they calibrate the smoothness parameter based on U.S. data so that the economy spends about 20% of the time in recessions. They estimate that the effects of government spending are large and positive (1.7 over 20 quarters) when the economy is in a recession and smaller (very close to one) when the economy is not in a re-

cession.<sup>3</sup> They control for the state of the business cycle by using a moving average of output growth as the threshold variable, and they impose that the threshold around which the behavior changes is equal to the mean of output growth.<sup>4</sup> Mittnik and Semmler (2011) estimate a bivariate threshold model for output and employment where the threshold output is lagged output growth and threshold is predetermined and equal to the mean of output growth. In their model, the responses of employment to output shocks are much larger in the low regime than in the high regime.

Our analysis differs from these two other nonlinear studies of fiscal policy and aggregate demand in a number of important ways. First, we estimate the threshold that determines state-dependent effects from the data, whereas Auerbach and Gorodnichenko (2012a) and Mittnik and Semmler (2011) both assume their thresholds *a priori*. Based on our estimated thresholds, we find evidence that the U.S. economy spends the majority of its time in the low-utilization state, a possibility not allowed for by the other two studies. Second, we consider a threshold SVAR model with a discrete change in regime instead of the smooth transition specification considered by Auerbach and Gorodnichenko (2012a). Although the smooth transition specification is potentially more general, estimating the smoothness parameter for such a model can be challenging, as evidenced by the fact that Auerbach and Gorodnichenko (2012a) fix this parameter (as well as the threshold) in their estimation. The difficulty is that the likelihood function for a smooth transition model is flat when the true smoothness parameter is large in the sense of implying a relatively discrete threshold, making maximum likelihood estimation and even Bayesian estimation unreliable. We circumvent this econometric problem by considering a discrete threshold only, which still allows us to focus on the primary question of whether there are state-dependent effects of fiscal policy. Third, we investigate the role of capacity constraints in generating potential state-dependent effects by considering various measures of economic slack as threshold variables, rather than just the growth rate of output, which was the focus of the other two studies. We also include each measure of slack as an endogenous variable in the SVAR model to allow the possibility that the variable is important for understanding the effects of fiscal policy, but does not necessarily relate to a state-dependent effect. Fourth, unlike the other two studies, we formally test for nonlinearity by comparing nonlinear models to

---

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

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

their linear counterparts using Bayesian marginal likelihood analysis.

## 3 Empirical Methods

### 3.1 Model

The basic vector autoregression (VAR) model is linear, and cannot capture nonlinear dynamics such as regime switching and asymmetric responses to shocks. For our analysis, we consider a nonlinear version of a VAR model that extends the threshold autoregressive model of Tong (1978, 1983) to a multivariate setting. Threshold models work by splitting a time series process endogenously into different regimes. Within each regime the process is described by a linear model. Specifically, we specify a threshold version of a reduced-form VAR model as follows:

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

where  $Y_t$  is a vector containing the first difference of the logarithm of real government spending, the first difference of the logarithm of net taxes, the first difference of the logarithm of real GDP, and an mean-adjusted measure of capacity utilization, as discussed in more detail below. This is the baseline version of our model. However, we also consider alternative versions of the model that incorporate the private-sector components (i.e., consumption, investment, exports, and imports) and other outcome variables of interest, again discussed in more detail below.

The lag polynomial matrices  $\Phi_1^1(L)$  and  $\Phi_1^2(L)$  capture the dynamics of the system. The disturbances  $\varepsilon_t$  are assumed to be independent and Gaussian with mean zero. Rather than assuming that the disturbances are strictly i.i.d., we set the covariance matrix of  $\varepsilon_t$  equal to  $\Omega$  until 1984Q1 and equal to  $\lambda\Omega$  afterwards to capture the Great Moderation. Because the focus of this paper is not on determining the exact break date in volatility and because there is near consensus in the literature about the general timing of the volatility break (see, for example Kim and Nelson, 1999, or McConnell and Perez-Quiros, 2000), we set the break date exogenously. By using a scale factor  $\lambda$  and a constant variance-covariance matrix  $\Omega$ , we are also assuming that the correlations between the endogenous variables do not change over time or over regimes. The threshold variable  $q_{t-d}$  determines the prevailing regime;  $\gamma$  is the threshold parameter at which the regime switch occurs. The indicator function  $I[\cdot]$  equals 1 when the  $q_{t-d}$  exceeds the threshold  $\gamma$  and 0 otherwise. The integer  $d$  is the delay lag for the threshold switch; that is, if the threshold variable  $q_{t-d}$  crosses  $\gamma$  at time  $t-d$ , the dynamics actually change at time  $t$ . For the threshold variable, we consider capacity utilization, other measures of economic slack, and a selection of other macroeconomic variables, as discussed in more detail below.

## 3.2 Data

In addition to capacity utilization, we also consider the output gap estimated by the CBO, the unemployment rate, output growth, and employment growth to measure economic slack. The traditional Keynesian theory summarized above implies that the threshold variable should measure the level of economic activity and intensity of resource use. For this purpose, the output gap, the level of capacity utilization or the unemployment rate would seem to be good choices. However, we also consider first differences of these variables and output and employment growth to check the robustness of the results and to explore whether threshold effects might relate to growth (as in Auerbach and Gorodnichenko, 2012a, and Mitnik and Semmler, 2011) rather than to levels.

Government spending and net taxes are defined as in Blanchard and Perotti (2002). The full sample period is 1967Q1-2011Q1. All output components are measured in real terms and are seasonally adjusted by the source. The series for output, its components, including government spending, and tax revenues were obtained from NIPA-BEA, and the capacity utilization series was obtained from the Federal Reserve Statistical Releases website. We also consider data for U.S. federal government debt, the Federal Funds Rate, inflation based on the CPI (seasonally adjusted), and non-farm payroll employment, which were all obtained from the Federal Reserve Bank of St. Louis FRED website. The monthly series for capacity utilization, the unemployment rate, the Federal Funds Rate, CPI, and employment are all converted to a quarterly frequency by using simple arithmetic means.

We use growth rates rather than log-levels in the VAR because the logarithms of real GDP and output components exhibit nonstationarity. Johansen cointegration tests suggest the absence of cointegrating relationship between government spending and taxes, between government spending, taxes, and output, as well as between government spending, taxes, and output components (consumption, investment, exports, and imports). However, we obtained roughly similar results when we allowed spending and taxes to be cointegrated.

## 3.3 Specification Issues

The lag length for the VAR model is chosen based on AIC (for the baseline linear VAR model, estimated using maximum likelihood as a starting point). Both AIC and SIC select four lags, which is also the number of lags used by Blanchard and Perotti (2002), Ramey (2011b), and most other studies that use the linear SVAR approach. Unlike Mitnik and Semmler (2011), who allow the number of lags to differ across regimes, we assume that the number of lags is the same in each regime,<sup>5</sup> and we consider a model with two regimes.

---

<sup>5</sup>It is certainly possible to extend the nonlinear SVAR model to accommodate more than two discrete regimes, or even an infinite number of regimes (by using a smooth transition model). However, it would make computation very burdensome and possibly imprecise, both because of the larger number of parameters that would need to be estimated and because of the identification issues for smooth transition models discussed in the previous section.



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

Economic theory implies several possible choices for the threshold variable. As discussed above, traditional Keynesian theory suggest that the dynamics may depend on the state of the economy, while DSGE models imply that the effects of government spending depend on the interest rate. A recent literature also suggests that the dynamics may depend on debt (see, for example, Reinhart and Rogoff, 2009, and Eggertsson and Krugman, 2010, for two very different views on the impact of debt on the efficacy of fiscal policy). Because we do not want to fix the threshold variable *a priori*, we consider a large set of possible threshold variables and select the preferred threshold variable using Bayesian model comparison. The threshold variables that we consider are

1. lagged output: output growth, long differences in the natural log of output, moving averages of differences in the natural log of output<sup>6</sup>
2. lagged CBO output gap
3. lagged capacity utilization: level, level adjusted for long-run change in mean, first differences, and first differences of the mean-adjusted series
4. lagged unemployment rate: level, differences, mean-adjusted level, differences in the mean-adjusted series
5. lagged debt-to-GDP ratio: total Federal debt and total Federal debt held by the public, both as a percent of GDP
6. lagged real interest rate: level and change in the ex ante real interest rate based on the Federal Funds Rate and CPI inflation under the assumption of static expectations

Both capacity utilization and the unemployment rate appear to have changes in their long-run mean levels, which would make those series unsuitable for use in a stationary VAR model. Standard tests for a structural break at an unknown breakdate reject the null of no break in mean for both capacity utilization and the unemployment rate. Meanwhile, there is some debate about whether the

---

<sup>6</sup>Because we were already estimating a large number of parameters, the weights for the moving averages were fixed exogenously. We considered an arithmetic mean of the past 4 differences, and  $q_{l,t-d} = \frac{1}{l-d+1} \sum_{j=d}^l threshold\_var_{t-j}$  for  $l = 1, d = 4$ .

unemployment rate has a unit root or whether there were just exogenous structural breaks in its mean (see, for example, Papell et al., 2000). For both series, therefore, we consider the level, first differences, the mean-adjusted levels, and the differences of the mean-adjusted levels as possible threshold variables. Table 1 summarizes the results of the test for structural breaks in mean for capacity utilization and the unemployment rate. A structural break test for capacity utilization identifies a highly significant break (F statistic of 41.7) in the level of capacity utilization in 1974Q1, which coincides with the well-known productivity slowdown. The structural break tests also identify three breaks in mean for the unemployment rate. Notably, the mean-adjusted capacity utilization series is strongly correlated with other commonly-used measures of economic activity, as shown in Figure 1, so it appears to be a highly representative measure of economic slack.<sup>7</sup>

### 3.4 Estimation and Inference

Because the threshold VAR model is highly parameterized, we make inferences about the threshold, the coefficients, the threshold variable, and the delay parameter using Bayesian methods; in particular, we use a multi-block Metropolis-Hastings (MH) algorithm described in detail in the appendix to sample from marginal posterior distributions for parameters and calculate marginal likelihoods for models. The advantages of using a Bayesian approach in this setting are twofold. First, it allows us to capture the uncertainty about the parameter values when constructing the impulse response functions. Second, despite the presence of nuisance parameters in the nonlinear models, comparing the linear to the nonlinear model and examining the presence of nonlinear effects is straightforward in the Bayesian framework.

