The Fiscal Multiplier and the State of Public Finances*

Szabolcs Deák†
Bocconi University
European University Institute

Andreja Lenarčič‡
Bocconi University

December 28, 2012

JOB MARKET PAPER

Abstract

Can fiscal policy always stimulate output? We address this question empirically by estimating a regime-switching VAR using U.S. data where the size of the fiscal multiplier is conditional on the state of public finances. We make two contributions. First, we address the question what proxy we should use for the state of public finances to define the regimes of our model. We estimate several model specifications which differ in the conditioning variable and find that a model with the debt-to-GDP ratio as a conditioning variable fits the data better than other nonlinear specifications or the linear model. Second, we estimate fiscal multipliers conditional on the debt-to-GDP ratio. We find strong asymmetries in the response of output across regimes and a negative relationship between the fiscal multiplier and the debt-to-GDP ratio. This implies that the use of deficit financed fiscal stimulus is characterized by diminishing returns since it increases the debt-to-GDP ratio and, consequently, leads to a decrease in the fiscal multiplier.

Keywords: fiscal policy, fiscal multiplier, public debt, regime-switching threshold autoregression

JEL Classification: E62, C34, H60

*Early drafts of this paper were circulated under the title "The Macroeconomic Effects of Fiscal Policy Shocks in Good Times and Bad". We are indebted to Carlo Favero, Ana Galvão, Francesco Giavazzi, and Massimiliano Marcellino for sharing their data and computer codes. We are also grateful to Gianni Amisano, Michael Ehrmann, Carlo Favero, Francesco Giavazzi, Róbert Lieli, Helmut Lütkepohl, Roberto Perotti, Morten Ravn, Luca Sala and seminar participants at Università Bocconi, EIEF, ECB, MNB and EUI for helpful comments. Parts of this paper was written while the authors were visiting the European Central Bank, whose hospitality is gratefully acknowledged.

†Jean Monnet Postdoctoral Fellow, Robert Schumann Center for Advanced Studies, European University Institute, Florence. I am very grateful to my advisor, Marco Maffezzoli, committee members, Carlo Favero and Francesco Giavazzi, and my mentor during my postdoctoral fellowship, Massimiliano Marcellino, for their guidance and encouragement. E-mail: szabolcs.deak@eui.eu

‡PhD candidate, Department of Economics, Università Commerciale Luigi Bocconi, Milan. E-mail: andreja.lenarcic@phd.unibocconi.it

1 Introduction

The recent financial and economic crisis has put considerable strains on public finances in many countries. Large government bailouts, declining tax revenues, and the workings of the automatic stabilizers have all contributed to a series of large deficits and, consequently, to a quick accumulation of debt. At the same time, policy makers turned towards fiscal policy as an alternative tool to spur a recovery as the scope of conventional monetary policy to provide stimulus became limited by the zero lower bound on the nominal interest rate. However, the literature provides little empirical evidence on the efficacy of fiscal stimulus implemented when the state of public finances is deteriorating.

In this paper we investigate empirically whether the ability of fiscal policy to stimulate output depends on the state of public finances. We use a regime switching structural VAR to estimate state-dependent fiscal multipliers—defined as the dollar response of output to an exogenous dollar change in government expenditures or revenues—for the United States for the period 1960-2007.

We estimate state-dependent fiscal multipliers in two steps. First, we address the question which variable we should use as a proxy for the state of public finances to define the regimes of our model and, consequently, to condition the fiscal multipliers on. We estimate several alternative specifications of our regime switching structural VAR that differ in the conditioning variable we use and compare them in terms of their fit to the data. Motivated by the flow and intertemporal budget constraints of the government we use conditioning variables related to the cost of servicing the debt, the stationarity of debt dynamics, the stock of accumulated debt, and the primary deficit. Furthermore, we use conditioning variables that are meant to capture nonlinearities arising from sovereign default risk premium. While we believe that investors considered sovereign default risk negligible in the United States over our sample period, we do not want to exclude this possibility a priori.

We find that the model with the debt-to-GDP ratio as a conditioning variable fits the data best. The estimated threshold value of debt that triggers the regime switch is at 42.5 percent of GDP and it splits the sample such that approximately one quarter of the observations are during bad times, a term we will use to refer to the regime characterized by high debt-to-GDP ratio.[1] We perform two nonlinearity tests to confront our benchmark nonlinear model to the linear specification and both tests clearly reject the linear model in favor of the benchmark nonlinear one.

Second, we compute state-dependent fiscal multipliers based on the impulse responses of our benchmark nonlinear model. The main distinguishing feature of our modeling approach is the endogenous determination of debt dynamics following a fiscal policy shock.

[1]We use the series *Federal Debt Held by the Public* to compute the debt-to-GDP ratio in our model. The observations range between 22.74 and 50.60 percent of GDP in our sample period.

We adopt the specification of Favero and Giavazzi (2007) who augment a linear VAR with the flow government budget constraint. We extend their specification to model endogenous regime switches in a nonlinear VAR and, as far as we can ascertain, ours is the first paper in the literature that can endogenously reconstruct the evolution of debt after a fiscal policy shock in a nonlinear VAR.

The main result of our paper is that the use of deficit financed fiscal stimulus is characterized by diminishing returns. The strong asymmetries in the response of output across regimes indicate a negative relationship between the fiscal multiplier and the debt-to-GDP ratio. Every additional dollar spent by the government increases output by 0.97 dollar less in bad times than in good, while the corresponding difference for a dollar decrease in government revenues is 1.54 dollar. Diminishing returns arise since deficit financed fiscal stimulus increases the debt-to-GDP ratio and, consequently, leads to a decrease in the fiscal multiplier.

At least three existing strands of models formalize how the effects of fiscal policy shocks vary with the debt-to-GDP ratio. First, potential nonlinearities arise in the presence of sovereign default risk. Bi (2010), Ghosh, Kim, Mendoza, Ostry, and Qureshi (2011) and Juessen, Linnemann, and Schabert (2011), among others, develop models where the interest rate on government bonds increases nonlinearly with the level of government liabilities. At low levels of government debt, interest rates are unresponsive to the level of debt. As the economy approaches its fiscal limit—the level of debt above which it cannot be rolled-over—financial markets start to demand a risk premium on government bonds that is steeply increasing in the level of debt. We find that the data strongly rejects sovereign default risk as the source of nonlinearities in our sample. Hence, the estimated threshold value for debt-to-GDP ratio in our benchmark nonlinear model should not be interpreted as a fiscal limit above which public finances are not sustainable.

Second, Blanchard (1990), Sutherland (1997) and Perotti (1999) illustrate that the wealth effect generated by fiscal policy shocks depends on the level of sovereign debt when taxes are distortionary. A tax cut, for example, represents an intertemporal reallocation of tax distortions in these models via the intertemporal government budget constraint. Because tax distortions are assumed to be convex in the tax rate, the same tax cut induces a higher change in the present discounted value of the tax distortions, and hence a stronger wealth effect, the higher the tax rate is today. The latter is, in turn, a positive function of the initial debt via the intertemporal government budget constraint.

Third, Bohn (1998) provides empirical evidence that fiscal policy in the US is characterized by spending reversals—defined as a systematic negative response of deficit to the level of public debt—that become stronger as the debt-to-GDP ratio increases. Corsetti, Meier, and Müller (2009) demonstrate that incorporating these features into a standard New Keynesian model gives rise to state-dependent fiscal multipliers. A positive expenditure shock in this framework increases aggregate demand in the short-run and at the same time creates expectations of government spending below trend in the future. These expectations lower the long-run interest rate and modify the time profile of the fiscal multiplier: the short-run spending multiplier is increasing while the long-run multiplier is decreasing in the strength of debt stabilization.

The different time profiles of the output response that we find in the two regimes indicate that the theory of Corsetti, Meier, and Müller (2009) based on spending reversals provides the explanation for the asymmetric response of output. Consistently with the prediction of their model the output response to a government expenditure shock in bad times is higher at shorter and lower at longer horizons compared to good times. Moreover, similarly to Bohn (1998), we find evidence that spending reversals are becoming stronger as the debt-to-GDP ratio increases. Consequently, the persistence of the fiscal expansion is lower during bad times. In order to determine whether the observed differences in the fiscal multiplier can be explained by the less persistent fiscal expansion only, we perform a counterfactual analysis in which we control for the differences in the estimated fiscal rules. We find that the estimated differences in the multipliers prevail even after imposing the same fiscal rules in both regimes, which suggests that the differences are driven by the asymmetric response of the private sector.

The main policy implication of our results is that the effects of fiscal stimuli should not be evaluated using linear models when the debt-to-GDP ratio is high. The predictions of our model are similar to the linear model at moderate levels of the debt-to-GDP ratio: fiscal expansion has positive effects on output over time. However, at high levels of debt-to-GDP ratio fiscal policy has no significant effect on output. In a counterfactual exercise we have simulated the effects of The Recovery Act[2] using both our linear and nonlinear models. While both models seem to agree regarding the cost of the stimulus package in terms of accumulated debt, the expansionary effects on output are considerably smaller and insignificant in the nonlinear model.

Although we limit our analysis to estimating fiscal multipliers conditional on the state of public finances, a growing body of the literature shows that other determinants are also important.[3] In particular, Auerbach and Gorodnichenko (2010) provide empirical evidence that the expenditure multiplier in the US is about 2.5 during recessions, while Christiano, Eichenbaum, and Rebelo (2011) find expenditure multipliers around 2.3 when monetary policy is constrained at the zero lower bound. However, it is hard to reconcile

[2]The American Recovery and Reinvestment Act of 2009.

[3]These determinants include the business cycle (Auerbach and Gorodnichenko, 2011; Fazzari, Morley, and Panovska, 2012; Batini, Callegari, and Melina, 2012; Canzoneri, Collard, Dellas, and Diba, 2011; Michaillat, 2012), the exchange rate regime (Corsetti, Meier, and Müller, 2010; Ilzetzki, Mendoza, and Végh, 2010), financial factors (Corsetti, Meier, and Müller, 2010; Kirchner, Cimadomo, and Hauptmeier, 2010; Afonso, Baxa, and Slavík, 2011) and openness to trade (Ilzetzki, Mendoza, and Végh, 2010; Kirchner, Cimadomo, and Hauptmeier, 2010). In addition, a number of papers have analyzed the behavior of the fiscal multiplier when the monetary policy is constrained at the zero lower bound (Cogan, Cwik, Taylor, and Wieland, 2010; Mertens and Ravn, 2010; Eggertsson, 2011; Bilbiie, Monacelli, and Perotti, 2012; Fernández-Villaverde, Gordon, Guerrón-Quintana, and Rubio-Ramírez, 2012).

these findings with the current experience of the US economy. Both the Economic Stimulus Act of 2008 and The Recovery Act provided large fiscal stimuli when the US economy was in a recession and the zero lower bound was binding. Yet the recovery has been weak which indicates that the fiscal multiplier has been much lower during the aftermath of the crisis than what has been suggested by these papers and more in line with our estimates.

The closest to our work are the papers by Choi and Devereux (2005) and Kirchner, Cimadomo, and Hauptmeier (2010) who study how the ability of government expenditures to stimulate output changes with the state of public finances. Choi and Devereux (2005) condition the effects of government expenditure shocks on the real treasury bill rate in order to capture nonlinearities related to the cost of financing the debt in the United States. They find that the ability of fiscal policy to boost economic growth is decreasing in the interest rate which gives rise to very different policy implications from ours. The real interest rate on treasury bills has been very low since the onset of the financial crisis hence their results support the increased use of fiscal measures. Kirchner, Cimadomo, and Hauptmeier (2010) use time-series data for the Euro Area and, in accordance with our results, find a negative relationship between the expenditure multiplier and the debt-to-GDP ratio. According to their results every percentage point increase in the debt-to-GDP ratio decreases the spending multiplier by 0.01 points.[4]

To our best knowledge, we are the first to estimate a threshold value of the debt-to-GDP ratio in a nonlinear model for the US economy. Panel data studies by Perotti (1999), Corsetti, Meier, and Müller (2010), Ilzetzki, Mendoza, and Végh (2010) and Reinhart and Rogoff (2010) all find that higher indebtedness leads to weaker or even contractionary response of output to fiscal stimulus. But contrary to our approach, these studies impose threshold values in the neighborhood of 100 percent of the debt-to-GDP ratio to the whole cross-section of countries uniformly to define the regimes in their models. As such, they are not applicable to the US economy that has historically recorded lower debt-to-GDP ratios compared to other countries in their sample.