To provide an accurate approximation of the target posterior distribution of the parameters, we follow the standard approach in the applied literature and we use a tailored multivariate Student- $t$  distribution as the proposal distribution. Our prior for  $\Phi$  is a normal distribution, truncated to ensure stationarity.  $\Omega$  is inverse-Wishart,  $\lambda$  is Gamma, and  $\gamma$  is uniform over  $[q_l, q_h]$  where  $q_l$  and  $q_h$  are the highest and the lowest observed values of the the threshold variable.<sup>8</sup> The full technical details of the posterior sampler and the priors are relegated to the appendix.

A crucial empirical question is whether the effects of government spending really do differ across regimes defined by economic slack. In a frequentist setting, to test for the presence of nonlinear effects, we would want consider the null hypothesis  $H_0 : \Phi_0^2 = \Phi_1^2 = 0$  that the coefficients are equal against the alternative that at least one of the elements of the matrices  $\Phi_0^2$ ,  $\Phi_1^2$  is not

<sup>7</sup>Morley and Piger (2012) also find that capacity utilization is closely related to their asymmetric measure of the business cycle based on a model-averaged forecast-based trend/cycle decomposition given a wide range of linear and nonlinear time series models of quarterly U.S. real GDP.

<sup>8</sup>Using a truncated univariate Student- $t$  prior for  $\gamma$  with mean equal to the maximum likelihood estimate and 5 degrees of freedom (relatively flat over the observed values) leads to very similar posterior estimates.

zero. This testing problem is tainted by the fact that the threshold  $\gamma$  is not identified under the null. If the errors are i.i.d., a test with near-optimal power against alternatives distant from the null hypothesis is the  $\sup LR$  test, but the asymptotic distribution of the test statistic is nonstandard and has to be approximated using Hansen’s (1996, 1997) bootstrap procedure. Because the model is very parameter-rich, bootstrapping the asymptotic distribution is computationally prohibitive. Also, it should be noted that the 1984Q1 structural break in the variance-covariance matrix of the disturbances makes it unclear how well Hansen’s procedure would perform in this setting. The Bayesian approach circumvents such problems by providing a direct method for comparing models based on the posterior odds ratios.<sup>9</sup> In particular, we estimate the threshold VAR model using the MH algorithm and then we compare its marginal likelihood to that for a restricted linear version of the VAR model, specified as follows:

$$Y_t = \Phi_0^0 + \Phi_1^0(L)Y_{t-1} + \varepsilon_t \quad (2)$$

The marginal likelihoods are calculated using Chib and Jeliazkov’s (2001) algorithm and we compare models based on Bayes factors, which are the ratio of marginal likelihoods and are equal to posterior odds ratios under even prior odds (i.e., equal prior probabilities on all models under consideration).

To estimate the effects of shocks to government spending, we calculate impulse response functions for output and other outcome variables in response to a shock to government spending. We do this for two reasons. First, the impulse responses give us the magnitude of the response of output and its components to government spending, so they can be used to define the multiplier. Second, when it comes to designing policies, the response of output is much more important than the coefficient estimates. Because the impulse responses are nonlinear functions of the coefficients, a small asymmetry in the coefficients might correspond to a large asymmetry in the impulse responses or vice versa. It is important to note that rejecting nonlinearity implies that the impulse responses are necessarily different across regimes, but because the responses are complicated highly nonlinear functions of the coefficients, the degree of this asymmetry can only be evaluated by looking at the impulse response functions, rather than solely by looking at the coefficients.

For the nonlinear model, we construct two sets of impulse responses. In the first case, the economy is assumed to remain in a given state forever. Because the model is linear within a state, the impulse response functions can be obtained by using the estimated VAR coefficients for the given regime. In the second case, the state of the economy is allowed to evolve because the threshold variable itself responds to government spending shocks. When we allow the system to evolve and switch between regimes, the impulse response function depends on

---

<sup>9</sup>It should be noted, however, that a bootstrap version of the  $\sup LR$  test for a simpler version of the model with only government spending, net taxes, and real GDP as endogenous variables and still using capacity utilization as the threshold variable is significant at the 5% level (under the assumption that the structural break does not distort the test). The results are available from the authors upon request.

the initial state and possibly on the size and the sign of the shock. Following Koop, Pesaran and Potter (1996), we consider generalized impulse response functions (GIRFs) in order to obtain the responses when the threshold variable is allowed to respond endogenously. A GIRF is defined as the change in conditional expectation of  $Y_{t+k}$  as a result of a shock at time  $t$ :

$$E[Y_{t+k}|shock_t, \Psi_{t-1}] - E[Y_{t+k}|\Psi_{t-1}] \quad (3)$$

where  $\Psi_{t-1}$  is the information set at time  $t-1$ . Calculating the GIRFs requires specifying the nature of the shock and the initial conditions  $\Psi_{t-1}$ , and then the conditional expectations  $E[Y_{t+k}|shock_t, \Psi_{t-1}]$  and  $E[Y_{t+k}|\Psi_{t-1}]$  are computed by simulating the model. Similar to Kilian and Vigfusson (2011), we consider an orthogonal exogenous shock identified from the SVAR model rather than a forecast error from the reduced-form VAR, as was considered in Koop, Pesaran and Potter (1996).

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

Because threshold models imply that the predicted responses from the model to a shock depend on a particular history, we first simulate the responses for the evolving model, averaging over all histories when the threshold variable is above the estimated threshold and averaging over states when it is below. Then we compare those results to those obtained when we simulate the GIRFs for the recent histories between 1984-2011 when the threshold is above the threshold and when it is below, including the “New Economy” rapid expansion in the late 1990s and the “Great Recession” in the late 2000s. To capture the uncertainty about the parameter values, the credibility intervals for the impulse response functions are obtained by simulating the GIRFs for all iterations of the MH algorithm.

## 4 Empirical Results

### 4.1 Model Comparisons

As discussed in Section 3.4, our formal model comparisons are based on marginal likelihoods and the implied Bayes factors. Table 2 reports marginal likelihood values for the baseline model with different threshold variables, including the restricted case of no threshold effect. The implied Bayes factors strongly favor nonlinearity when considering threshold variables related to economic slack. There is also support for nonlinearity when the threshold variable is lagged output growth, as considered in Auerbach and Gorodnichenko (2012a) and Mettnik and Semmler (2011), although neither of those studies formally tested for nonlinearity. By contrast, there is no support for nonlinearity when considering the debt-to-GDP ratio or the real interest rate. For the debt-to-GDP ratio, the estimated threshold is near the boundary of the parameter space considered, so the lack of support for nonlinearity in this case might reflect the relatively low levels of the debt-to-GDP ratio in U.S. economy since 1967, at least compared to the levels observed in other countries that have suffered debt crises. For the real interest rate, the estimated threshold is about 2 percent, which is often thought to be close to the long-run “neutral” rate. However, this estimate is quite imprecise, consistent with the lack of support for a threshold effect relating to the interest rate.

Based on the marginal likelihood results in Table 2, the preferred threshold variable for the baseline model is the first lag of capacity utilization (adjusted for a one-time structural break in the mean, as discussed in Section 3.3). The estimated threshold for the baseline model is slightly below the mean of the adjusted capacity utilization series.<sup>10</sup> The mean-adjusted capacity utilization series and its estimated threshold are plotted in Figure 2. The high-utilization regime dates estimated with capacity utilization as the threshold variable largely coincide with the high-utilization regime dates found when using the CBO’s output gap estimate as a measure of slack and with the regime dates obtained when using output or unemployment as measures of slack.

Notably, more than 60% of the historical observations for mean-adjusted capacity utilization fall below the mode of posterior distribution for the threshold parameter, while close to 50% fall below the posterior mean. This result is important because it implies that, for a majority of the time since the middle 1960s, the U.S. economy has operated in a regime in which government spending shocks have relatively large effects on output. Since 2000, almost all observations have been in the low regime. A possible interpretation of these results is that demand, not supply, has been the proximate constraint on aggregate output for much of the sample period. This result also distinguishes our approach from Auerbach and Gorodnichenko (2012a), as their approach imposes that only 20% of the observations fall in a recessionary regime.

Although the marginal likelihood results in Table 2 strongly favor nonlin-

---

<sup>10</sup>The sample mean for unadjusted capacity utilization is 85.2 for 1967Q1-1974Q1 and 80.0 for 1974Q2-2011Q1.

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

When we estimate the effects of government spending on output components and other variables, we substitute the outcome variable of interest (i.e., consumption, investment, exports, imports, the unemployment rate, employment, and inflation) for output in the baseline VAR model, using the first lag of mean-adjusted capacity utilization as the threshold variable.<sup>11</sup> As with the baseline model, we find strong evidence of nonlinearity for these models. Table 3 reports the marginal likelihood values (and the modes of the likelihood) for linear and nonlinear specifications of these alternative VAR models. In every case, the nonlinear specification is preferred. In particular, the implied Bayes factors always prefer the nonlinear specification, with posterior odds only favoring linear specifications under extremely high prior odds of more than 10 to 1 on the linear specifications. Table 3 also reports the estimated thresholds in mean-adjusted capacity utilization for these alternative VAR models and shows that they are quite consistent across the different models.

## 4.2 Responses of Output to a Government Spending Shock

As discussed above, we identify government spending shocks in the SVAR model by assuming output and its private-sector components can respond to government spending within a quarter, but government spending does not respond to output within the same quarter. The results are similar when we consider alternative identification schemes: specifically, we obtain almost identical results when we reorder taxes and government spending so that spending can

---

<sup>11</sup>In preliminary analysis, we also considered these effects by adding each series as a fifth variable to the baseline model. The point estimates for the threshold and the median impulse responses were very similar for both specifications, but the 95% (and even the 75%) credibility intervals were very wide in the specification with five variables because there are too few observations per regime to precisely estimate a threshold VAR with so many variables without imposing very tight priors. Thus, the results presented in the rest of this paper are based on four-variable versions of the VAR model.

respond to tax shocks, when we use Blanchard and Perotti's (2002) identification scheme that imposes short-run tax elasticities, and when we add Ramey's narrative spending variable and order it first so that the rest of government spending can respond to military spending within a quarter.<sup>12</sup>

When constructing the impulse response functions to government spending, the initial shock to government spending is set to be equal to 1 percent of GDP. The shock to government spending initiates a dynamic path of adjustment for both government spending and other variables of interest. To make the interpretation of the results more straightforward and to facilitate comparison with the linear results obtained by Blanchard and Perotti (2002), we calculate the dynamic multipliers as ratios of the cumulative dollar-for-dollar change in the variable of interest to the cumulative dollar-for-dollar change in government spending. Meanwhile, when we examine the behavior of employment, unemployment, and inflation, the responses are presented as level responses to a shock to government spending equal to 1 percent of GDP.

Our primary results appear in Figure 3. The top row of the figure shows the impulse responses of output to a government spending shock for the two regimes, in both cases assuming that the economy remains in the same state forever. The response of output to spending shocks depends strongly on the regime. The shock to government spending pushes output up immediately in both the high and the low utilization regimes. However, in the low regime, output rises almost monotonically to a cumulative change in output equal to 1.6 times the cumulative change in government spending. Most of the effect takes place in the first three years. In the high-utilization regime, the pattern is substantially different. After the initial positive response, the cumulative change in output falls back towards zero. The long-term response is positive, but the multiplier is less than half of that when output is in the low-utilization regime.

When the economy is allowed to evolve from one state to another, the magnitude of the multiplier varies depending both on the state of the economy at the time of the government spending shock and on the actual history of other shocks. As shown in the bottom two rows of Figure 3, the output response for all low states peaks at 1.6 after two years and then the effects of the spending shock die out. The lower bound of the credibility interval for the low-regime impulse response is strongly positive, despite the fact that we use a fairly conservative 90% credibility interval. In comparison, the average response for all high states peaks at 0.8 after two years, and then it remains stable, but the credibility interval always covers zero.

Thus, our estimates clearly suggest that the effects of government spending on output are larger and more persistent when capacity utilization is low. In the following subsections, we examine the source of this result about the asymmetric response of output to fiscal policy shocks in more detail. In particular, we

---

<sup>12</sup>Owyang and Zubairy (2010) also find impulse response functions for SVAR models are broadly robust when considering different identification schemes, including sign restrictions. They consider a linear VAR model that includes U.S. state-level data and separates out military spending, as in Ramey (2011b).

look at the responses of output components in order to determine whether the asymmetry comes from an asymmetry in the response of fiscal variables to the government spending shock or if it is due to an asymmetric response in the components of private spending.

### 4.3 Responses of Fiscal Policy to a Government Spending Shock

From Figure 4, it is clear that the response of government spending to its own shock does not depend very strongly on the prevailing regime. In this case, the impulse response functions are shown as cumulative dollar-for-dollar changes in government spending relative to the size of the initial shock, because the ratio of government spending to itself is necessarily equal to one. For both regimes, the peak cumulative dollar-for-dollar change is roughly 1.3, which is consistent with the results obtained in the linear case by Blanchard and Perotti (2002) and similar to the results obtained by Auerbach and Gorodnichenko (2012a). Both the credibility intervals for the regimes overlap and the actual estimated responses are similar across regimes. The similar responses across regimes clearly indicate then that the asymmetric response of output is not due to higher or more persistent government spending in the low-utilization regime.

Figure 5 shows that the peak response of tax revenues to a government spending shock is roughly 0.8 in evolving regimes, with little effect of the initial state of the economy. In the fixed low-utilization regime, tax revenues appear to increase persistently after a government spending shock, while the response of tax revenues is smaller and dies off quickly when the economy starts and remains in the high regime. But, given the wide credibility intervals for the responses at long horizons, there is no obvious evidence of an asymmetry in the response of tax revenues that could explain the asymmetry in the response of output to a government spending shock.<sup>13</sup> Thus, the asymmetry in the response of output to a fiscal policy shock appears to be due to an asymmetry in the response of private spending, not the government sector of the economy.

### 4.4 Responses of Consumption and Investment to a Government Spending Shock

Figure 6 displays the responses of consumption to a government spending shock. The main result is that consumption increases in both regimes, but the magnitude of the response is much larger when the economy is in the low-utilization

---

<sup>13</sup>It is important to note that these results are for the responses of tax revenues, not tax rates. Tax revenues are correlated with income, so part of the increase in revenues comes from increases in income due to the positive government spending shock, indicating that spending could be partially self-financing (although further analysis would be necessary to examine this possibility given the wide credibility intervals). Another part of the increase in revenues could come from the endogenous response of tax rates to a government spending shock. The use of tax revenues also makes it difficult to interpret the responses of output and its components to changes in taxes because individuals and firms respond to marginal tax rates. Unfortunately, though, reliable data for marginal tax rates are only available at an annual frequency.



regime.<sup>14</sup> When starting from a low-utilization state, but allowing the state to evolve, the long-run response levels off after three years at close to 0.8, averaging over all histories (the effect is even larger when averaged over recent histories). Consumption is much less responsive when the economy starts in a high-utilization state. The peak response in this case is only around 0.4, and becomes insignificant after a year. Thus, it appears that the asymmetric response of output to government spending is at least partly due to an asymmetry in the magnitude of the response of consumption. Meanwhile, the findings of a positive response of consumption in both regimes is consistent with the linear results obtained by Blanchard and Perotti (2002), Pappa (2009), and Woodford (2011). Also, accounting for anticipated government spending by including Ramey’s military spending variable and ordering it first in the linear or nonlinear versions of the SVAR does not change the response or the significance of the response in either case.

Figure 7 displays the responses of investment, which also appear to be asymmetric depending on the state of the economy. In the fixed low regime, investment increases in response to government spending, with a peak response of 0.4, although the credibility interval includes zero. When the economy is assumed to remain in the high-utilization state forever, investment drops significantly in response to a spending shock, with a cumulative decline equal to 0.9 after five years. Allowing the economy to evolve from one regime to another, the responses of investment are weakly positive when the economy starts from a low-utilization state and not different from zero when the economy starts from a high-utilization state. These results indicate the relevance of crowding out in the high-utilization state, but provide no support for crowding out in the low-utilization state. Furthermore, these results may help explain the “investment puzzle” in linear studies such as Blanchard and Perotti (2002) because the negative response in the linear VAR is roughly a weighted average of the responses in the nonlinear model. Specifically, the apparent neoclassical behavior of investment found in these studies appears to reflect crowding out when capacity utilization is high only.

Overall, the strong state-dependence in the responses of consumption and investment suggests that a lot of the asymmetry in the response of output is due to different responses of these key components to government spending depending on the degree of resource utilization.

## 4.5 Responses of Other Macroeconomic Variables to a Government Spending Shock

Figure 8 displays the fixed-regime responses of exports and imports, which are very similar across regimes. Both exports and imports weakly increase, and the increase in exports roughly cancels out the increase in imports. Furthermore, the credibility intervals for the state-dependent responses overlap, suggesting little support for asymmetric responses of imports and exports to government

---

<sup>14</sup>This result is robust to considering consumption of nondurables and services only.

spending. Indeed, the increase in imports and the decrease in exports always essentially cancel each other, regardless of the priors or the measure of slack we consider. For simplicity, we do not report responses when allowing the state of the economy to evolve, although not surprisingly those responses display very little asymmetry given the results for the fixed regimes.

Figure 9 shows that the unemployment rate decreases in response to a spending shock in both states. In the low-utilization regime, the unemployment rate decreases monotonically, falling by a total of 2.5 percentage points after five years. The effect of a spending shock on the unemployment rate is weaker and less persistent when the economy is in the high-utilization regime. The impact response is essentially zero, and the maximum response (in magnitude) is a 1.3 percentage point decline. When analyzing the magnitude of the responses, it is important to keep in mind that the impulse responses were constructed using a relatively large spending shock (1 percent of the GDP), which might explain the large responses of the unemployment rate.

The responses of employment also exhibits state-dependence that is consistent with the responses of the unemployment rate. In Figure 10, when the economy is in a low-utilization state, employment increases by 1 percent after two years, and the long run response is equal to 0.8 percent. When the economy starts from a high-utilization state, the effect of a government spending shock on employment is only slightly positive and transitory. The credibility intervals for employment, however, are quite wide, and zero effects are not outside the 90% interval for either regime. This result is due to the fact that we use a conservative 90% interval and the fact that employment only builds up slowly after the shock. In particular, the estimated employment effects are only large at longer horizons, while credibility intervals are always wide at longer horizons for SVAR models.

Figure 11 displays the responses of inflation. In the fixed low regime, the response is less than 0.1 percentage points after 20 quarters versus 0.2 percentage points in the fixed high regime. Furthermore, the response of inflation in the fixed low regime is short-lived as the response dies off after only a few quarters. If we assume that the economy stays in the high-utilization regime forever, higher government spending has a more persistent effect on inflation. The estimated responses are consistent with the idea that government spending crowds out resource use, thus increasing marginal costs when the economy is close to capacity. But it should be noted that the response of inflation is rather small in both regimes and the credibility intervals are quite wide (due to the VAR polynomial having a root that was relatively close to one). Generalized impulse responses can still be used when the roots are close to one, but credibility intervals tend to be quite wide in this case. The responses in the evolving regimes are not displayed, as they are very similar and show little significant response of inflation to a government spending shock.

## 4.6 Counterfactual Analysis

One of the main criticisms of the ARRA fiscal stimulus is that output growth remained anemic two years after it was first implemented and that the unemployment rate remained persistently high. However, in order to fully evaluate whether the stimulus had any effect on the economy, it is important to compare what output and the unemployment rate would have been if there had been no stimulus in the first place. Taking into account that in the months before the stimulus package was implemented, interest rates were already approaching the zero bound and that employment was dropping precipitously, the absence of fiscal stimulus could have resulted in even lower output growth and the unemployment rate rising even higher than its 10.1 percent peak.

In order to evaluate the effects of the stimulus, we make use of our preferred model to perform counterfactual analysis for the period 2009-2010 in which we compare the implied path of output and employment without increases in government spending to the actual path of output and spending. We do this by orthogonalizing the shocks for each period and setting the orthogonalized government spending shocks equal to zero for the period 2009-2010.<sup>15</sup> Figure 12 displays the results of the counterfactual experiments. If we set the spending shocks between 2009 and 2010 equal to zero, the economy would have needed one more quarter to recover (i.e., the simulated output growth would not have become positive until 2009Q3) and the recovery would have been even more sluggish, with the maximum simulated growth rate reaching 0.4 percent at the end of 2009 and then dropping down to zero by the end of 2010.