The paper is organized as follows. Section 2 presents the empirical methodology used to estimate the reduced form models and the strategy to choose the benchmark specification. Section 3 describes our identification strategy and how the fiscal multiplier is computed. Section 4 discusses the results relevant to the choice of the benchmark specification and section 5 presents the estimated effects of fiscal policy in the linear and nonlinear models. Section 6 examines the robustness of our results. Section 7 concludes.

[4]This is considerably lower than our corresponding estimate of 0.06 which is due to both a smaller historical variation in the debt-to-GDP ratio in the US and a smaller elasticity of the spending multiplier with respect to public debt in the Euro Area.

2 Empirical methodology

This section outlines the empirical methodology used to estimate the reduced form model. First, we describe both the linear and the nonlinear models used in the paper and their estimation. We employ a multivariate endogenous threshold autoregressive model (ET-VAR) to model explicitly the dependence of the transmission mechanism on initial conditions that differ across regimes. We discuss the methodology in terms of the two regime model, but everything can be readily extended to a model with more regimes. Then, we briefly describe how the endogenous variables of the model are constructed and we give details of the threshold variables we consider as proxies for the state of public finances. Finally, we discuss our strategy to choose our benchmark specification.

2.1 The (almost) linear model

In order to keep our exercise comparable to the existing literature, we estimate a type of structural VAR estimated by Blanchard and Perotti (2002) and Perotti (2004). In particular, we adopt the specification of Favero and Giavazzi (2007) because it combines a linear VAR with the flow government budget constraint within the same model and can be easily extended to model endogenous regime switches in a nonlinear VAR. While the model can be estimated as any other linear VAR the nonlinearity of the flow government budget constraint makes the model behave as a nonlinear model when computing the impulse response functions.

The model consists of two parts. The first is a standard linear autoregression of the form:

$$\mathbf{Y}_t = \mathbf{c} + \mathbf{\Phi}\mathbf{X}_t + \mathbf{\Gamma}\mathbf{D}_t + \varepsilon_t \tag{1a}$$

where $\mathbf{Y}_t$ is an $n \times 1$ vector of endogenous variables, $\mathbf{X}_t = [\ \mathbf{Y}'_{t-1} \ \cdots \ \mathbf{Y}'_{t-p}\ ]'$ is an $np \times 1$ vector of lagged values, $\mathbf{D}_t = [\ d_{t-1} \ \cdots \ d_{t-k}\ ]'$ is a $k \times 1$ vector of lagged values of the level of the debt-to-GDP ratio, $\mathbf{c}$, $\mathbf{\Phi}$ and $\mathbf{\Gamma}$ are $n \times 1$, $n \times np$ and $n \times k$ coefficients matrices. The second part of the model is the flow government budget constraint:

$$d_t = \frac{1+i_t}{1+\Delta p_t}\frac{1}{1+\Delta y_t}d_{t-1} + \frac{\exp(g_t) - \exp(t_t)}{\exp(y_t)} \tag{1b}$$

where i_t is the nominal interest rate on government debt, Δp_t is inflation, y_t is log real GDP, g_t is log government expenditure net of interest payments and t_t is log government revenues net of interest receipts.

To close the model, Favero and Giavazzi (2007) defines the vector of endogenous variables to be $\mathbf{Y}_t = [\ g_t \ \ t_t \ \ y_t \ \ \Delta p_t \ \ i_t\ ]'$. This way the dynamics of all the variables that enter the right-hand side of (1b) are modeled by the autoregression in (1a) and the model endogenizes debt dynamics in a way that is consistent with the flow government budget constraint.[5]

Notice that the flow government budget constraint (1b) is an identity and has no parameters to be estimated. Since (1a) is linear in its coefficients it can be estimated as any other linear VAR. However, the endogenous variables depend on the lagged values of the debt-to-GDP ratio that is, in turn, a nonlinear function of the lagged endogenous variables. This implies that we need resort to simulation to compute the impulse response functions of the endogenous variables.

2.2 The nonlinear model

Multivariate threshold autoregressive models (TVARs) combine two piecewise linear models with different sets of coefficients over two subsamples (regimes) into a nonlinear VAR (Tsay, 1998). The two regimes are determined by an observed threshold variable, a value of that threshold variable that separates the two regimes and a delay parameter. The linear model of the previous section generalizes to the following nonlinear model:

$$\mathbf{Y}_t = \begin{cases} \mathbf{c}^{(1)} + \mathbf{\Phi}^{(1)}\mathbf{X}_t^{(1)} + \mathbf{\Gamma}^{(1)}\mathbf{D}_t + \varepsilon_t^{(1)} & \text{if } z_{t-d} \leq r \\ \mathbf{c}^{(2)} + \mathbf{\Phi}^{(2)}\mathbf{X}_t^{(2)} + \mathbf{\Gamma}^{(2)}\mathbf{D}_t + \varepsilon_t^{(2)} & \text{if } z_{t-d} > r \end{cases} \tag{2a}$$

$$d_t = \frac{1+i_t}{1+\Delta p_t}\frac{1}{1+\Delta y_t}d_{t-1} + \frac{\exp(g_t) - \exp(t_t)}{\exp(y_t)} \tag{2b}$$

where z_t is the threshold variable, d is the delay parameter and r is the threshold value that triggers a regime switch. Each regime has a different set of coefficients indexed by superscript (j), possibly different number of endogenous lags (p_j) and we also allow for regime specific covariance matrices for the residuals, that is $\varepsilon_t^{(j)} \sim N(0, \mathbf{\Sigma}^{(j)})$.

For a given threshold variable and lag lengths p_j and k we concentrate the log likelihood in three steps to estimate the model. First, for given values of the threshold variable, r, and the delay parameter, d, the model reduces to two linear VARs. The (regime specific) coefficient and covariance matrices of the two piecewise linear models, $\hat{\mathbf{c}}^{(j)}(r,d)$, $\hat{\mathbf{\Gamma}}^{(j)}(r,d)$, $\hat{\mathbf{\Phi}}^{(j)}(r,d)$ and $\hat{\mathbf{\Sigma}}^{(j)}(r,d)$, can be estimated with least squares formula using observations from regime j.

Second, for a given value of the delay parameter d we estimate the threshold value by maximizing the concentrated log likelihood over a grid of values for the threshold value. Galvão (2006) shows that maximum likelihood works better than least squares estimation

[5] Chung and Leeper (2007), instead, include the debt in their VAR as one of the endogenous variables and impose the intertemporal budget constraint of the government as a set of cross-equation restrictions. The advantage of the specification of Favero and Giavazzi (2007) is that the flow government budget constraint (1b) is an identity and has no parameters to be estimated. We need to estimate the parameters of five equations only which makes the regime-switching model computationally more feasible.

when the covariance matrices are regime specific, hence $\hat{r}(d)$ is obtained as

$$\hat{r}(d) = \arg\min_{r \in R} \sum_{j=1}^{2} \frac{T_j}{2} \log |\hat{\mathbf{\Sigma}}^{(j)}(r,d)| = \arg\min_{r \in R} \sum_{j=1}^{2} \frac{T_j}{2} \log \left| \frac{1}{T_j} \sum_{i=1}^{T_j} \hat{\varepsilon}_t^{(j)}(r,d) \hat{\varepsilon}_t^{(j)}(r,d)' \right|$$

where $|\hat{\mathbf{\Sigma}}^{(j)}(r,d)|$ is the determinant of the estimated covariance matrix, T_j is the number of observations in regime j and R denotes the grid of values for the threshold value. We form this grid using all observations of the threshold variable excluding the lowest and highest 20 percent of the observations. This ensures that we have at least 20 percent of the observations in each regime.[6]

Third, the delay parameter is estimated by maximizing the concentrated log likelihood over the values $D = \{1, \ldots, 4\}$:[7]

$$\hat{d} = \arg\min_{d \in D} \sum_{j=1}^{2} \frac{T_j}{2} \log |\hat{\mathbf{\Sigma}}^{(j)}(\hat{r}(d), d)|$$

An alternative specification used by Auerbach and Gorodnichenko (2010) is the smooth transition autoregressive (STAR) model where the two regimes are not mutually exclusive as in the case of TVARs. The dynamics of the endogenous variables are driven by a weighted average of the two piecewise linear functions, where weights are given by the logistic function $[1 + \exp(-\gamma(z_{t-d} - r))]^{-1}$. It allows a smooth transition even with two regimes and may give a better approximation to the underlying nonlinearity than an ET-VAR model with several regimes. However, in short samples that are typically available in macroeconomics the curvature parameter, γ, and/or the threshold value, r, can be estimated only very imprecisely. Auerbach and Gorodnichenko (2010) overcome this problem by calibrating rather than estimating the curvature parameter using the NBER classification for recessions as an exogenous source of information. We have chosen the ET-VAR model over the STAR model because it allows us to estimate the parameters governing the transition between regimes instead of imposing them.[8]

2.3 Endogenous variables

As mentioned before we adopt the specification of Favero and Giavazzi (2007) both in terms of endogenous variables and identification strategy. The specification includes quar-

[6]The typical choice in applications is 10 or 15 percent. Our more conservative choice reflects the fact that we have a VAR with five endogenous variables and potentially up to four lags are included in the regressions. We also experimented both with higher and lower values, but the results are unaffected.

[7]The last two steps are equivalent to maximizing the concentrated likelihood over the two dimensional grid $D \times R$. We chose to estimate them in two steps only for convenience.

[8]When Auerbach and Gorodnichenko (2010) estimate all the parameters of their model jointly they find very high point estimate for the curvature parameter γ. These high values indicate that the data can be well described by a model where regime switches occur sharply at certain threshold values.

terly US data on (federal) government total expenditures net of interest payments (g_t), (federal) government total receipts net of interest receipts (t_t), and GDP (y_t), all in per capita real terms, the GDP deflator inflation rate (Δp_t) and the average nominal cost of financing the debt (i_t).[9] All variables are in logs except the interest rate, which enters in levels. The specification includes a constant, one lag of the debt-to-GDP ratio and up to four lags of the endogenous variables. The full sample goes from 1960:1 until 2007:4.[10]

While Favero and Giavazzi (2007) include the same set of variables in their VAR as Perotti (2004), some differences in the definition of the variables are worth pointing out. Transfer payments are considered as part of government expenditure, rather than being subtracted from government receipts. Also, the nominal cost of servicing the debt is used as the interest rate instead of the yield to maturity on long-term government bonds. It is defined as the ratio of net interest payments and the end of last period stock of government debt. Furthermore, since the definition of debt refers to federal government debt we use only federal government expenditures and receipts to construct the endogenous variables.[11]

2.4 Threshold variables considered

In this subsection we attempt to assemble a list of potential proxies for the state of public finances to use as threshold variables in our model. We collect these variables into five groups (Table 1). We motivate this grouping by the flow and the intertemporal budget constraints of the government.

[Table 1 about here.]

The first term on the right-hand side of the flow budget constraint (1b), which is repeated here for convenience

$$d_t = \frac{1+i_t}{1+\Delta p_t}\frac{1}{1+\Delta y_t}d_{t-1} + \frac{\exp(g_t)-\exp(t_t)}{\exp(y_t)} \tag{1b}$$

is the real cost of servicing the debt and we include three variables related to it. The real 3-month treasury bill rate is used by Choi and Devereux (2005) as a direct measure of the

[9] We use the GDP deflator for all variables to obtain the corresponding real values. For details on the construction of the variables and the data sources, see our Appendix or Favero and Giavazzi (2007).

[10] Some of the threshold variables discussed later have smaller samples as a consequence of differencing. Computing year-on-year differences, for example, forces us to drop three observations from the beginning of the sample. When we estimate several competing specifications we restrict the sample to be the same for all models.

[11] Favero and Giavazzi (2007) carefully check whether these differences in data definition alter the estimated effects of fiscal policy shocks. They conclude that the impulse responses are similar to the results of Perotti (2004) both in their full sample (1960:1-2006:2) and in their two subsamples (1960:1-1979:4 and 1980:1-2006:2). We carry out a similar comparison exercise in order to check the effects of data definition on the impulse responses in our slightly longer sample and arrive to the same conclusion.

cost of financing the debt. We also consider the real cost of financing the debt and the real 10-year government bond rate among our candidates. The former is directly related in our model to the cost of financing the debt and can be computed from our model variables as $(1 + i_t)/(1 + \Delta p_t)$. The latter is a more forward looking variable than the short term rate and can be a more sensitive proxy to expectations about the sustainability of fiscal policy.