The results for unemployment are similar. Without fiscal stimulus, the simulated estimate for the unemployment level is 0.7 percentage points higher in 2010Q2 than what the actual unemployment rate turned out to be. Furthermore, without the spending shocks the unemployment rate would have stayed higher (around 10.3 percent), without dropping below double digits at any point. Interestingly, our results based on the counterfactual analysis are somewhat similar to the CBO estimates for the effects of the ARRA stimulus and support the idea that, even if employment and output growth did not reach the high rates that are typical for a recovery, the fiscal stimulus was helpful in the sense that it prevented the recession from becoming deeper and longer.

## 5 Conclusions

In this paper, we have presented strong empirical evidence in favor of state-dependent effects of fiscal policy. In particular, the estimates from a threshold structural vector autoregressive model clearly identify different responses of the

---

<sup>15</sup>Taylor (2011) notes that the overall impact of the ARRA stimulus on government spending was not as large as advertised because state and local government saved a large portion of their stimulus grants. Thus, the implied effects on output and employment reported below should not be thought of in terms of a one-time exogenous \$700 billion dollar increase in federal government spending, but rather in terms of the much smaller exogenous changes in total government spending over the period under consideration.

economy to government spending shocks depending on whether the economy has high or low utilization of economic resources. We find that a rise in demand from the government sector causes large and persistently positive effects on output when the economy is operating with relatively low capacity utilization. This effect is much smaller and less persistent when capacity utilization is above an estimated threshold for our model.

It is particularly interesting to note that the estimated threshold for capacity utilization is such that a majority of observations for the U.S. economy over the past 40 years appear to have been in the low-utilization regime in which demand shocks have larger and more persistent effects, with any constraints from the supply side binding less tightly. We infer from this result that the normal state for the U.S. economy is one of significant excess capacity. Therefore, the proximate effect of a demand shock is more likely than not to be positive and persistent.

We find no evidence that higher government spending crowds out consumption. Indeed, consumption rises after positive government spending shocks in both the high- and low-utilization regimes, but the increase is almost twice as large during “normal” times (i.e., low utilization) than during “booms” (i.e., high utilization). Most of the increase in the private components of output comes from the increase in consumption. These results for consumption are consistent with the linear results obtained by Blanchard and Perotti (2002), Perotti (2008) and Pappa (2009), but are at odds with the simulation results obtained using most calibrated dynamic stochastic general equilibrium (DSGE) models. Only when allowing for a high proportion of rule-of-thumb consumers do Galí et al (2007) find such large responses of consumption in an estimated (not calibrated) DSGE model. Meanwhile, the state-dependent responses of consumption are potentially related to the results obtained by Kaplan and Violante (2012), who develop a life-cycle model that endogenizes the proportion of rule-of-thumb consumers in order to examine the effect of taxes on consumption when a large proportion of the consumers’ wealth is tied up in illiquid assets such as real estate. Historically, the number of credit-constrained consumers rises in recessions, and the Great Recession started with the crash of the housing market, which likely implied a large increase in the proportion of credit-constrained consumers in its aftermath. Even more directly along these lines, Canzoneri et al. (2012) calibrate a New Keynesian DSGE model with costly financial intermediation and show that countercyclical shocks to the spread between rates paid by borrowers and received by depositors implies countercyclical fiscal multipliers, although this is a fairly mechanical result given the assumptions of countercyclical spread shocks and the ability of government spending shocks to disproportionately lower borrowing costs when the level of output is lower.

Regardless of the exact mechanism behind the state-dependent effects of fiscal policy, the implications for policy are straightforward and significant. Higher government spending raises output, but this effect is both larger and more persistent when capacity utilization is low. At these times, including during recessions, higher government spending reduces economic slack and increases output, consumption, and investment. Although stimulus policy to reduce slack may

increase government debt, the effect is smaller than a simple calculation would suggest because higher government spending raises output, income, and therefore tax revenue, and the effect of spending stimulus on public debt is less than dollar for dollar.

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

## References

- [1] Andrews, D. and W. Ploberger (1994): “Optimal Tests when a Nuisance Parameter Is Present Only under the Alternative.” *Econometrica* 62(6), 1383-1414.
- [2] Auerbach, A. J. (2009): “Implementing the New Fiscal Policy Activism.” *American Economic Review: Papers and Proceedings* 99-2, 543-549.
- [3] Auerbach, A. J. and Y. Gorodnichenko (2012a): “Measuring the Output Responses to Fiscal Policy.” *American Economic Journal: Economic Policy* 4(2), 1-27.
- [4] Auerbach, A. J. and Y. Gorodnichenko (2012b): “Fiscal Multipliers in Recession and Expansion.” Working Paper.
- [5] Bachmann, R. and E. R. Sims (2012): “Confidence and the Transmission of Government Spending Shocks.” *Journal of Monetary Economics* 59, 235-249.
- [6] Barro, R. J. and C. J. Redlick (2011): “Macroeconomic Effects from Government Purchases.” *Quarterly Journal of Economics* 126(1), 51-102.

- [7] Blanchard, O., and R. Perotti (2002): “An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output.” *The Quarterly Journal of Economics* 117(4), 329-1368.
- [8] Calvo, G. A. (1983): “Staggered Contracts in a Utility Maximizing Framework.” *Journal of Monetary Economics* 12, 383-398.
- [9] Canzoneri, M., F. Collard, H. Dellas, and B. Diba (2012): “Fiscal Multipliers in Recessions.” Working Paper.
- [10] Chib, S. and I. Jeliazkov (2001): “Marginal Likelihood from the Metropolis-Hastings Output.” *Journal of the American Statistical Association* 96, 270-281.
- [11] Christiano, L. M. , M. Eichenbaum, and S. Rebelo (2011): “When Is the Government Spending Multiplier Large?” *Journal of Political Economy* 119(1), 78-121.
- [12] Cogan, J. F. , T. Cwik, J.B. Taylor, and V. Wieland (2010): “New Keynesian Versus Old Keynesian Government Spending Multipliers.” *Journal of Economic Dynamics and Control* 34, 281-295.
- [13] Davig, T. and E. M. Leeper (2009): “Expectations and Fiscal Stimulus.” *CAERP Working Paper* 006-2009.
- [14] Edelberg, W., M. Eichenbaum, and J. Fisher (1999): “Understanding the Effects of a Shock to Government Purchases.” *Review of Economic Dynamics* 2, 166-206.
- [15] Eggertsson, G.B. and P. Krugman (2010): “Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach.” Working Paper.
- [16] Galí, J., D. López-Salido, and J. Vallés (2007): “Understanding the Effects of Government Spending on Consumption.” *Journal of the European Economic Association* 5, 227-270.
- [17] Hansen, B. E. (1996): “Inference when a Nuisance Parameter is not Identified under the Null Hypothesis.” *Econometrica* 64, 413-430.
- [18] Hansen, B. E. (1997): “Inference in TAR Models.” *Studies in Non-linear Dynamics and Econometrics* 2(1), 1-14.
- [19] Inoue, A. and L. Kilian (2011): “Inference on Impulse Response Functions in Structural VAR Models.” Working Paper.
- [20] Kaplan G. and G.L. Violante (2012) “A Model of the Consumption Response to Fiscal Stimulus Payments.” NBER Working Paper.

- [21] Kilian, L. and R.J. Vigfusson (2011): “Are the Responses of the U.S. Economy Asymmetric in Energy Price Increases and Decreases?” *Quantitative Economics* 2, 419-453.
- [22] Kim, C-J. and C. R. Nelson (1999): “Has the US Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle.” *The Review of Economics and Statistics* 81(4), 608-616.
- [23] Koop, G. M.H. Pesaran, and S.M. Potter (1996): “Impulse Response Analysis in Nonlinear Multivariate Models.” *Journal of Econometrics* 74, 119-147.
- [24] Leigh, D. , P. Devries, C. Freedman, J. Guajardo, D. Laxton, A. Pescatori (2010): “Will it Hurt? Macroeconomic Effects of Fiscal Consolidation.” IMF Working Paper, October 2010.
- [25] Leeper, E.M., N. Traum, and T.B. Walker (2011): “Clearing up the Fiscal Multiplier Morass.” Working Paper.
- [26] Lo, C. M. and J. Morley (2011): “Bayesian Analysis of Nonlinear Exchange Rate Dynamics and the Purchasing Power Parity Persistence Puzzle.” Working Paper, University of New South Wales.
- [27] McConnell, M. M. and G. Perez-Quiros (2000): “Output Fluctuations in the United States: What Has Changed Since the Early 1980’s?” *American Economic Review* 90(5), 1464-1676.
- [28] Mittnik, S. and W. Semmler (2011): “Regime Dependence of the Fiscal Multiplier.” Working Paper.
- [29] Morley, J. and J. Piger (2012): “The Asymmetric Business Cycle.” *The Review of Economics and Statistics* 91(1), 208-221.
- [30] Mountford, A. and H. Uhlig (2009): “What Are the Effects of Fiscal Policy Shocks?” *Journal of Applied Econometrics* 26, 960-992.
- [31] Owyang, M.T. and S. Zubairy (2010): “Who Benefits from Increased Government Spending? A State-Level Analysis.” Working Paper.
- [32] Papell, D.H., C.J. Murray, and H. Ghiblawi (2000): “The Structure of Unemployment.” *The Review of Economics and Statistics* 82(2), 309-315.
- [33] Pappa, E. (2009): “The Effects of Fiscal Shocks on Employment and Real Wages.” *International Economic Review* 50(1), 217-244.
- [34] Parker, J. A. (2011): “On Measuring the Effects of Fiscal Policy in Recessions.” *Journal of Economic Literature* 49, 703-718.