The first two terms together, the relationship between the real interest rate and the growth rate of real GDP, determine the stationarity of the accumulation equation. When the real interest rate is low relative to the growth rate of the economy, the debt stock is falling behind GDP and the debt-to-GDP ratio is decreasing. When interest rates are high relative to the growth rate of the economy, then debt is growing faster than GDP and an expansionary fiscal shock accelerates the growth of the debt-to-GDP ratio even further. Hence, this difference is often used to evaluate fiscal sustainability (see for example Callen, Terrones, Debrun, Daniel, and Allard, 2003; IMF, 2009; ECB, 2011). We include three variables measuring the difference between the real interest rate and the growth rate of the economy where the interest rate is defined by the short-run real interest rate (3-month T-bill rate), the long-run real interest rate (10-year government bond rate) and the real cost of financing the debt.

We include three variables related to the driving force of equation (1b), the primary deficit-to-GDP ratio. The quarterly primary deficit-to-GDP ratio and the annual primary deficit-to-GDP ratio are used in Giavazzi and Pagano (1995) to define protracted and sizable budget cuts or expansions. The worse of the last two periods' quarterly primary deficit-to-GDP ratio follows Perotti's (1999) definition of the bad times dummy.

Iterating the flow budget constraint forward we obtain the intertemporal budget constraint of the government:

$$d_t = \sum_{s=1}^{\infty} \left[\prod_{j=1}^{s} \frac{1 + i_{t+j}}{1 + \Delta p_{t+j}} \frac{1}{1 + \Delta y_{t+j}} \right]^{-1} \frac{\exp(t_{t+s}) - \exp(g_{t+s})}{\exp(y_{t+s})} \tag{3}$$

which shows that the debt-to-GDP ratio can be interpreted as the PDV of the future surpluses of the government from the perspective of the current period. We include four variables related to it. The papers by Perotti (1999), Giavazzi, Jappelli, and Pagano (2000), Giavazzi, Jappelli, Pagano, and Benedetti (2005) and Corsetti, Meier, and Müller (2010) are all using the debt-to-GDP ratio as a conditioning variable.[12] Coenen, Straub, and Trabandt (2012) assume fiscal rules in their model where fiscal instruments react to, among others, the cyclical component of real per capita debt. Additionally, we consider the change in the debt-to-GDP ratio which has been used as a potential source of

[12]The time series for debt held by the public is available only from 1970:1 on the FRED website. We follow Favero and Giavazzi (2007) and use the observation for 1970:1 as an initial value and the debt dynamics equation (1b) to construct a debt-to-GDP series that covers our entire sample.

nonlinearity in a number of papers by Giavazzi and Pagano (1990, 1995); Alesina and Perotti (1996a,b); Giavazzi, Jappelli, and Pagano (2000); Giavazzi, Jappelli, Pagano, and Benedetti (2005) and more recently by Burriel, de Castro, Garrote, Gordo, and Prez (2009).[13] Finally, big and persistent year-on-year declines in the debt-to-GDP ratio were used in a recent paper by Nickel, Rother, and Zimmermann (2010).

The intertemporal budget constraint of the government is valid only if the government is willing and able to honor its debt. While we believe that investors considered sovereign default risk negligible in the United States over our sample period, we do not want to exclude this possibility a priori. We condition on five variables that are potential indicators of either the government's ability to service its debt or investors' perception of sovereign risk. Following Haugh, Ollivaud, and Turner (2009) we use the ratio of debt interest payments-to-government receipts to proxy for investor assessments of sovereign risk. As a joint proxy for the first three terms in (1b), we consider the ratio of debt interest payments-to-GDP as an indicator for government indebtedness. The last three variables in this category are related to the fiscal limit of the economy.[14] The models of Bi (2010), Ghosh, Kim, Mendoza, Ostry, and Qureshi (2011) and Juessen, Linnemann, and Schabert (2011) predict that the interest rate rises nonlinearly with the level of government liability. When the economy is below the fiscal limit, the sovereign risk premium is stable and interest rates vary little with government debt. As the economy approaches its fiscal limit the probability of sovereign default increases and financial markets start to demand a premium for government bonds. This mechanism increases the slope of the pricing rule of the interest rate that links government debt to the interest rate as the economy approaches its fiscal limit.[15]

2.5 Choosing between different specifications

We need to compare competing specifications along three different dimensions. First, given a choice of a threshold variable we need to select the lag lengths in the two regimes. Second, we need to compare models that differ in the threshold variable. Third, we need to confront the best fitting nonlinear model with the linear benchmark.

We base the first two choices on a penalized likelihood function. We follow the ap-

[13] Favero and Giavazzi (2007) include two lags of the debt-to-GDP ratio in their linear VAR as exogenous regressors and cannot reject the restriction that the two coefficients on d_{t-1} and d_{t-2} are equal in size with opposite signs thus it is the first difference of the debt-to-GDP ratio that enters their specification.

[14] The fiscal limit is defined as the level of debt that the government is able and willing to service. When the debt level exceeds the fiscal limit, a sovereign default occurs.

[15] For the nonlinear relationship described see Figure 6 in Bi (2010) and the discussion therein. Note that the interest rate rule depicted in the graph looks very similar to a threshold function. It shows a very sharp increase in the interest rate at the fiscal limit while it is linear away from it. Thus, the slope of the interest rate has only two values in her model. This is a direct consequence of the assumption that the default rate, δ, is constant and exogenous. If the default rate was an endogenous function of the macroeconomic fundamentals as in Juessen, Linnemann, and Schabert (2011), then the interest rate rule would take off more gradually.

proach of Artis, Galvão, and Marcellino (2007) and Galvão and Marcellino (2010) for comparing competing specifications that differ in terms of the threshold variable. They use information criteria based on a penalized likelihood function, where the penalty depends on the number of estimated parameters, in particular the Akaike, Hannan and Quinn, and Schwarz information criteria. The results of Gonzalo and Pitarakis (2002) show that the most reliable information criterion to choose among the models is the one with the heaviest penalty function, i.e. the Schwarz criterion.

Following standard practice in time series analysis, we also use the information criteria to select the best lag structure of our model. First, we set the lag length of the exogenous regressor d_t to $k = 2$. Our results show that all the information criteria are increasing with the lag length k, but for values above 2 the gain is negligible. Hence estimating additional parameters do not justify the additional computational costs while they decrease the power of the estimation substantially (Hansen, 1996). Second, in each regime the lag length can take a value from the set $P = \{1, \ldots, 4\}$. We allow for different lag lengths in the two regimes and thus for each threshold variable we estimate a model for all possible lag length combinations over the two dimensional grid $P \times P$ and select the combination that minimizes the given information criterion.

When comparing the best fitting nonlinear specification to the linear model we rely on two statistical tests. The first test is a variable addition test, which considers as the nonlinear alternative the specification

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}\mathbf{X}_t + \mathbf{\Gamma}\mathbf{D}_t + \mathbf{\Psi}\begin{bmatrix}\mathbf{X}_t' & \mathbf{D}_t'\end{bmatrix}' z_{t-d} + \varepsilon_t$$

given the value of the delay parameter, d (see for example Artis, Galvão, and Marcellino, 2007; Teräsvirta, 1998). If the true model is linear, then the coefficients in $\mathbf{\Psi}$ are jointly insignificant which can be tested using an LR test.

The second test was proposed by Tsay (1998) and it uses predictive residuals from an arranged model to construct a test statistic. Given a threshold variable, z, and a value for the delay parameter, d, the observations in the linear model are arranged according to the increasing ordering of z_{t-d}. Then a series of linear models for each value $m = m_0, \ldots, T$ are estimated using observations $i = 1, \ldots, m$ from the arranged sample to construct one step ahead prediction errors.[16] If the data were generated by a threshold model, then this arrangement transforms the model into a structural break model with observations for regime 1 at the beginning and for regime 2 at the end of this arranged sample and a structural break at an unknown date in between. In this case the predictive errors are correlated with the regressors of the arranged model. If, instead, the data were generated by a linear model, then the predictive residuals are uncorrelated with the

[16] m_0 is the sample size of the first model and T is the size of the full sample. On the selection of m_0 and for details of the procedure see Tsay (1998).

arranged regressors and the coefficients of the regression of the predictive residuals on the arranged regressors should be jointly insignificant under the null.

There are several alternatives to these two tests in the literature (see for example Andrews and Ploberger, 1994; Hansen, 1999; Altissimo and Corradi, 2002). We have chosen these two tests because they rely on the choice of the threshold variable to construct a test statistic, but neither the value nor the distribution of the test statistic depend on the threshold value, r, a nuisance parameter which is present only under the alternative. Thus, both of these tests are simple and have familiar limiting distributions. Furthermore, as pointed out by Galvão and Marcellino (2010), applying these alternative tests in a multivariate setting may be misleading when the variance of the disturbances is regime specific.

3 Estimating the fiscal multiplier

Obtaining estimates for the fiscal multiplier requires two steps. First, we need to identify the structural shocks in the estimated models. Second, we need to compute impulse response functions to the identified fiscal policy shocks.

3.1 Identification

We identify the structural shocks for the fiscal variables separately for each of the regimes using the Blanchard and Perotti (2002) identification approach, extended by Perotti (2004) for the five variable VAR. We impose the relationship

$$\left[\begin{array}{cc|ccc} 1 & 0 & -\alpha_{gy}^{(j)} & -\alpha_{g\Delta p}^{(j)} & -\alpha_{gi}^{(j)} \\ 0 & 1 & -\alpha_{ty}^{(j)} & -\alpha_{t\Delta p}^{(j)} & -\alpha_{ti}^{(j)} \\ \hline a_{31}^{(j)} & a_{32}^{(j)} & 1 & 0 & 0 \\ a_{41}^{(j)} & a_{42}^{(j)} & a_{43}^{(j)} & 1 & 0 \\ a_{51}^{(j)} & a_{52}^{(j)} & a_{53}^{(j)} & a_{54}^{(j)} & 1 \end{array}\right] \left[\begin{array}{c} e_1^{(j)} \\ e_2^{(j)} \\ \hline e_3^{(j)} \\ e_4^{(j)} \\ e_5^{(j)} \end{array}\right] = \left[\begin{array}{cc|ccc} b_{11}^{(j)} & 0 & 0 & 0 & 0 \\ b_{21}^{(j)} & b_{22}^{(j)} & 0 & 0 & 0 \\ \hline 0 & 0 & b_{33}^{(j)} & 0 & 0 \\ 0 & 0 & 0 & b_{44}^{(j)} & 0 \\ 0 & 0 & 0 & 0 & b_{55}^{(j)} \end{array}\right] \left[\begin{array}{c} u_1^{(j)} \\ u_2^{(j)} \\ \hline u_3^{(j)} \\ u_4^{(j)} \\ u_5^{(j)} \end{array}\right]$$

where $e_i^{(j)}$ and $u_i^{(j)}$ denote the reduced form innovation and the structural shock of the ith equation in regime j, respectively. The elasticities $\alpha_{gy}^{(j)}$, $\alpha_{g\Delta p}^{(j)}$, $\alpha_{gi}^{(j)}$, $\alpha_{ty}^{(j)}$, $\alpha_{t\Delta p}^{(j)}$ and $\alpha_{ti}^{(j)}$ represent the automatic response of fiscal variables to economic activity and are computed using external information.

We follow Perotti (2004) and make the same assumptions about the values of $\alpha_{gy}^{(j)}$, $\alpha_{g\Delta p}^{(j)}$, $\alpha_{gi}^{(j)}$, and $\alpha_{ti}^{(j)}$, while we update his methodology to compute the values of $\alpha_{ty}^{(j)}$ and $\alpha_{t\Delta p}^{(j)}$ based on the work of Caldara (2011). In order to compute the latter values Perotti (2004) uses the OECD estimates for the elasticity of personal income tax liabilities with respect to earnings (Girouard and André, 2005). These estimates, however, are only available for every fifth year from 1979. We follow Caldara (2011) and use the NBER estimates from the TAXSIM model for this elasticity which are available at an annual frequency over our entire sample period (Feenberg and Coutts, 1993). Furthermore, we estimate the elasticity of transfers with respect to output instead of assuming the same value as Perotti (2004) (see Appendix B for the details).

The values of the elasticities are shown in Table 2. We obtain the values of α_{ty} and $\alpha_{t\Delta p}$ for the linear model by computing their sample mean for the entire sample. These values are lower than those used by Perotti (2004) due to the differences in our methodology. For the nonlinear model we compute the subsample means pertaining to each of the regimes in order to capture the potential differences in the workings of the automatic stabilizers.[17] While the differences in the subsample means of the elasticities seem to be small, recall that we also allow for regime specific covariance matrices for the residuals. Hence regime specific identification can be a potential source of asymmetric impulse responses across regimes and we examine whether this is indeed the case.

[Table 2 about here.]

We have chosen the Blanchard and Perotti (2002) approach among the alternative identification schemes primarily because it facilitates the comparability of our results with a large pool of literature on the effects of fiscal policy shocks. Moreover, the use of other identification approaches in the literature has limitations in a threshold VAR model. The narrative approaches of Ramey and Shapiro (1998) and Romer and Romer (2009) has too few observations to obtain reliable results after splitting the sample into two.[18] The other alternative identification method is the sign restrictions approach of Canova and Pappa (2006) and Mountford and Uhlig (2008). This identification method requires very few restrictions and can be easily motivated by the theoretical literature when applied to a linear VAR. However, it is not readily applicable to our regime switching setting. Several papers provide empirical support in favor of contractionary fiscal expansions (Perotti, 1999; Giavazzi and Pagano, 1990, 1995; Alesina and Perotti, 1996a,b, among others), i.e. macroeconomic variables can respond qualitatively differently to fiscal shocks conditional on the state of public finances. We therefore prefer to be agnostic and avoid putting any a priori restrictions on the sign of the impulse responses.

Finally, we have to point out that our chosen identification approach is not free of criticism. It was criticized, for example, for the sensitivity of its results to the elasticities used (Caldara and Kamps, 2008) or its inability to accommodate fiscal foresight (e.g.

[17]This is the same approach as in Baum and Koester (2011), for example. However, they restrict the estimated covariance matrices to be the same across regimes, which implies that their identification is regime specific only due to the regime specific elasticities.

[18]We have estimated the model using the narrative record of Romer and Romer (2009). The smaller regime in our nonlinear model contains only 6 episodes from their dataset which prevents reliable identification. Mertens and Ravn (2011) have reconciled the two approaches in the proxy SVAR framework for the three variable VAR of Blanchard and Perotti (2002), but the lack of reliable identification also applies to their approach in our model for the same reason.

Ramey, 2009; Leeper, Walker, and Yang, 2009). Our primary goal in this paper is to compare impulse responses obtained from our best fitting nonlinear model and from the benchmark linear VAR. We do not have reason to believe that these models are affected asymmetrically by the shortcomings of the identification method used and we can safely compare these two sets of results. Furthermore, recent papers by Chahrour, Schmitt-Grohé, and Uribe (2010), Perotti (2011) and Caldara and Kamps (2012) show that the method of Blanchard and Perotti (2002) does a reasonable job to identify the structural shocks of the model even in the presence of foresight.

3.2 Computing impulse response functions and their confidence intervals

As we pointed out already, the endogenous variables of the linear model depend on the lagged values of the debt-to-GDP ratio that are nonlinear functions of the lagged endogenous variables. It implies that (1a) has no moving average representation and in order to derive impulse responses we need to compute generalized impulse responses even for the linear model (see Appendix C for the details).

When computing impulse responses for the nonlinear model we can distinguish two cases based on whether we allow for the possibility of a regime switch following the structural shock. If the possibility of a regime switch is excluded, then the impulse response functions depend only on the regime when the shock hits. It is computationally much less demanding since each piecewise linear model has its own set of impulse responses that can be computed independently from each other. However, the differences in the responses of the model variables across regimes are overestimated, as Auerbach and Gorodnichenko (2010) point out, under the assumption that the regime at the time of the shock's arrival prevails forever. It is especially true at longer horizons since the importance of initial conditions should diminish over time.

Allowing for regime switches is in particular important for threshold VAR models since the coefficients of a TVAR model are functions of the observations in their respective subsamples only. Assuming that the same regime prevails forever implies that also the impulse responses depend on those subsamples only. If we allow for the possibility of a regime switch following the structural shock instead, then the impulse response functions depend on the parameter estimates of both regimes as well as on the dynamics of the threshold variable. These impulse responses thus utilize the entire sample and alleviate the disadvantage of TVARs arising from splitting the sample.

In order to simulate regime switches we need to model the dynamics of the threshold variable. This is the main reason we have chosen the specification of Favero and Giavazzi (2007) for our analysis. It links the dynamics of our threshold variable candidates to the endogenous variables of the model through the flow government budget constraint.

Many papers in the fiscal VAR literature report impulse responses that give the dollar response of each variable to a dollar shock to one of the fiscal variables. In a linear model this is achieved either by rescaling the impulse responses to a one standard deviation shock or by using a shock equivalent to 1 percent of GDP in order to obtain impulse responses that can be interpreted as a dollar value response without any further transformations. Since the impulse response functions of a linear model scale up proportionally with the size of the structural shock these two approaches yield the same result. In a nonlinear model, however, the size and the sign of the shock can matter depending on how different the estimated dynamics in the two regimes are. For this reason we use a structural shock equivalent to 1 percent of GDP during all our simulations.[19] This allows us to interpret the impulse responses of output as dollar value multipliers without any further transformation.[20]

We use a bootstrap approach to compute confidence intervals. We build time series $\mathbf{y}_t$ for the endogenous variables, d_t for the debt-to-GDP ratio and z_t for the threshold variable based on the estimated parameters of the model and resampled residuals. We repeat the first two steps of the estimation procedure for each generated series: we keep the estimated delay parameter $\hat{d}$ fixed, but reestimate the coefficients $\hat{\mathbf{\Phi}}^{(j)}(r,d)$ and $\hat{\mathbf{\Sigma}}^{(j)}(r,d)$ and the threshold value $\hat{r}(d)$ each time. Since parameter uncertainty about the threshold value can be a major weakness to regime switching models this approach attempts to address this issue. We simulate impulse responses for each replication and use the empirical percentiles to obtain our confidence intervals.[21]

4 The best fitting nonlinear model

This section reviews the results relevant to the choice of the benchmark model specification. We compare the estimation results of the competing nonlinear specifications first. Then, we confront the best fitting nonlinear model to the linear one by means of the two nonlinearity tests. Finally, we inspect the fit of our selected nonlinear specification.

As discussed in section 2, we estimate a nonlinear model with two regimes for every candidate threshold variable. For each estimated nonlinear model we calculate the AIC, SC and HQC information criteria to choose among the competing model specifications in terms of lag structure and alternative threshold variables. We report the values of the three different information criteria for the estimated models in Table 3. Given the large number of model specifications it is reassuring that the data prefer the model with

[19]In the terminology of Erceg and Lindé (2010) we compoute the average multiplier $\frac{g}{y}\frac{\Delta y}{\Delta g}$ with $\Delta g = 0.01y$ as opposed to the marginal multipler $\frac{g}{y}\frac{dy}{dg}$ delivered by linear models.

[20]We investigate the extent to which the fiscal multiplier changes with the size and the sign of the structural shock in Section 5.3.

[21]For a detailed description and discussion see Artis, Galvão, and Marcellino (2007) and Galvão and Marcellino (2010). We use 2000 replications in our plots throughout the paper.

the debt-to-GDP ratio as a threshold variable independently of the information criterion used.[22]

[Table 3 about here.]

The different information criteria suggest quite different number of lags, ranging from $(p_1 = 1, p_2 = 1)$ by SC to $(p_1 = 3, p_2 = 4)$ by AIC. We select the most parsimonious model specification chosen by the SC criterion for the same reason we fixed the lag length for the exogenous regressor at $k = 2$: in order to minimize the probability of overfitting due to the high cost of estimating additional parameters both in terms of computer and estimation power. This leaves us enough degrees of freedom even in the regime with the smaller number of observations for the estimation the coefficients of the model.

We plot the log likelihood function against the threshold variable to assess the fit of the model and, in particular, how strong the threshold effect is in the estimated model (Figure 1).[23] The log likelihood displays a sharp spike at the estimated threshold value indicating a strong threshold effect. The dotted horizontal line marks the location of the 1 percent confidence interval around the estimated threshold value based on the LR-statistics approach of Hansen (2000). The small confidence interval implies a tightly estimated threshold value.

[Figure 1 about here.]

The test statistics for the two nonlinearity tests, the arranged regression test of Tsay (1998) and the variable addition test of Teräsvirta (1998), are shown in Table 4. We already have estimated the delay parameter to be $d = 2$, but we perform the tests for delay values up to 4 in order to make sure that the result is independent of the estimated value. Both tests for all delay values reject the linear model in favor of the benchmark nonlinear one even at 1 percent significance.

[Table 4 about here.]

Based on these results we feel confident to use the model with the debt-to-GDP ratio as a threshold variable as our benchmark specification. Figure 2 plots the evolution of the threshold variable and the estimated threshold value, that splits the sample into two parts. The estimated threshold value implies that approximately three quarters (75.5 percent) of the observations belong to the lower regime (good times), as they correspond

[22] In fact the data prefer the model with the debt-to-GDP ratio as the threshold variable for almost all possible combinations of the lag lengths. Thus, this result is not driven by the specific lag structure chosen or the particular way one or the other information criterion penalizes the likelihood of the models. It implies that the choice of the threshold variable depends on the log likelihood only and is independent of the choice of the lag structure which is selected by the information criterion.

[23] We plot the log likelihood function concentrated with respect to the regime specific coefficient and co-variance matrices ($\hat{\mathbf{c}}^{(j)}(r, d)$, $\hat{\mathbf{\Gamma}}^{(j)}(r, d)$, $\hat{\mathbf{\Phi}}^{(j)}(r, d)$ and $\hat{\mathbf{\Sigma}}^{(j)}(r, d)$) that is maximized to obtain the estimate for the threshold value (r) given lag lengths (p_j and k) and the delay parameter (d).

to the debt-to-GDP ratio (lagged two periods) lower than 42.54 percent. The rest of the observations belong to the high debt-to-GDP ratio regime (bad times) denoted with the shaded area in the plot.

[Figure 2 about here.]

Table 5 reports the estimated feedback from the debt-to-GDP ratio to the endogenous variables. Our results are very similar to Favero and Giavazzi (2007) in the sense that the coefficients on the first and second lags are of opposite sign but the same magnitude. This is not only true for the linear model, but also for both regimes of the nonlinear model. We have tested the restriction that the sum of the two lag coefficients in each equation is zero, i.e. the first difference of the debt-to-GDP enters the specification with one lag only, but rejected the restriction for both models.[24]

[Table 5 about here.]

Interestingly, the models using a threshold variable related to the sovereign risk premium provide the poorest fit (Table 3). Furthermore, our nonlinearity tests do not reject the linear model for any of these variables as a threshold variable. This confirms our prior belief that investors considered sovereign default risk negligible in the United States over our sample period and implies that the estimated threshold value should not be interpreted as a fiscal limit above which public finances are not sustainable. This evidence is also in line with the results of Ghosh, Kim, Mendoza, Ostry, and Qureshi (2011) that the current debt level in the US is well below its “debt limit” beyond which fiscal solvency is in doubt; or to use their terminology, the US has considerable fiscal space, defined as the distance between the current debt level and its debt limit.[25]

Before we turn our attention to the estimated fiscal multipliers we deem important to stress that our results do not imply that the debt-to-GDP ratio is the only possible source of nonlinearity among the threshold variables considered here. The information criteria do not provide us a measure of nonlinearity only rank the competing specifications in terms of likelihood. Hence, our selected threshold variable is only the most likely one given our specification and an authentic one given the results of our nonlinearity tests. Other specifications or even questions within the same specification may well require the use of the other threshold variables considered here. In fact, the nonlinearity tests reject the linear specification also in favor of some of the other nonlinear models we have disregarded here.

[24]The test statistics for the linear and the nonlinear model are 25.5 and 160.4, respectively, following a chi-squared distribution with 5 and 10 degrees of freedom. The restriction is rejected in both cases even at 1 percent significance level.