- [35] Perotti, R. (2008): “In Search of the Transmission Mechanism of Fiscal Policy.” *NBER Macroeconomic Annual 2007*, ed. by D. Agemogly, K. Rogoff, and M. Woodford, pp. 169-226. University of Chicago Press, Chicago, Illinois.
- [36] Ramey, V. A. (2011a): “Can Government Purchases Stimulate the Economy?” *Journal of Economic Literature* 126(1).
- [37] Ramey, V. A. (2011b): “Identifying Government Spending Shocks: It’s All in the Timing.” *The Quarterly Journal of Economics*.
- [38] Ramey, V. A. and M. Shapiro (1998): “Costly Capital Reallocation and the Effects of Government Spending.” *Carnegie Rochester Conference on Public Policy* 48, 145-194.
- [39] Reinhart, C.M. and K.S. Rogoff (2009): “This Time Is Different: Eight Centuries of Financial Folly.” Princeton University Press.
- [40] Romer, C. and J. Bernstein (2009): “The Job Impact of the American Recovery and Reinvestment Plan.” *Executive Office of The President of the United States, Council of Economic Advisers Report*, January 8, 2009.
- [41] Romer, C. and D. Romer (2009): “The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks.” *American Economic Review* 100, 763-801.
- [42] Sims, C.A. (2001): “Comment on Sargent and Cogley’s ‘Evolving US Postwar Inflation Dynamics’.” *NBER Macroeconomics Annual* 16, 373-379.
- [43] Taylor, J.B. (2011): “An Empirical Analysis of the Revival of Fiscal Activism in the 2000s” *Journal of Economic Literature* 49(3), 686-702.
- [44] Tong, H. (1978): “On a Threshold Model in Pattern Recognition and Signal Processing.” C. H. Chen, Amsterdam: Sighoff and Nordhoff.
- [45] Tong, H. (1983): “Threshold Models in Non-linear Time Series Analysis.” New York: Springer-Verlag.
- [46] Van Brusselen, P. (2009): “Fiscal Stabilization Plans and the Outlook for the World Economy.” *NIME Policy Brief* 01-2009.
- [47] Woodford, M. (2011): “Simple Analytics of the Government Expenditure Multiplier.” *American Economic Journal: Macroeconomics* 3(1), 1-35.



## A Bayesian Estimation

For the baseline linear model, we assume that the prior for the VAR parameters is multivariate normal, the prior for the variance matrix is an inverse Wishart distribution, and the prior for the scale parameter  $\lambda$  is a Gamma distribution. Under these assumptions, we can sample directly using a Gibbs step. Specifically, recall that the linear model is given by (1):

$$Y_t = \Phi_0 + \Phi(L)Y_{t-1} + \lambda_t \varepsilon_t$$

where  $\Phi(L)$  is an autoregressive matrix polynomial with roots strictly outside the unit circle,  $\lambda_t=1$  for  $t = 1967q1, \dots, 1983q4$ ,  $\lambda_t=\lambda$  for  $t = 1984q1, \dots, 2010q4$ , and  $\varepsilon_t$  is i.i.d. Gaussian random variable with mean 0 and variance-covariance matrix  $\Omega$  that does not change over time. Then, letting  $\Phi = \text{vec}(\Phi_0) | \text{vec}(\Phi_1) \dots | \text{vec}(\Phi_j)$ , we assume that the prior for  $\Phi$  is a normal distribution, truncated to the stationarity region, with mean equal to 0, and variance-covariance matrix equal to  $V_n$ . The scaling parameter  $\lambda$  is assumed to have a gamma prior with parameters  $\alpha$  and  $\beta$ , and we impose an inverted Wishart prior with  $\nu_0 = 25$  degrees of freedom and a scale matrix  $R_0$ . For brevity, let  $rhs_t = [1 \ y_{t-1,1} \ \dots \ y_{t-1,k} \ \dots \ y_{t-p,k}]$  and

$$x_t = \begin{bmatrix} rhs_t & 0 & 0 & 0 \\ 0 & rhs_t & 0 & 0 \\ 0 & 0 & rhs_t & 0 \\ 0 & 0 & 0 & rhs_t \end{bmatrix}.$$

It is straightforward to see that  $\Phi | \Omega, \lambda, y$  is Gaussian with variance

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

and mean

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

Similarly,  $\Omega | \Phi, \lambda, y \sim IW(\nu_1, R_1)$ , where  $\nu_1 = \nu_0 + T - 4$  and  $R_1 = [R_0^{-1} + \sum_{t=5}^T (y_t - x_t \Phi)' \lambda_t^{-1} (y_t - x_t \Phi)]^{-1}$ . The inverse Wishart distribution is a standard distribution, so we can sample  $\Omega$  conditional on the other parameters directly. Conditional on the other parameters and the data,  $\lambda$  has a gamma distribution with parameters  $\alpha_1 = \alpha + T_p$  and  $\beta_1 = \beta + 0.5 \sum_{t=t_1}^T (y_t - x_t \Phi)' \Omega^{-1} (y_t - x_t \Phi)$  where  $t_1 = 1984Q1$ , and  $T_p = 108$  (the number of quarters from 1984Q1 to 2010Q4).

The threshold model is given by

$$Y_t = \Phi_0^1 + \Phi_1(L)Y_{t-1} + (\Phi_0^2 + \Phi_2(L)Y_{t-1})I[q_t \leq \gamma] + \lambda_t \varepsilon_t$$

where  $q_t$  is the threshold variable,  $\gamma$  is the threshold around which the dynamics of the model changes, and the other variables are defined similarly to the

variables in the linear model. Let  $rhs_{1t} = [rhs_t \ rsh_t * I[q_t \leq \gamma]]$  and assume that the prior for  $\Phi = vec(\Phi_0^1) | vec(\Phi_1) | vec(\Phi_2^0) | vec(\Phi_2)$  is normal with mean 0 and variance 0, truncated so that  $\Phi_1(L)$  and  $\tilde{\Phi}(L) = \Phi_1(L) + \Phi_2(L)$  have roots strictly outside the unit circle (i.e. so that the VAR is stationary in each regime).

Similar to the linear case, it is straightforward to show that  $\Phi | \Omega, \lambda, \gamma, y$  is Gaussian with variance  $V = (V_n^{-1} + \sum_{t=p+1}^T x'_{1t} (\lambda_t \Omega)^{-1} x_{1t})^{-1}$  and mean  $\mu = V^{-1} (\sum_{t=p+1}^T x'_{1t} (\lambda_t \Omega)^{-1} y_t)$ , where  $x_{1t}$  is defined the same way as  $x_t$ , except we use the vector  $rhs_{1t}$  is used in place of  $rhs_t$ . Likewise, the conditional distribution of  $\Omega$  is inverse Wishart, and the conditional posterior distribution of  $\lambda$  is gamma, and we can sample from these posteriors using a Gibbs step. The conditional distribution for  $\gamma$  is nonstandard, and it has to be sampled using an MH step.

Following the standard approach in the literature, the proposal density is Student- $t$  with 15 degrees of freedom. To obtain the mode for the proposal distribution for the first draw, we use concentrated maximum likelihood (ML) and grid search over the middle 70% of the sample range for the threshold variable in order to obtain the posterior mode of the parameter  $\gamma$ . Because  $\varepsilon_t$  is assumed to be Gaussian, the ML estimators can be obtained by using least squares estimation. For this maximization  $\gamma$  is restricted to a bounded set  $\Gamma = [\underline{\gamma}, \bar{\gamma}]$  that covered the middle 70% of the threshold variable.

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

$$\hat{\gamma} = \arg \max_{\gamma \in \Gamma_n} \text{lik}_n(\gamma | q, d) \quad (4)$$

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

Note, however, that the grid search makes it infeasible to obtain the variance of the estimate of  $\gamma$ . To address this issue, we use the approach proposed by Lo and Morley (2011). In particular, we obtain a measure of the curvature of the posterior distribution with respect to  $\gamma$  by inverting the likelihood ratio statistics for the threshold parameters, based on the assumption that the parameter estimate is normally distributed and the LR statistics is  $\chi^2(1)$ . We use the 95% CI for the likelihood ratio statistics to obtain a corresponding standard error for  $\gamma$ . Again, we do not attempt to perform frequentist inference. This is simply a fast way of obtaining a plausible value for the curvature. It is important to note that this approach is only an approximation, and it is not asymptotically accurate

for obtaining standard errors of the ML estimate for  $\gamma$ , but it is much faster than bootstrapping and inverting the LR test in order to obtain the scale for the first draw of the MH algorithm. Because this approximation affects only the first draw, it will not affect the estimates of the parameters if the burn-in is large enough.

At the  $i^{th}$  iteration, the transition density for  $\gamma^{(i+1)}$  is a Student-t distribution with mean equal to  $\gamma^{(i)}$  and variance equal to  $\kappa \hat{\sigma}_\gamma^2$ , where  $\hat{\sigma}_\gamma^2$  is obtained by inverting the LR test. The parameter  $\kappa$  is calibrated on the fly to ensure acceptance rate between 20 and 60%.

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

	Type of Prior	Mean	Variance/ Scale
$\Phi$	Multivariate Gaussian	0	$100 * I_k$
$\Omega$	Inverse Wishart	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$25 * \mu$
$\lambda$	Gamma	1	0.75
$\gamma$	Uniform	0.165	$0.365^2$

## B Generalized Impulse Response Function

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

$$GIRF_y(k, shock_t, \Psi_{t-1}) = E[Y_{t+k}|shock_t, \Psi_{t-1}] - E[Y_{t+k}|\Psi_{t-1}] \quad (5)$$

where  $k$  is the forecasting horizon and  $\Psi_{t-1}$  denotes the initial values of the variables in the model. The impulse response is then computed by simulating the model. The shock to government spending is normalized to be equal to 1 percent of GDP (at the time the shock occurs).

The  $GIRF_y$  response for a given draw  $\Theta^{(i)}$  of the MH algorithm is generated using the following steps:

1. Pick a history  $\Psi_{t-1}$ . The history is the actual value of the lagged endogenous variable at a particular date.
2. Pick a sequence of 4-dimensional forecast errors  $\varepsilon_{t+k}$ ,  $k = 0, 1, \dots, 20$ . The forecast errors are simulated assuming an independent Gaussian process with mean zero and variance-covariance matrix equal to  $\lambda_t^{(i)} * \Omega^{(i)}$ .
3. Using  $\Psi_{t-1}^T$  and  $\varepsilon_{t+k}$ , simulate the evolution of  $Y_{t+k}$  over  $l + 1$  periods. Denote the resulting path  $Y_{t+k}(\varepsilon_{t+k}, \Psi_{t-1})$  for  $k = 0, 1, \dots, l$ .
4. Using the Cholesky decomposition of  $\Omega_t$  to orthogonalize the shocks, solve for the government spending shock at time  $t$ , replace it with a shock equal to 1 percent of GDP, and reconstruct the implied vector of forecast errors. Denote the implied vector of forecast errors as  $\varepsilon_t^{shock_t}$ , the sequence of forecast errors as  $\varepsilon_{t+k}^{shock_t}$ , and the resulting simulated evolution of  $Y_{t+k}$  over  $l + 1$  periods as  $Y_{t+k}(\varepsilon_{t+k}^{shock_t}, \Psi_{t-1})$  for  $k = 0, 1, \dots, l$ .
5. Construct a draw of a sequence of impulse responses as  $Y_{t+k}(\varepsilon_{t+k}^{shock_t}, \Psi_{t-1}) - Y_{t+k}(\varepsilon_{t+k}, \Psi_{t-1})$  for  $k = 0, 1, \dots, l$ .
6. Repeat steps 2 to 5 for  $B$  times, with  $B = 500$ , and average the sequences of responses to obtain a consistent estimate of the impulse response function conditional on the history.
7. To obtain the average response for a subset of histories, repeat steps 1-6 for a the subset of histories of interest (we compute it for all low states, all high states, the rapid expansion of the late 1990s and the Great recession), and report the response averaged over all histories.

Because the impulse responses are nonlinear functions of the parameters, their distribution is nonstandard and it is not necessarily symmetric around the mean. In this case, reporting the median value is unlikely to be adequate, as the median may not be a valid measure of central tendency, and the median impulse response may not correspond to a well-defined structural model. In order to circumvent this problem, we use the approach proposed by Inoue and Kilian (2011). For a given history, we evaluate the impulse response function for each draw of the MH algorithm, drawing the entire impulse response function for periods 1 through 20. Then we average over histories, and we evaluate the posterior likelihood of the impulse response for that draw of the algorithm (averaged over the histories of interest). The impulse response function with the highest average posterior likelihood is then used for inference. To construct the  $(1 - \alpha) * 100\%$  credibility interval, we order the posterior likelihood values, and we include the impulse responses whose posterior likelihood was in the upper  $(1 - \alpha) * 100$  percentile. This method results in a “credibility cloud” with a shot gun pattern because we draw entire impulse responses rather than responses for each individual point in time. For easy interpretation, we report only the outer points of the cloud.

## C Tables

Table 1: Structural Breaks in Capacity Utilization and the Unemployment Rate

Variable	Break Date(s)	F-statistics	p-value
Capacity Utilization	1974Q1	41.66	<0.001
Unemployment Rate	1974Q3, 1981Q4, 1994Q4	32.34, 16.43, 6.21	<0.01, <0.01, 0.03

Notes: The break dates were obtained using a sequential Quandt-Andrews test. The estimated break dates coincide with the break dates obtained using Bai-Perron’s sequential procedure under the assumption that the mean is the only parameter that has structural breaks.

Table 2: Marginal Likelihood Values and Estimated Thresholds for the Baseline Model with Different Threshold Variables

Threshold Variable	$q_{t-d}$	$ml$	Threshold
None	-	-997.18	-
Output Growth	$\Delta y_{t-2}$	-720.69	1.33 (0.12)
Output Gap	$gap_{t-1}$	-682.52	-0.59 (0.41)
Capacity Utilization, not adjusted for breaks	$cap_{t-1}$	-800.52	81.10 (1.42)
Capacity Utilization, adjusted for breaks	$\hat{c}ap_{t-1}$	-673.69	-0.21 (0.37)
Unemployment Rate, not adjusted for breaks	$un_{t-1}$	-703.25	4.83 (0.33)
Unemployment Rate, adjusted for breaks	$\hat{u}n_{t-2}$	-760.63	-0.26 (0.11)
Debt-to-GDP Ratio	$(b/y)_{t-2}$	-1020.23	47.2 (1.15)
Real Interest Rate	$r_{t-2}$	-1004.42	2.10 (1.35)

Notes: “ $ml$ ” denotes the natural logarithm of the marginal likelihood. The reported threshold variables correspond to particular variable and lag with the highest ML for a given type of threshold variable, as listed in Section 3.3. The preferred debt measure is total federal debt outstanding. The real interest rate is based on the Federal Funds rate and CPI inflation under the assumption of static expectations. The threshold estimate is the posterior mean (with standard deviation in parentheses).

Table 3: Marginal Likelihoods and Likelihoods Evaluated at Posterior Modes for the Linear and Nonlinear Specifications with Different Outcome Variables

Outcome Variable	$ml_{Linear}$	$l_{Linear}$	$ml_{Nonlinear}$	$l_{Nonlinear}$	Threshold
Output Growth	-997.18	-1461.28	-673.69	-1338.53	-0.21 (0.37)
Consumption Growth	-851.95	-1270.47	-576.91	-1209.77	-0.54 (0.37)
Investment Growth	-2011.98	-2561.17	-1473.50	-2216.87	-1.39 (1.32)
Export Growth	-6414.28	-6962.87	-4060.67	-4812.77	-1.00 (0.12)
Import Growth	-7011.89	-7519.62	-4311.90	-5062.58	-0.18 (0.37)
Unemployment Rate	-813.43	-1205.82	-549.33	-1197.53	-0.46 (0.39)
Employment Growth	-802.77	-1190.28	-544.12	-1199.82	-0.51 (0.45)
Inflation	-1563.58	-2064.27	-1156.22	-1995.35	-0.52 (0.35)

Notes: “ $ml$ ” denotes the natural logarithm of the marginal likelihood and “ $l$ ” denotes the natural log likelihood evaluated at parameter values set to the modes of their marginal posterior distributions. The threshold variable is mean-adjusted capacity utilization.

## D Figures

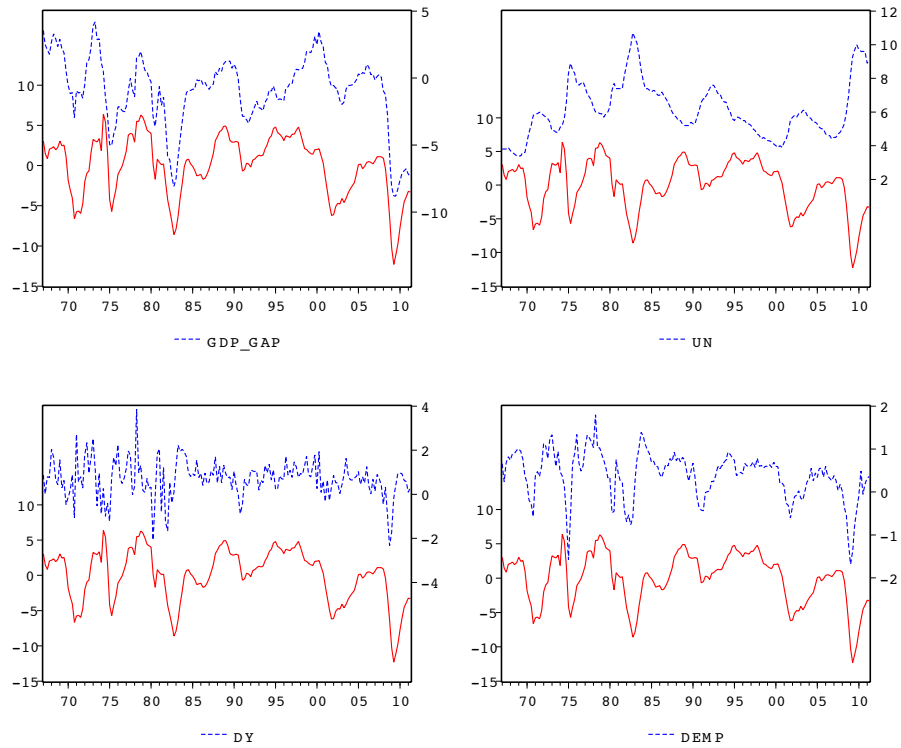


Figure 1: Mean-Adjusted Capacity Utilization versus Other Measures of Slack

Notes: Capacity Utilization (solid red) is mean adjusted for a structural break. GDP\_GAP (dashed blue) is the percentage difference between actual real GDP and Congressional Budget Office estimate of potential output. UN (dashed blue) is the civilian unemployment rate. DY (dashed blue) is the 100 times the log change in real GDP. DEMP (dashed blue) is 100 times the log change in employment.

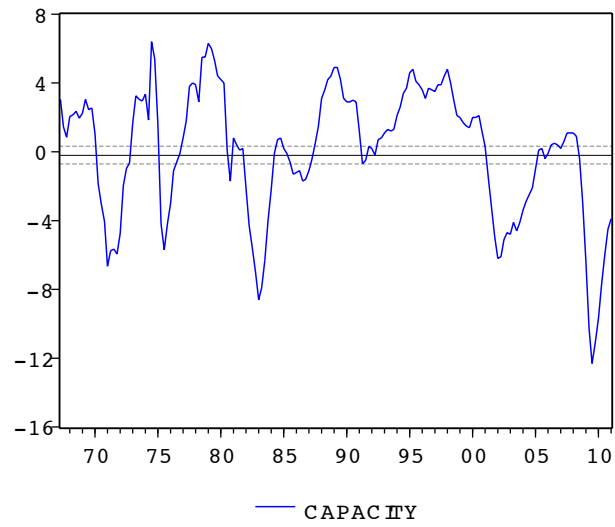


Figure 2: Mean-Adjusted Capacity Utilization and Estimated Threshold (with 90% credibility interval)