[25]We provide further evidence against nonlinearities arising from sovereign risk premium in Section 5.2.

5 Can fiscal policy always stimulate output?

This section presents the results of our impulse response analysis. We compare the fiscal multipliers between the linear model and the two regimes of the nonlinear model first. Then, we examine if our results are consistent with spending reversals that become stronger as the debt-to-GDP ratio increases. Additionally, we implement a policy counterfactual to understand the source of the asymmetric response to fiscal policy shocks across regimes. We conclude this section by reviewing some implications of our model.

5.1 Estimated fiscal multipliers

The main focus of this section is to characterize the size of the state-dependent fiscal multipliers based on the impulse responses from our models. For this reason, and to conserve space, we present the responses of three variables only: i) GDP, the main variable of interest, ii) the debt-to-GDP ratio, the threshold variable in our model, and iii) the deficit-to-GDP ratio, the main driving force of our threshold variable. Recall that we use expansionary fiscal shocks equivalent to 1 percent of GDP in all our simulations; a positive expenditure and a negative revenue shock. This allows us to interpret the impulse responses of GDP as dollar value multipliers without any further transformation.

Government spending shock. We plot the responses of the three variables to a positive expenditure shock in Figure 3. The three columns plot the same impulse responses, but differ in the confidence intervals. The shaded areas represent a one standard deviation confidence band around the impulse responses of the linear model, good times and bad times in the first, second and third columns respectively.

[Figure 3 about here.]

At almost every horizon the output response is stronger in good times than in bad with the response of the linear model in between. The peak value of the expenditure multiplier is 1.45 in the linear model. This result is in line with the estimated values for the US in the literature (see Ramey, 2011, and the references therein). Compared to other papers in the SVAR literature, it is slightly above the estimated value by Blanchard and Perotti (2002) and Favero and Giavazzi (2007), close to the findings of Perotti (2004) and below the results of Caldara and Kamps (2008) and Caldara and Kamps (2012). The output response peaks at 1.58 after 5 quarters in good times, slightly above the linear model. The output response in bad times is stronger on impact than in good times, but steadily decreases after that and becomes weaker after 4 quarters. Its peak value is only 0.69 and it is only significantly different from zero for two periods after impact.

The response of the deficit-to-GDP ratio is the least persistent during bad times out of the three cases. It is already the weakest on impact due to regime-specific identification and it returns to baseline faster than in good times or in the linear model. This finding is consistent with the empirical evidence of Bohn (1998) and the assumption of Corsetti, Meier, and Müller (2009) that spending reversals become stronger as the debt-to-GDP ratio increases. Moreover, the finding that output response to an expenditure shock in bad times is higher at shorter while it is lower at longer horizons compared to good times is also consistent with the prediction of the New Keynesian model with spending reversals.

The response of the debt-to-GDP ratio is positive in all three cases reflecting the expansionary nature of the expenditure shock. Interestingly, fiscal expansion is the least costly in terms of debt accumulation in bad times with a peak response of only 0.6 percent of GDP 10 quarters after the shock. This is due to the dynamics of both output and deficit. The stronger impact response of output implies a weaker impact response of the debt-to-GDP ratio, while the less persistent deficit response outweighs the weaker response of output afterwards and leads to a weaker debt accumulation.

Tax receipt shock. We plot the responses to a negative revenue shock in Figure 4. The three columns again share the same impulse responses, but differ in the depicted confidence bands.

[Figure 4 about here.]

We find that also the revenue multiplier is decreasing in the the debt-to-GDP ratio. The main differences with respet to the effects of an expenditure shock are that the output response is stronger in good times than in bad already on impact, and the revenue multiplier in bad times becomes slightly negative between quarters 4 and 17. However, it is never significantly different from zero. The peak value of the revenue multiplier is 1.55 in the linear model.[26] Blanchard and Perotti (2002) estimate the peak value to be only 0.74, Favero and Giavazzi (2007) find 0.9, while Caldara and Kamps (2012) report a value of 1.6. The output response in good times is very similar to that of the linear model; it is stronger up to 14 quarters after impact and becomes weaker afterwards. The revenue multiplier is positive on impact in bad times, but turns negative after 3 quarters only to return to positive 15 quarters later. Its peak value is only 0.32, however it is never significantly different from zero and it is outside of the confidence intervals of both the linear model and the other regime for almost the entire horizon.

As opposed to the case of an expenditure shock, the response of the deficit-to-GDP ratio is the most persistent out of the three cases during bad times. It is increasing up to 1.29 percent of GDP above baseline before it starts reverting to it and it is stronger than in good times for almost every horizon. This finding seems to indicate that spending reversals become weaker as the debt-to-GDP ratio increases. Given that the deficit responses to the two fiscal shocks have quite different implications for the relationship between the

[26] We define the revenue multiplier as the response of output divided by the size of the shock (expressed in terms of output) multiplied by minus one, $-\frac{t}{y}\frac{\Delta y}{\Delta t}$, in order to express the expansionary effect of revenue based fiscal stimulus and to facilitate comparison with expenditure multipliers.

strength of the spending reversals and the degree of indebtedness, we will return to this question in more detail later.

The response of the debt-to-GP ratio is positive again in all three cases since we are considering an expansionary revenue shock. The response of debt is much more pronounced during bad times with a peak response of 2.36 percent of GDP 15 quarters after impact. This is driven by both the stronger deficit and weaker output responses in that regime.

Measuring the difference An important question is whether any two of the three output responses (linear model, good times and bad) are significantly different from each other. The output response to a revenue shock in the linear model and in good times never seem to be significantly different since they both are inside each other's confidence intervals. The answer is not so obvious when one impulse response lies outside of the confidence intervals of the other, but the opposite is true the other way around. However, we cannot claim that they are different even if both impulse responses lie outside of the confidence intervals of one another if the distributions forming the confidence intervals are correlated; as is the case for the two regimes of the nonlinear model.

To take into account the correlation between the distributions we compute the pairwise differences between the impulse responses within each replication of our bootstrap approach.[27] Figure 5 plots the differences in the fiscal multipliers derived from the two models and the confidence intervals around these differences. The graphs reflect our previous discussion that both the revenue and the expenditure multipliers, with the exception of the expenditure multiplier on impact, are higher in good times than in bad with the multipliers of the linear model in between. The expenditure multiplier never seems to be significantly different between the two models or across regimes, while the revenue multiplier in bad times is significantly different from the other two cases for most of the horizons.

[Figure 5 about here.]

To sum up, we find strong asymmetries in the response of output across regimes which imply a negative relationship between the fiscal multiplier (both the expenditure and the revenue multipliers) and the debt-to-GDP ratio. This result implies that the use of deficit financed fiscal stimulus is characterized by diminishing returns since it increases the debt-to-GDP ratio and, consequently, leads to a decrease in the fiscal multiplier.

[27] The distributions forming the confidence intervals of the linear model and any of the two regimes of the nonlinear model are uncorrelated by construction. Taking pairwise differences this way constitutes a random matching between these two populations.

5.2 Inspecting the asymmetric response to fiscal policy shocks

In this section we examine the observed differences between the fiscal multipliers in the two regimes. First, we return here to the question whether spending reversals become stronger or weaker as the debt-to-GDP ratio increases. Then, we look at the structural shocks to see if regime specific identification can contribute to these differences. Then, we carry out a counterfactual analysis to determine if we can attribute the differences to the dynamic paths of one or more variables.

Spending reversals We return here to the question whether our results imply a positive or a negative relationship between the strength of the spending reversals and the degree of indebtedness. Recall that our results from the impulse response analysis show that the response of deficit to an expenditure shock is less persistent during bad times suggesting a positive relationship (Figure 3). On the other hand, following a revenue shock the response of deficit shows a more pronounced fiscal expansion in bad times indicating that spending reversals become weaker in that regime (Figure 4).

Bohn (1998) uses annual data to estimate a fiscal reaction function for the US in a single equation framework. He finds that the primary deficit-to-GDP ratio responds negatively to changes in the debt-to-GDP ratio, which implies that the debt-to-GDP ratio should be mean-reverting. Furthermore, his results show that the marginal response is increasing in the debt-to-GDP ratio, which implies that spending reversals are stronger during bad times.

However, we cannot interpret the impulse response of deficit as the response to changes in debt for two reasons. First, we could interpret our identified fiscal shocks as an innovation to debt if the implied fiscal expansion was purely deficit financed. But government expenditures, for example, respond endogenously to a revenue shock in our model and it is unlikely that the response is such that the impulse response corresponds to the case of a deficit financed revenue shock. Second, Bohn (1998) finds a positive correlation between current deficit and lagged debt while an expenditure shock in our model affects debt starting only contemporaneously.

Debt enters as an exogenous regressor into our model as well as a threshold variable. This offers us two ways to assess how debt accumulation affects the evolution of deficit. The left panel of Figure 6 shows how the dynamics of deficit changes with the debt-to-GDP ratio as a threshold variable. It plots the mean evolution of deficit along the simulated baseline scenarios in the two regimes. The graph corroborates the first finding of Bohn (1998) since the deficit of good times is replaced by a surplus in bad times in order to stabilize the debt-to-GDP ratio.

The right panel of Figure 6 shows how deficit changes with debt within the regimes as well as how its dynamics shifts between regimes. It plots the "impulse response" of deficit to a "debt shock" in both regimes.[28] The negative response at each horizon during bad times implies that debt stabilization starts already on impact. During good times, however, the initially positive response of deficit to an increase in debt shows an accelerating debt accumulation which turns into negative to stabilize the debt stock only after several periods. The marked shift in the impulse responses show that the marginal response of deficit to changes in debt is decreasing in the debt, which corroborates Bohn's second finding.

[Figure 6 about here.]

Regime specific identification A potential source for the differences in our impulse responses is the impulses themselves. We could observe different impulse responses across regimes even with the same lag coefficients if the variables responded to different shocks due to regime specific identification. Table 6 compares the reduced form equivalents of the different structural shocks used in the simulations across regimes. The revenue shocks are practically identical in both regimes while we can see a difference between the reduced form equivalents of the expenditure shocks. In particular, the impacts on tax revenues and output are higher in bad times.

[Table 6 about here.]

These differences on impact can prevail if the estimated processes are persistent. To determine the extent to which these differences are important we carried out a counterfactual analysis where we used the same reduced form shocks in both regimes to derive impulse responses. The first column in Figure 7 repeats the third column of Figure 5 for convenience and shows the difference between the fiscal multipliers across regimes in the benchmark model. The second and third columns show the results from the experiments using the same reduced form shock in both regimes; the identified good times shock in the second and the bad times shock in the third column, respectively. While regime specific identification contributes to the differences in the multipliers across regimes, qualitatively the results are very similar to the benchmark case.

[Figure 7 about here.]

[28]Since we do not model innovatins to debt explicitely, we compute the response of deficit to a debt shock as the derivative of the deficit-to-GDP ratio with respect to the debt-to-GDP ratio. Deficit is a nonlinear function of the endogenous variables, hence its response at each horizon depends not only on the coefficient estimates, but also on the value of the variables when the shock arrives. The impact response, for example, is given by

$$\frac{\partial deficit_t}{\partial d_{t-1}} = \frac{\partial \frac{\exp(g_t)-\exp(t_t)}{\exp(y_t)}}{\partial d_{t-1}} = (\gamma_{g,d_{t-1}} - \gamma_{y,d_{t-1}})\frac{\exp(g_t)}{\exp(y_t)} - (\gamma_{t,d_{t-1}} - \gamma_{y,d_{t-1}})\frac{\exp(t_t)}{\exp(y_t)}$$

where $\gamma_{v,d_{t-1}}$ is the estimated coefficient of d_{t-1} in the equation of variable v. We have plotted the mean value of the responses within each regime.

Counterfactual analysis The previous results imply that the asymmetric response of output stems from the differences in the estimated dynamics (coefficients) of the two regimes. This section presents the results of two counterfactual experiments to analyze the contribution of the fiscal rules and the interest rate rule to these observed asymmetries.