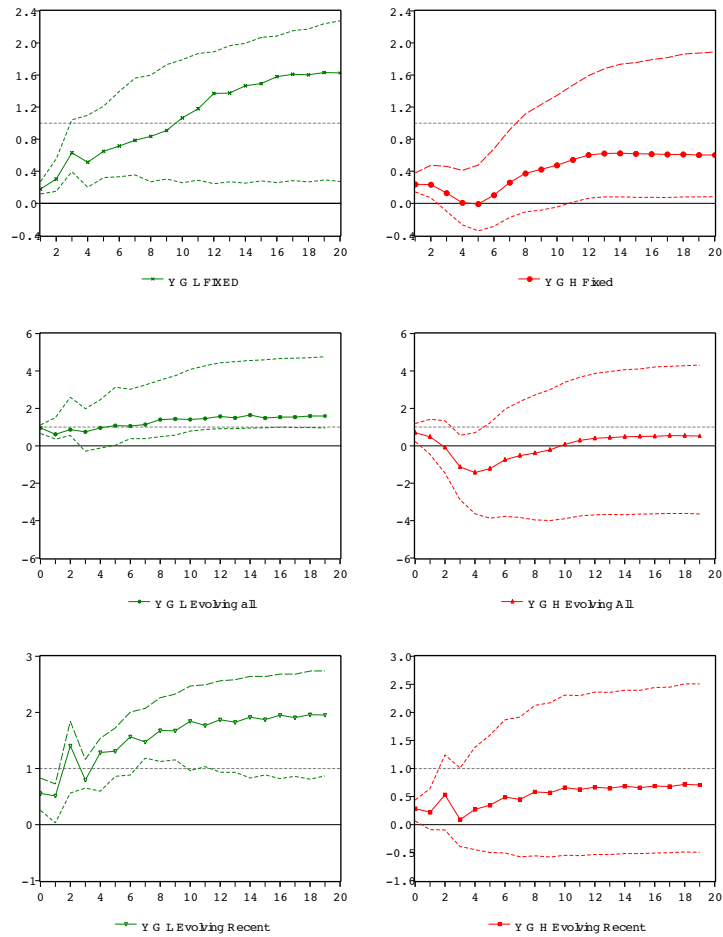


Figure 3: Responses of Output to a Government Spending Shock

Notes: Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: Low Initial State, Right: High Initial State. Top: Fixed States, Middle: Evolving States, averages over all histories (1967-2011), Bottom: Evolving States, averages over recent histories (1984-2011)

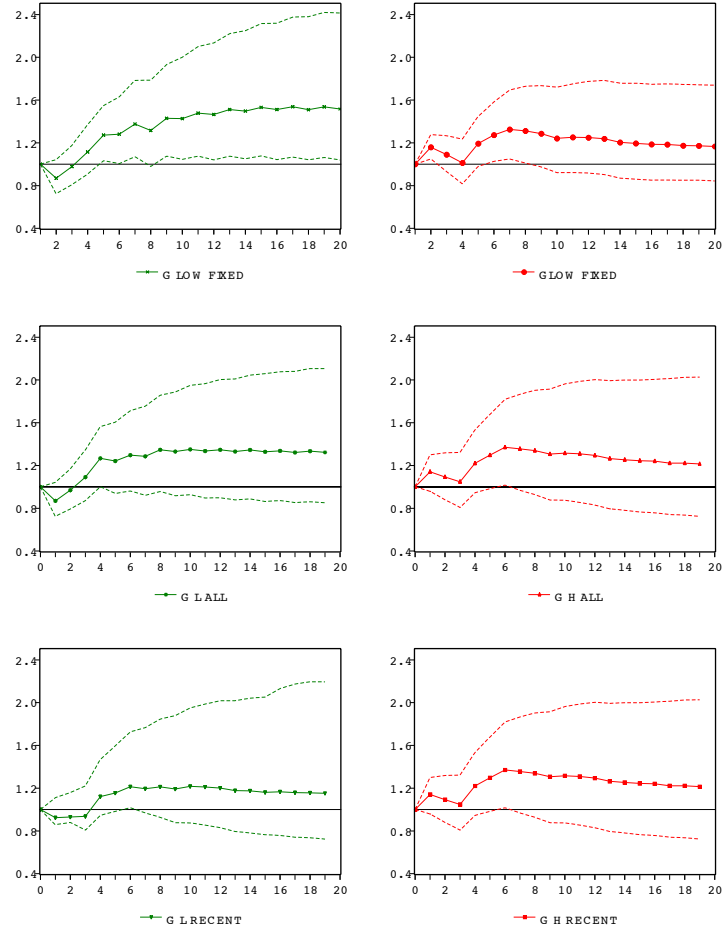


Figure 4: Responses of Government Spending to a Government Spending Shock

Notes: Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: Low Initial State, Right: High Initial State. Top: Fixed States, Middle: Evolving States, averages over all histories (1967-2011), Bottom: Evolving States, averages over recent histories (1984-2011)

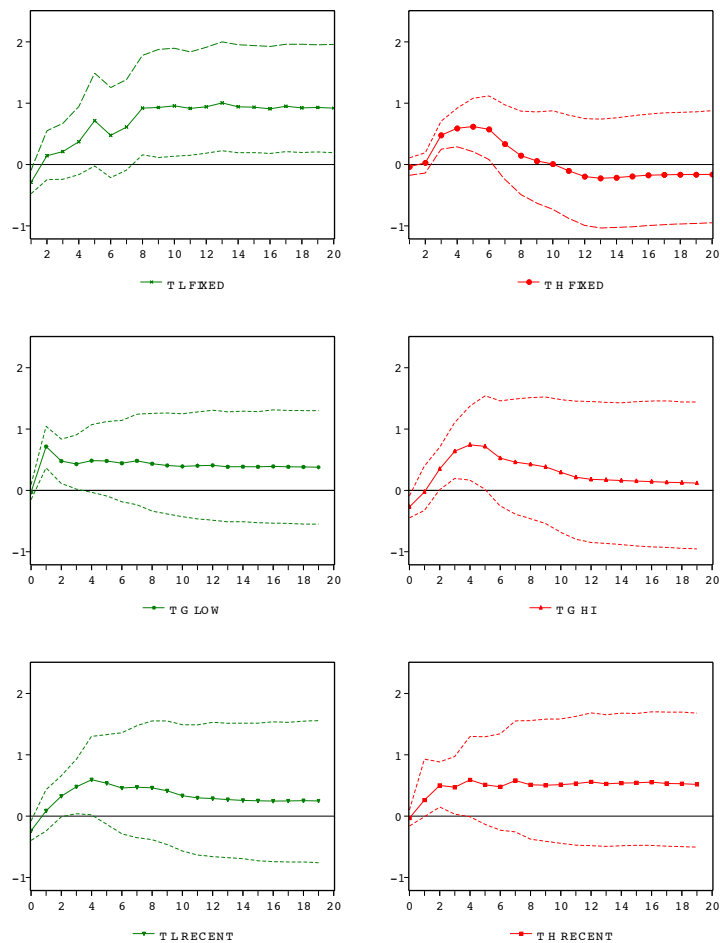


Figure 5: Responses of Tax Revenues to a Government Spending Shock

Notes: Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: Low Initial State, Right: High Initial State. Top: Fixed States, Middle: Evolving States, averages over all histories (1967-2011), Bottom: Evolving States, averages over recent histories (1984-2011)

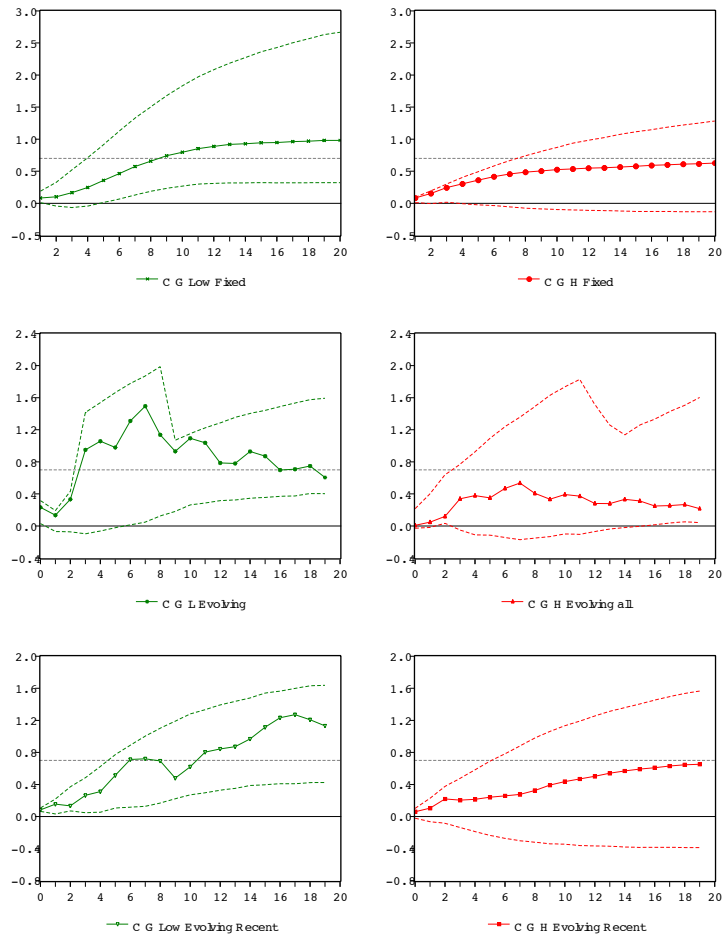


Figure 6: Responses of Consumption to a Government Spending Shock

Notes: Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: Low Initial State, Right: High Initial State. Top: Fixed States, Middle: Evolving States, averages over all histories (1967-2011), Bottom: Evolving States, averages over recent histories (1984-2011)

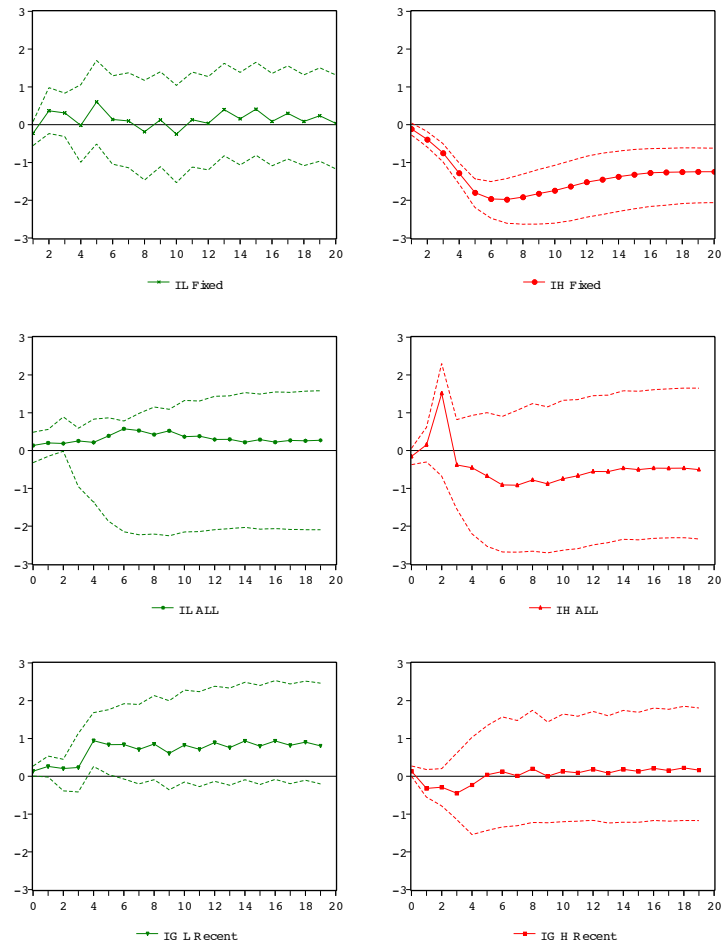


Figure 7: Responses of Investment to a Government Spending Shock

Notes: Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: Low Initial State, Right: High Initial State. Top: Fixed States, Middle: Evolving States, averages over all histories (1967-2011), Bottom: Evolving States, averages over recent histories (1984-2011)