In the first experiment we examine the contribution of spending reversals. As we discussed earlier we have found evidence of spending reversals in our sample that are becoming stronger as the debt-to-GDP ratio increases. These reversals have a direct and an indirect effect on the multipliers. First, the differences in the fiscal rules can translate directly into asymmetries in the fiscal multipliers. A less persistent fiscal expansion implies a smaller multiplier during bad times even if output responded the same way to the shock. Second, as Corsetti, Meier, and Müller (2009) point out, an expansionary fiscal shock with spending reversals creates expectations of government spending below trend in the future which alters the response of the private sector.

In order to determine how much of the asymmetric response of output we can attribute to the direct effect of the differences in the fiscal rules, we implement the following counterfactual experiment. We compare the output responses in the two regimes in a model where we control for the differences in the fiscal rules. We impose the coefficients of the first two equations of one of the regimes to the other regime as well and allow for the estimated regime specific coefficients only in the other three equations; namely the GDP, inflation and interest rate equations. Comparing the three columns in Figure 8 shows that asymmetries in the fiscal multipliers prevail even if we impose the same fiscal rules in both regimes. Hence, the asymmetric output response is not driven by differences in policy, which leaves the explanation that it is the response of the private sector that matters.

[Figure 8 about here.]

The second experiment provides further evidence that the estimated threshold value for the debt-to-GDP ratio should not be interpreted as a fiscal limit above which public finances are not sustainable. We compare the output responses in the two regimes in a model where we allow for regime specific coefficients only in the interest rate equation while we impose the same coefficients in all the other equations. The second and third columns in Figure 8 show that the interest rate equation alone is not able to generate any observable asymmetries in the fiscal multipliers. Hence, the asymmetric output response does not arise from nonlinearities related to the behavior of the sovereign risk premium.

[Figure 9 about here.]

5.3 Model implications

This section reviews some direct or indirect implications of our model. First, we examine an issue that is specific to the impulse response analysis of nonlinear models. We vary the size and the sign of the structural shocks to see if it leads to any noticeable differences in the fiscal multipliers presented before. Second, we implement a policy counterfactual to analyze the different policy implications of the two models for The Recovery Act. Finally, we assess the fit of our model in terms of how well it replicates the regime switching probabilities estimated directly from the dynamics of the threshold variable.

Size of the structural shock The impulse response functions of a linear model scale up proportionally with the size of the structural shock. Hence, one can obtain output response functions that represent dollar value multipliers in two equivalent ways. The first possibility is to rescale the output response function to a one standard deviation shock while the second option is to use a shock equivalent to 1 percent of GDP in order to obtain an output response function that can be interpreted as a dollar value multiplier without any further transformations. In a nonlinear model, however, the average and the marginal multipliers are not equal and these two alternatives lead to different results.[29]

We have varied both the expenditure and revenue shocks from a fiscal expansion equivalent to 5 percent of GDP to a contraction with the same magnitude (Table 7). The general conclusion from the exercise is that the more contractionary a shock is the larger its effect on output is. This lends further support to our earlier argument that there are diminishing returns to the use of deficit financed fiscal stimulus. Intuitively, since our impulse response analysis has shown that the fiscal multiplier is a decreasing function of the debt-to-GDP ratio, the more expansionary the fiscal shock is the more debt is accumulated and the smaller the multiplier gets.

The results also show that expenditure cuts are associated with a higher output loss than revenue increases at shorter horizons during bad times. At longer horizons our results are consistent with Alesina, Favero, and Giavazzi (2012) who find that revenue based fiscal consolidations have been associated with mild and short-lived recessions while tax-based adjustments have been associated with prolonged and deep recessions.

[Table 7 about here.]

The effects of The Recovery Act So far we have focused on comparing the output effects of a one-time expenditure or revenue shock equivalent to 1 percent of GDP. In this policy counterfactual we try to evaluate the different policy implications of the two models for the effects of The Recovery Act.

The Recovery Act represents a series of both expenditure and revenue shocks over 7 quarters starting in 2009:2. These shocks are all smaller than 1 percent of GDP but accumulate to 5.8 percent of GDP (Table 8).[30] While we excluded the financial crisis

[29] The average multiplier is defined as $\frac{g}{y}\frac{\Delta y}{\Delta g}$ and the marginal multiplier is given by $\frac{g}{y}\frac{dy}{dg}$ (Erceg and Lindé, 2010).

[30] This is admittedly a very crude way to simulate the effects of The Recovery Act. However, we are not interested in the quantitative effects but in the different implications of the two models.

period from our sample, the first quarter of the stimulus is in bad times according to our estimated threshold value.[31] Consequently, we simulate impulse responses from the nonlinear model starting in that regime.

[Table 8 about here.]

Figure 10 plots the impulse responses from both the linear and the nonlinear models. According to the linear model output is significantly above the baseline already on impact and the difference is steadily increasing over the entire horizon of our simulation. By the end of the fifth year output is 7.2 percent above the baseline. Deficit is increasing until the end of the stimulus and reaches a peak deficit of 3.7 percent of GDP above baseline. Consequently debt is accumulated quickly and it peaks by 6.0 percent of GDP above the baseline 16 quarters after the first shock.

The nonlinear model gives a different view on the effects of The Recovery Act. It seems to be much less effective in stimulating output since the response remains around 1 percent above the baseline for 15 quarters after the first shock. Then it slowly rises and by the end of the fifth year GDP is 1.9 percent higher than it would have been without the fiscal expansion. The response of deficit and debt are very similar to that of the linear model peaking by 3.6 and 6.5 percent of GDP above their respective baselines. Consequently, according to the nonlinear model the stimulus is far less successful at stimulating the economy while the consequences on accumulated debt are the same.

[Figure 10 about here.]

Regime switching probabilities Recall that the dynamics of the threshold variable is determined by the flow government budget constraint (1b) which is an identity and has no estimated parameters. The evolution of the debt-to-GDP ratio in our simulations hence is driven indirectly by shocks to the endogenous variables. This gives us a way to assess how well our benchmark model approximates the nonlinear dynamics in the data. We can compare the dynamics of the threshold variable in our simulations to the dynamics obtained from a model where innovations to the debt-to-GDP ratio are explicitly modeled.

The key contribution of the threshold variable to the dynamics of our nonlinear model is to determine the regime. To assess how well the benchmark model can replicate this aspect of the data we plotted regime switching probabilities derived from a univariate model of the debt-to-GDP ratio and from our benchmark nonlinear model in Figure 11.[32] The left panel plots the regime switching probabilities at each horizon for simulations starting in good times: the probability at a given horizon is approximated by the fraction

[31] We have also reestimated the model using a sample period extended until 2009:4 to include the first period of the stimulus. The inference regarding the regime in 2009:2 remains the same.

[32] The univariate model is an autoregressive model of the debt-to-GDP ratio in levels with four lags. We have experimented also with a model in first differences, but the results remained qualitatively similar.

of the simulations being in bad times, i.e in which the debt-to-GDP ratio is above the estimated threshold value. The right panel computes the probabilities for bad times in a similar way.

The dynamics of the threshold variable are reproduced by the benchmark nonlinear model remarkably well despite being a piecewise linear approximation only. In bad times the probabilities are always very close and never significantly different from each other. In good times the regime switching probabilities are slightly higher in the benchmark model than in the data, but again are not significantly different for most of the horizon considered. We can observe two asymmetries between the panels which can both be explained by the differences of the two subsamples constituting the two regimes.

First, the wider confidence intervals in bad times are due to the smaller subsample which translates into larger parameter uncertainty. Second, the higher probability of a regime switch in bad times can be explained by the asymmetric distribution of the threshold variable around the estimated threshold value. The distance of the highest debt-to-GDP ratio in bad times (50.60 percent) from the threshold value is 2.5 times the distance of the lowest observation in good times (22.74 percent) from the threshold, while the same ratio for the median observations in the two regimes is 2.2. Hence, the threshold variable observations are more dispersed below the threshold value than above leading to an asymmetry in the probability of regime switches.

[Figure 11 about here.]

6 Robustness

In this section we consider alternative specifications to assess whether our results are due to the particular empirical strategy chosen.

Debt-to-GDP ratio as an exogenous regressor One might argue that including the debt-to-GDP ratio as an exogenous regressor could bias our results in favor of the benchmark model chosen. Given that our model specification is only an approximation, using the same variable both as a regressor and a threshold variable can help the nonlinear model to capture the underlying nonlinearities in the data better. To this end we have reestimated all the models using the same variable both as a regressor and a threshold variable, but no clear pattern emerged how the information criteria have changed compared to our earlier results. Some of the competing models have their fit improved but the benchmark model still provides the best fit (results are not reported to conserve space).

Identification There is a considerable disagreement about the value of the elasticities in the fiscal SVAR literaure, in particular about the output elasticity of revenues. Caldara (2011), for instance, uses a lower elasticity than our estimate in this paper.[33] Mertens and

[33]He estimates one of the components of the elasticity, the elasticity of private consumption with

Ravn (2011), on the other hand, show that the narrative record can be reconciled with an elasticity around 3 within their proxy SVAR framework. Since Caldara and Kamps (2008) have demonstrated that the fiscal multipliers computed from a SVAR using the identification approach of Blanchard and Perotti (2002) can be sensitive to the values of the elasticities used, we consider alternative values to both the output and price elasticities of revenues, α_{ty} and α_{tp}, to see how robust our results are.

We find that expenditure multiplers are robust to alternative values of these elasticites since the expenditure shock is ordered first. Revenue multipliers, on the other hand, are in general increasing in these elasticities which is in line with the findings of Caldara (2011). The output response in bad times increases more than in good times and the differences in the multiplers across regimes becomes smaller and less significant.

Structural break Since the observations with high debt-to-GDP ratio are concentrated in the second part of our sample, it could very well be the case that the differences in the two regimes captured with our benchmark specification are caused by a structural break. Several papers studying the effects of fiscal policy shocks split their sample around 1980 to allow for the possibility of a structural break (see for example Perotti, 2004; Favero and Giavazzi, 2007; Caldara and Kamps, 2008, among others) and find that impulse responses in the second subsample are less pronounced than in the first subsample. They attribute their findings to changes in the conduct of fiscal and/or monetary policies or the developments in the variance of shocks known in the literature as the Great Moderation.

To confront our nonlinear model with the possibility of a structural break we estimate two alternative models. First, we estimate a structural break model (SB-VAR) by splitting our sample into two subsamples (1960:1-1979:4 and 1980:1-2007:4) at roughly the same date as other papers in the literature. Second, we estimate an endogenous structural break model (ESB-VAR) where we also estimate the most likely date of the structural break in the model based on a penalized likelihood approach. In particular, we estimate a series of SB-VARs with a possible structural break as early as 1978:1 and as late as 1987:4. The estimated endogenous structural break that emerges from this exercise is between 1979:4 and 1981:1 depending on the information criterion used.

We list the values of information criteria for both models in Table 9, where we also repeat the results of our benchmark specification for convenience. While we find that both alternative models provide a comparable fit to most of the alternative nonlinear specifications considered in this paper, none of the two structural break models fit the data better than the benchmark specification.

[Table 9 about here.]

Three regimes While our nonlinearity tests rejected the linear model in favor of the nonlinear specification we have no a priori reason to exclude the possibility of more than

respect to output, to be around 0.6 instead of assuming a value of 1 as Perotti (2004) does.

two regimes. In fact, the nonlinearity tests we have used can only tell us whether the data is generated by a linear or nonlinear process, but they do not specify the number of regimes to be used for the nonlinear model. If nonlinearities are strong in our sample, then more regimes should improve the fit of our models given that threshold models are only piecewise linear approximations of the underlying nonlinear process.

We estimate a nonlinear model with three regimes for all the threshold variable candidates we have considered earlier.[34] The debt-to-GDP ratio emerges as the preferred threshold variable also in the case of three regimes irrespective of which information criteria we use. To keep comparability with the two regime case we focus again on the most parsimonious version of the model selected by the SC criterion.

We need to estimate two threshold values, $r = (r_1, r_2)$, that split the observations into three regimes. Hence, the model is estimated over a two dimensional grid and the log likelihood we have used to assess the fit of the model with two regimes in Section 4 becomes a three dimensional function.[35] Figure 12 shows the marginal log likelihood functions obtained as cross-sections of this three dimensional log likelihood surface. In panel (a) we fix the lower threshold value at its point estimate and plot the log likelihood as a function of the higher threshold value only. The graph is essentially identical to Figure 1 which depicted the log likelihood of the model with two regimes only and the estimated upper threshold value is equal to the estimated threshold value of the two regime model.[36] The marginal log likelihood displays a sharp spike at the estimated threshold value indicating again a strong threshold effect.