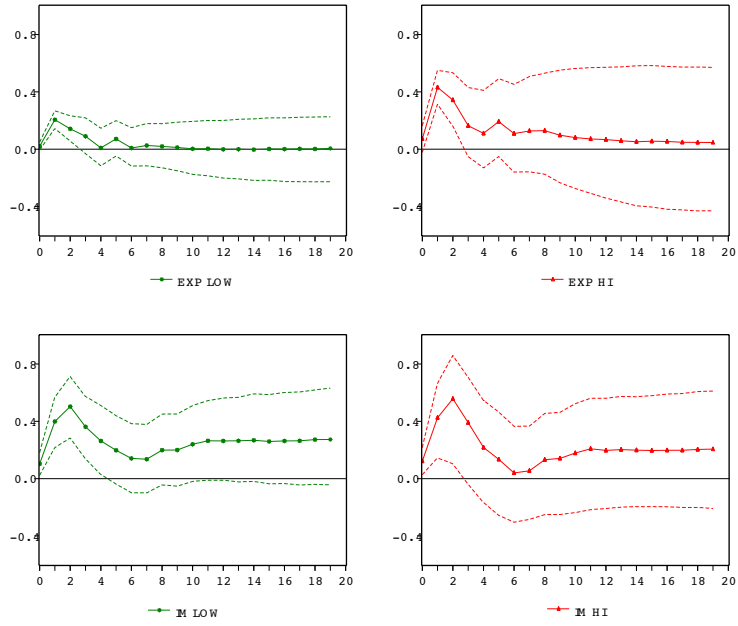


Figure 8: Fixed-Regime Responses of Exports and Imports to a Government Spending Shock

Notes: Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: Low Regime, Right: High Regime. Top: Exports, Bottom: Imports

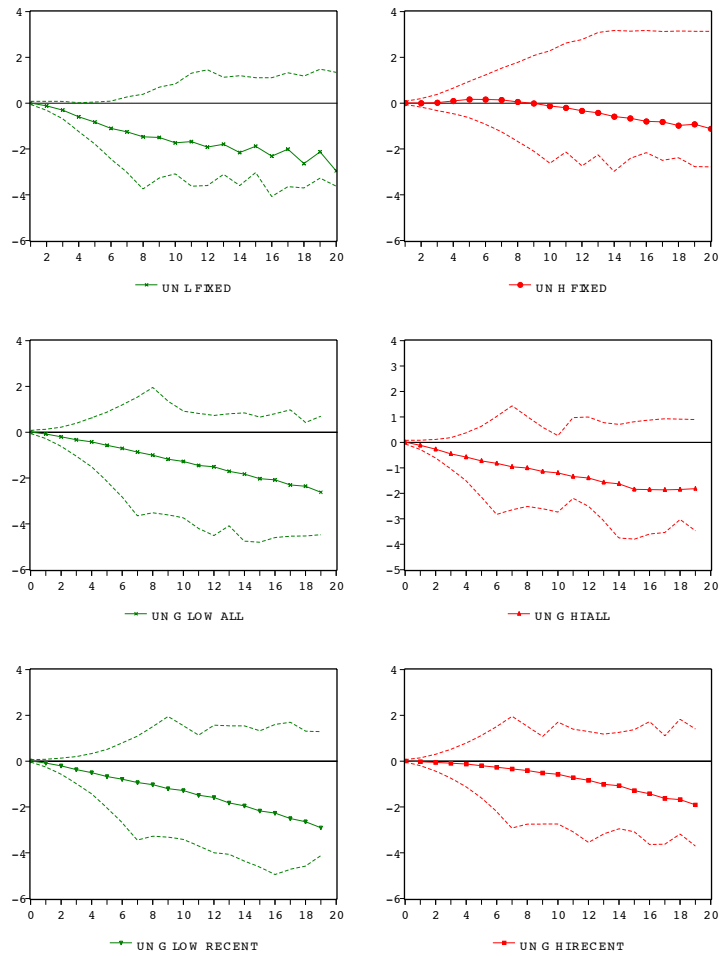


Figure 9: Responses of the Unemployment Rate to a Government Spending Shock

Notes: Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: Low Initial State, Right: High Initial State. Top: Fixed States, Middle: Evolving States, averages over all histories (1967-2011), Bottom: Evolving States, averages over recent histories (1984-2011)

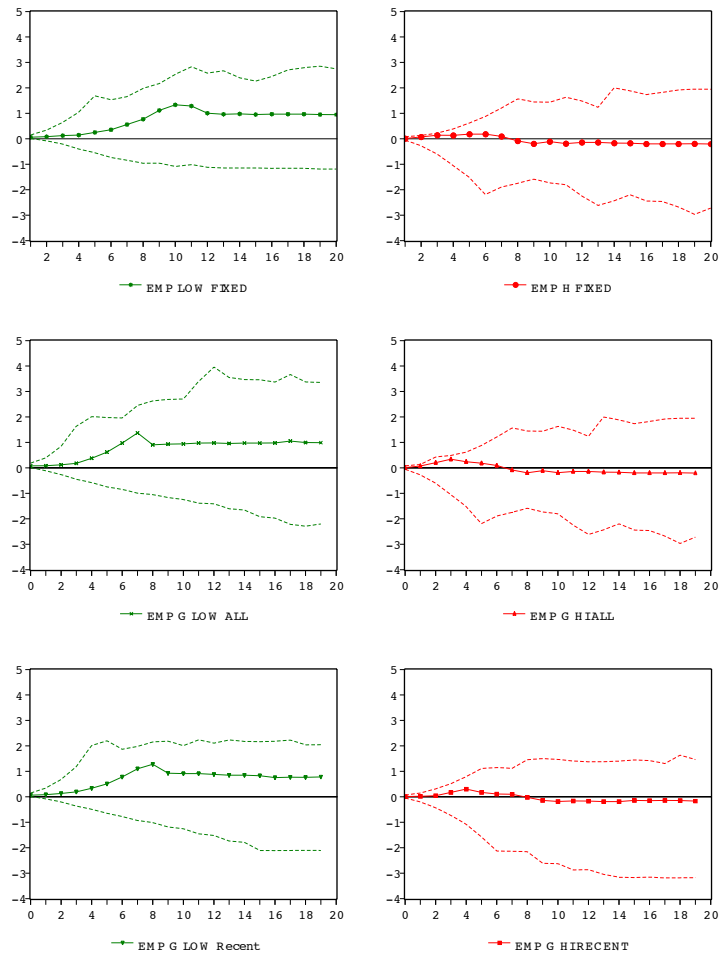


Figure 10: Responses of Employment to a Government Spending Shock

Notes: Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Left: Low Initial State, Right: High Initial State. Top: Fixed States, Middle: Evolving States, averages over all histories (1967-2011), Bottom: Evolving States, averages over recent histories (1984-2011)



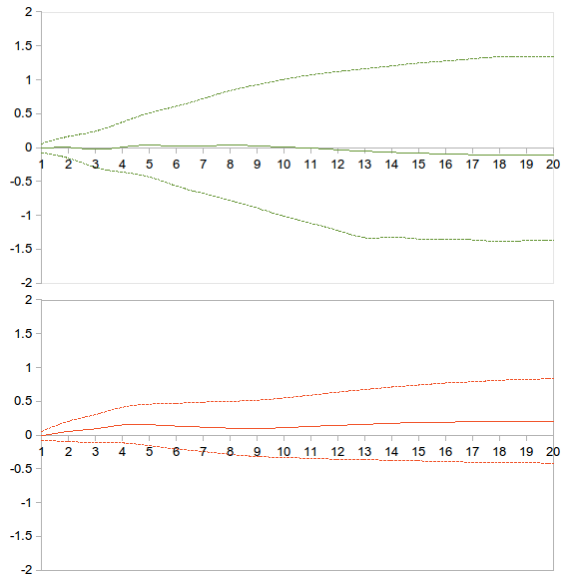


Figure 11: Fixed-Regime Responses of Inflation to a Government Spending Shock

Notes: Modal responses (solid) with equal-tailed 90% credibility bands (dashed). Top: Low Regime, Bottom: High Regime.

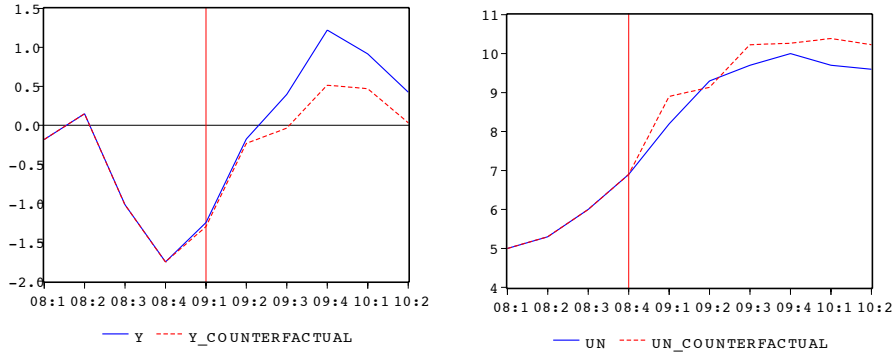


Figure 12: Counterfactual Output and Unemployment Rate following the ARRA Stimulus

Notes: Actual Path (solid blue), Median Simulated Counterfactual Path (dashed red).