In panel (b) we plot the log likelihood as a function of the lower threshold value only while fixing the higher threshold. The estimated lower threshold is 27.86 percent, but the loglikelihood is practically flat. This indicates a large uncertainty around the estimate of the lower threshold value, which is a crucial parameter for a regime-switching model. Based on the model's inability to distinguish the two lower regimes from each other, we have decided to use the model with two regimes for our analysis.

[34]The estimation procedure is a straightforward generalization of the case with two regimes; it only takes more time given that the dimensionality of the estimation procedure is exponentially increasing in the number of regimes. Similarly to the two regimes case we fix the lag length of the exogenous regressor at $k = 2$ and consider values for both the lag lengths of the endogenous variables and the delay parameter up to 4. We need to estimate two threshold values, $r = (r_1, r_2)$, that split the observations into three regimes. We estimate the model over a two dimensional grid $r \in R \times R$ of all possible combinations that ensure that each regime contains at least 20 percent of the observations and calculate the log likelihood for each of them. Recall that we do not constrain the model to have the same lag lengths in the two regimes and thus the "one-step-at-a-time" approach described by Hansen (1999) cannot be used here to reduce the computational burden.

[35]It is the log likelihood function $\mathscr{L}(r_1, r_2)$ concentrated with respect to the regime specific coefficient and covariance matrices ($\hat{\mathbf{c}}^{(j)}(r_1, r_2, d)$, $\hat{\mathbf{\Gamma}}^{(j)}(r_1, r_2, d)$, $\hat{\mathbf{\Phi}}^{(j)}(r_1, r_2, d)$ and $\hat{\mathbf{\Sigma}}^{(j)}(r_1, r_2, d)$) that is maximized to obtain the estimate for the threshold values (r_1 and r_2) given lag lengths (p_j and k) and the delay parameter (d).

[36]Hansen (1999) shows that the estimated threshold value in a threshold model with two regimes is a consistent estimate for one of the two threshold values if the data generating process is a threshold model with three regimes.

[Figure 12 about here.]

7 Conclusions

Motivated by the recent deterioration of the state of public finances and a rekindled interest in the ability of fiscal policy to stimulate aggregate demand, we use a regime-switching empirical model to estimate state-dependent fiscal multipliers that depend on the state of public finances in the United States. First, we estimate several model specifications in order to see which conditioning variable we should use to proxy the state of public finances for the purpose of defining the regimes in our model. We use conditioning variables related to the cost of servicing the debt, the stationarity and the stock of accumulated debt, to the primary deficit, and to the sovereign risk premia. Comparing these model specifications in terms of their fit we find that the model with the debt-to-GDP ratio as a conditining variable fits the data best. The estimated threshold value of debt that triggers the regime switch is at 42.5 percent of GDP and it splits the sample such that approximately one quarter of the observations are in the regime characterized by high debt-to-GDP ratio.

Second, using the impulse response functions of our regime-switching model we compute fiscal multipliers conditional on the debt-to-GDP ratio. The strong asymmetries that we find in the response of output across regimes reveal a negative relationship between the fiscal multiplier and indebtedness. This finding implies that the use of deficit financed fiscal stimulus is characterized by diminishing returns since an expansionary fiscal policy shock increases the debt-to-GDP ratio and, consequently, leads to a smaller fiscal multiplier. While we estimate positive fiscal multipliers in both regimes, our results show that the ability of fiscal policy to stimulate aggregate demand has been weaker in the aftermath of the financial and economic crisis than other papers in the literature have found. Moreover, our results lend support to a policy that reduces the debt stock during booms to ensure that the fiscal multiplier is higher during recessions when fiscal stimulus is most needed.

While we find that the fiscal multiplier varies with the degree of indebtedness, related empirical papers provide evidence that fiscal multipliers depend also on other factors, including business cycle conditions, exchange rate regime and the degree of openness to trade. Future research should concentrate on creating a common econometric framework that encompasses these different factors and allows to investigate their interactions and their relative contributions to the state-dependent effects of fiscal policy. Furthermore, it would be instructive to construct DSGE models that could account for the state-dependency of fiscal multipliers and would provide a platform for policy experiments.

References

Afonso, A., J. Baxa, and M. Slavík (2011): "Fiscal Developments and Financial Stress: A Threshold VAR Analysis," Working Paper 1319, ECB.

Alesina, A., C. Favero, and F. Giavazzi (2012): "The Output Effect of Fiscal Consolidations," Working Paper 18336, NBER.

Alesina, A., and R. Perotti (1996a): "Fiscal Expansions and Fiscal Adjustments in OECD Countries," Working Paper 5214, NBER.

——— (1996b): "FiscalAdjustments in OECD Countries: Composition and Macroeconomic Effects," Working Paper 5730, NBER.

Altissimo, F., and V. Corradi (2002): "Bounds for Inference with Nuisance Parameters Present Only Under the Alternative," *Econometrics Journal*, 5(2), 494–519.

Andrews, D. W. K., and W. Ploberger (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62(6), 1383–1414.

Artis, M., A. B. Galvão, and M. Marcellino (2007): "The Transmission Mechanism in a Changing World," *Journal of Applied Econometrics*, 22(1), 39–61.

Auerbach, A., and Y. Gorodnichenko (2010): "Measuring the Output Response to Fiscal Policy," mimeo.

——— (2011): "Fiscal Multipliers in Recession and Expansion," mimeo.

Batini, N., G. Callegari, and G. Melina (2012): "Successful Austerity in the United States, Europe and Japan," Working Paper WP/12/190, IMF.

Baum, A., and G. B. Koester (2011): "The impact of fiscal policy on economic activity over the business cycle — evidence from a threshold VAR analysis," Discussion Paper: Economic Studies 03/2011, Deutsche Bundesbank.

Bi, H. (2010): "Sovereign Default Risk Premia, Fiscal Limits, and Fiscal Policy," mimeo.

Bilbiie, F., T. Monacelli, and R. Perotti (2012): "Tax Cuts vs. Spending Increases: Welfare and the Zero Lower Bound," .

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), 1329–1368.

Blanchard, O. J. (1990): "Suggestions for a New Set of Fiscal Indicators," OECD Economics Department Working Papers 79, OECD.

Bohn, H. (1998): "The Behavior Of U.S. Public Debt And Deficits," *The Quarterly Journal of Economics*, 113(3), 949–963.

Burriel, P., F. de Castro, D. Garrote, J. Gordo, Esther Paredes, and J. J. Prez (2009): "Fiscal Policy Shocks in the Euro Area and the US," Working Paper 1133, ECB.

Caldara, D. (2011): "The Analytics of SVARs: A Unified Framework to Measure Fiscal Multipliers," mimeo.

Caldara, D., and C. Kamps (2008): "What are the effects of fiscal policy shocks? A VAR-based comparative analysis," Working Paper 877, ECB.

——— (2012): "The Analytics of SVARs: A Unified Framework to Measure Fiscal Multipliers," Finance and Economics Discussion Series 2012-20, Borad of Governors of the Federal Reserve System.

Callen, T., M. Terrones, X. Debrun, J. Daniel, and C. Allard (2003): "Public debt in Emerging markets: Is it too high?," in *World Economic Outlook*, pp. 113–152. International Monetary Fund.

Canova, F., and E. Pappa (2006): "The elusive costs and the immaterial gains of fiscal constraints," *Journal of Public Economics*, 90(8–9), 1391–1414.

Canzoneri, M., F. Collard, H. Dellas, and B. Diba (2011): "Fiscal Multipliers in Recessions," mimeo.

Chahrour, R., S. Schmitt-Grohé, and M. Uribe (2010): "A model-based evaluation of the debate on the size of the tax multiplier," Working Paper 16169, NBER.

Choi, W. G., and M. B. Devereux (2005): "Asymmetric Effects of Government Spending: Does the Level of Real Interest Rates Matter?," Working Paper WP/05/7, IMF.

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

Chung, H., and E. M. Leeper (2007): "What Has Financed Government Debt?," Working Paper 13425, NBER.

Coenen, G., R. Straub, and M. Trabandt (2012): "Fiscal Policy and the Great Recession in the Euro Area," *Amercian Economic Review*, 102(3), 71–76.

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(3), 281–295.

Corsetti, G., A. Meier, and G. J. Müller (2009): "Fiscal Stimulus with Spending Reversals," Discussion Paper 7302, CEPR.

——— (2010): "What determines government spending multipliers?," mimeo.

ECB (2011): "Ensuring Fiscal Sustainability," *ECB Monthly Bulletin*, April, 61–77.

Eggertsson, G. B. (2011): "What Fiscal Policy is Effective at Zero Interest Rates?," in *NBER Macroeconomics Annual 2010, Volume 25*, NBER Chapters, pp. 59–112. National Bureau of Economic Research, Inc.

Erceg, C. J., and J. Lindé (2010): "Is There a Fiscal Free Lunch in a Liquidity Trap," International Finance Discussion Papers 1003, Board of Governors of the Federal Reserve System.

Favero, C., and F. Giavazzi (2007): "Debt and the Effects of Fiscal Policy," Working Paper 317, IGIER.

Fazzari, S. M., J. C. Morley, and I. Panovska (2012): "State-Dependent Effects of Fiscal Policy," Research Paper 2012-27, UNSW Australian School of Business.

Feenberg, D., and E. Coutts (1993): "An Introduction to the TAXSIM Model," *Journal of Policy Analysis and Management*, 12(1), 189–194.

Fernández-Villaverde, J., G. Gordon, P. Guerrón-Quintana, and J. F. Rubio-Ramírez (2012): "Nonlinear Adventures at the Zero Lower Bound," .

Galvão, A. B. (2006): "Structural Break Threshold VARs for Predicting US Recessions Using the Spread," *Journal of Applied Econometrics*, 21(4), 463–487.

Galvão, A. B., and M. Marcellino (2010): "Endogenous Monetary Policy Regimes and the Great Moderation," mimeo.

Ghosh, A. R., J. I. Kim, E. G. Mendoza, J. D. Ostry, and M. S. Qureshi (2011): "Fiscal Fatigue, Fiscal Space and Debt Sustainability in Advanced Economies," Working Paper 16782, NBER.

Giavazzi, F., T. Jappelli, and M. Pagano (2000): "Searching for non-linear effects of fiscal policy: Evidence from industrial and developing countries," Working Paper 7460, NBER.

Giavazzi, F., T. Jappelli, M. Pagano, and M. Benedetti (2005): "Searching for Non-Monotonic Effects of Fiscal Policy: New Evidence," Working Paper 11593, NBER.

Giavazzi, F., and M. Pagano (1990): "Can Severe Fiscal Contractions Be Expansionary? Tales of Two Small European Countries," Working Paper 3372, NBER.

——— (1995): "Non-Keynesian Effects of Fiscal Policy Changes: International Evidence and the Swedish Experience," Working Paper 5332, NBER.

Girouard, N., and C. André (2005): "Measuring Cyclically-adjusted Budget Balances for OECD Countries," OECD Economics Department Working Papers 434, OECD.

Gonzalo, J., and J.-Y. Pitarakis (2002): "Estimation and model selection based inference in single and multiple threshold models," *Journal of Econometrics*, 110(2), 319–352.

Hansen, B. E. (1996): "Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis," *Econometrica*, 64(2), 413–30.

——— (1999): "Testing for Linearity," *Journal of Economic Surveys*, 13(5), 551–576.

——— (2000): "Sample Splitting and Threshold Estimation," *Econometrica*, 68(3), 575–603.

Haugh, D., P. Ollivaud, and D. Turner (2009): "What drives sovereign risk premiums?," Economic Departments Working Papers 718, OECD.

Ilzetzki, E., E. G. Mendoza, and C. A. Végh (2010): "How Big (Small?) are Fiscal Multipliers?," mimeo.

IMF (2009): "Fiscal Rules—Anchoring Expectations for Sustainable Fiscal Policies," .

Juessen, F., L. Linnemann, and A. Schabert (2011): "Understanding default risk premia on public debt," .

Kirchner, M., J. Cimadomo, and S. Hauptmeier (2010): "Transmission of Government Spending Shocks in the Euro Area: Time Variation and Driving Forces," Working Paper 1219, ECB.

Leeper, E. M., T. B. Walker, and S.-C. S. Yang (2009): "Fiscal Foresight and Information Flows," Working Paper 14630, NBER.

Mertens, K., and M. O. Ravn (2010): "Fiscal Policy in an Expectations Driven Liquidity Trap," CEPR Discussion Papers 7931, C.E.P.R. Discussion Papers.

——— (2011): "A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers," .

Michaillat, P. (2012): "Fiscal Multipliers Over the Business Cycle," CEP Discussion Paper 1115.

Mountford, A., and H. Uhlig (2008): "What are the Effects of Fiscal Policy Shocks?," Working Paper 14551, NBER.

Nickel, C., P. Rother, and L. Zimmermann (2010): "Major public debt reductions: Lessons from the past, lessons for the future," Working Paper 1241, ECB.

Perotti, R. (1999): "Fiscal Policy in Good Times and Bad," *The Quarterly Journal of Economics*, 114(4), 1399–1436.

——— (2004): "Estimating the effects of fiscal policy in OECD countries," Working Paper 276, IGIER.

——— (2011): "Expectations and Fiscal Policy: An Empirical Investigation," .

Ramey, V. (2011): "Can Gocernment Purchases Stimulate the Economy?," *Journal of Economic Literature*, 49(3), 673–685.

Ramey, V. A. (2009): "Identifying Government Spending Shocks: It's All in the Timing," Working Paper 15464, NBER.

Ramey, V. A., and M. D. Shapiro (1998): "Costly capital reallocation and the effects of government spending," *Carnegie-Rochester Conference Series on Public Policy*, 48(1), 145–194.

Reinhart, C. M., and K. S. Rogoff (2010): "Growth in a Time of Debt," *American Economic Review*, 100(2), 573–78.

Romer, C. D., and D. H. Romer (2009): "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," mimeo.

Sutherland, A. (1997): "Fiscal crises and aggregate demand: can high public debt reverse the effects of fiscal policy?," *Journal of Public Economics*, 65, 147–162.

Teräsvirta, T. (1998): "Modelling Economic Relationships with Smooth Transition Regressions," in *Handbook of Applied Economic Statistics*, ed. by A. Ullah, and D. E. A. Giles, pp. 507–552. Marcel Dekker.

Tsay, R. S. (1998): "Testing and Modelling Multivariate Threshold Models," *Journal of the American Statistical Association*, 93(443), 1188–1202.

A Data

We use quarterly data for the US economy from 1960:1-2009:4. The source of all the variables are the NIPA accounts (available on the Bureau of Economic Analysis website), except for the time series for the stock of US public debt which is obtained from the FRED database (available on the Federal Reserve of St.Louis website). The definition of the variables is as follows:

— g_t: log of real per capita federal government total nominal expenditures minus net interest payments. Quarterly observations computed as

$$g_t = \log \frac{G - INT\_PAY}{POP * GDPDEF}$$

where

- G: federal government total nominal expenditure from line 40 in Table 3.2. It is seasonally adjusted at annual rates.
- INT_PAY: federal government nominal interest payments from line 29 in Table 3.2. It is seasonally adjusted at annual rates.
- GDPDEF: The price index for GDP from line 1 in Table 1.1.4. It is seasonally adjusted and the base year is 2005.
- POP: midperiod population from line 39 of Table 2.1.

— t_t: log of real per capita federal government total receipts . Quarterly observations computed as

$$t_t = \log \frac{T - INT\_REC}{POP * GDPDEF}$$

where

- T: federal government total nominal receipts from line 37 in Table 3.2. It is seasonally adjusted at annual rates.
- INT_REC: federal government nominal interest receipts from line 13 in Table 3.2. It is seasonally adjusted at annual rates.

— y_t: log of per capita GDP. Quarterly observations computed as

$$y_t = \log \frac{GDP}{POP * GDPDEF}$$

where

- GDP: nominal GDP from line 1 in Table 1.1.5. It is seasonally adjusted at annual rates.

— Δp_t: GDP deflator inflation rate. Quarterly observations computed as

$$\pi_t = \log GDPDEF_t - \log GDPDEF_{t-1}$$

— i_t: nominal cost of financing the debt. Quarterly observations computed as

$$i_t = \frac{INT\_PAY_t}{DEBT_{t-1}}$$

where

- DEBT: Federal Debt Held by the Public recomputed by the authors.

B Computing output and price elasticities

The elasticities α_{ty}, $\alpha_{t\Delta p}$, α_{ti}, α_{gy}, $\alpha_{g\Delta p}$ and α_{gi} represent the automatic response of fiscal variables to economic activity and are computed using external information. We compute these elasticities based on the approach of Perotti (2004). But we also incorporate features developed by Caldara (2011) to improve the identification approach of Blanchard and Perotti (2002):

1. α_{ty} represents the output elasticity of government revenues. It is computed as the weighted average of five components (Personal Income Tax, Social Security Contributions, Corporate Income Tax, Indirect Taxes and Transfers):

$$\alpha_{ty} = \sum_{i=1}^{5} \eta_{T_i,y} \frac{T_i}{T}$$

where $\eta_{T_i,y}$ is the elasticity of taxes type i to output and the weights are the share of taxes type i from tax revenues. The five elasticities are computed as

$$\eta_{T_i,y} = \eta_{T_i,TB_i} \eta_{TB_i,y}$$

where η_{T_i,TB_i} is the elasticity of taxes type i to their tax base and $\eta_{TB_i,y}$ is the elasticity of the tax base to output. We follow Perotti (2004)and assume that the output elasticity of the tax base for Indiferet Taxes is one. The output elasticities of the remaining tax bases are estimated using the following specification:

$$\Delta \log(TB_i) = \alpha_{i,0} + \eta_{TB_i,y} \Delta \log(y_t) + \varepsilon_{i,t}$$

where y_t is real GDP and $\varepsilon_{i,t}$ is assumed to be autocorrelated. The tax bases for the five categories are defined as follows:

(a) Personal Income Tax and Social Security Contributions: Compensation of employees.

(b) Corporate Income Tax: Proprietors' income with IVA and CCAdj plus Corporate profits with IVA and CCAdj plus Rental income of individuals.

(c) Transfers: Civilian Unemployment rate.

The elasticity of tax type i to their tax base is computed for the five categories as follows:

(a) Personal Income Tax: The OECD estimates the elasticity of this tax type to its base (Girouard and André, 2005), but the estimates are only available for every fifth year from 1979. We follow Caldara (2011) and use the NBER estimates from the TAXSIM model for this elasticity which are available at an annual frequency over our entire sample period (Feenberg and Coutts, 1993).[37]

(b) Social Security Contributions, Corporate Income Tax and Indirect Taxes: We use the OECD estimate for these categories.

(c) Transfers: Most of the transfers (pensions, disability benefits) do not respond automatically to changes in output and assumed to have a zero elasticity. However, we assume that unemployment insurance has an automatic response. Hence the elasticity is computed as the weighted average of these two components, i.e. the share of unemployment benefit from transfers.

2. $\alpha_{t\Delta p}$ represents the price elasticity of government revenues and computed similarly to the output elasticity. In particular, it is compted as the weighted average of the following five sub-elasticities:

   (a) Personal Income Tax: As shown by Perotti (2004) the price elasticity of real revenues, holding constant employment, output and the real wage is given by the elasticity of Personal Income Taxes to their tax base minus 1.

   (b) Social Security Contributions: It is constructed similarly to the price elasticity of Personal Income Tax.

   (c) Corporate Income Tax and Indirect Taxes: Following Perotti (2004) they are assumed to be 0.

   (d) Transfers to Individuals: Following Perotti (2004) it is assumed to be -1.

3. α_{ti} represents the interest rate elasticity of government revenues. Following Perotti (2004) it is assumed to be 0.

[37]We use the elasticities corresponding to the Federal Income Tax system available at `http://users.nber.org/~taxsim/allyup/ally.html`.

4. α_{gy} represents the output elasticity of government expenditures. Following Perotti (2004) it is assumed to be 0

5. $\alpha_{g\Delta p}$ represents the price elasticity of government expenditures. Following Perotti (2004) it is assumed to be -0.5.

6. α_{gi} represents the interest rate elasticity of government expenditures. Following Perotti (2004) it is assumed to be 0.

C Computing generalized impulse responses

Linear model Suppose we are interested in computing responses up to horizon $t+s$ to the structural shock ε which hits our system at time t:[38]

$$GI_y(\varepsilon) = E\{\mathbf{y}_{t,t+s}|\varepsilon\} - E\{\mathbf{y}_{t,t+s}|0\}$$

where $\mathbf{y}_{t,t+s}$ denotes the history of the endogenous variables between period t and $t+s$. We obtain $GI_y(\varepsilon)$ through the following two steps:

1. We compute the impulse responses

$$GI_y(\varepsilon, \mathbf{\Theta}_t) = E\{\mathbf{y}_{t,t+s}|\varepsilon, \mathbf{\Theta}_t\} - E\{\mathbf{y}_{t,t+s}|0, \mathbf{\Theta}_t\}$$

   that are conditional on the history of model variables, both the endogenous variables and the debt-to-GDP ratio, leading up to this period, that is $\mathbf{\Theta}_t = \begin{bmatrix} \mathbf{X}_t' & \mathbf{D}_t' \end{bmatrix}'$:

   (a) Draw a sample from the estimated residuals of the model, $\varepsilon_{t,t+s}$.

   (b) Generate a baseline simulation $\mathbf{y}^{bs,l}_{t,t+s}$ for the endogenous variables and $d^{bs,l}_{t,t+s}$ for the debt-to-GDP ratio by solving the model equations (1a) and (1b) forward conditional on the initial condition $\mathbf{\Theta}_t$. The superscript bs stands for baseline simulation, while l denotes the number of the current replication (see point 1e). Use the resampled residuals from step 1a for the simulation.

   (c) Generate an alternative simulation $\mathbf{y}^{as,l}_{t,t+s}$ for the endogenous variables and $d^{as,l}_{t,t+s}$ for the debt-to-GDP ratio by solving the model forward conditional on $\mathbf{\Theta}_t$. The superscript as stands for alternative simulation. Use the resampled residuals from step 1a for the simulation with the first realization perturbed by the shock, i.e. replace ε_t with $\varepsilon_t + \varepsilon$.

[38] Artis, Galvão, and Marcellino (2007) refer to ε as the "extraordinary" shock to distinguish it from the shocks they draw from the estimated normal distribution of the residuals to simulate time series from their model. We resample the estimated residuals both for the linear and the nonlinear VAR instead to simulate time series from our model. Therefore, we will refer to ε as a structural shock, or shock for short, as opposed to the residuals.

(d) Compute the impulse response of replication l as $\mathbf{y}^{as,l}_{t,t+s} - \mathbf{y}^{bs,l}_{t,t+s}$. Notice that this response is conditional on the particular history $\varepsilon_{t,t+s}$.

(e) Repeat steps 1a to 1d for $l = 1, \dots, L$ to average out the effects of future histories $\varepsilon_{t,t+s}$, which affect both the baseline and the alternative scenarios similarly.[39] Notice that we use the same draws to compute the two simulations, which guarantees that the only source of difference between them is the shock ε.

2. To obtain impulse responses $GI^{(j)}_y(\varepsilon)$ we average $GI_y(\varepsilon, \mathbf{\Theta}_t)$ over all histories within our sample.

Nonlinear model Suppose that the threshold variable is a function of the model variables:

$$z_t = f(\mathbf{Y}_t, \mathbf{X}_t, d_t, \mathbf{D}_t) \tag{2c}$$

We obtain $GI^{(j)}_y(\varepsilon)$, responses up to horizon $t+s$ to the structural shock ε which hits our system at time t through the following steps:

1. We compute the impulse responses

$$GI^{(j)}_y(\varepsilon, \mathbf{\Theta}^{(j)}_t) = E\{\mathbf{y}_{t,t+s}|\varepsilon, \mathbf{\Theta}^{(j)}_t\} - E\{\mathbf{y}_{t,t+s}|0, \mathbf{\Theta}^{(j)}_t\}$$

that are conditional on the history of model variables leading up to this period, that is $\mathbf{\Theta}^{(j)}_t = \begin{bmatrix} \mathbf{X}^{(j)\prime}_t & \mathbf{D}^{(j)\prime}_t & z^{(j)}_{t-d} & \dots & z^{(j)}_t \end{bmatrix}'$:

(a) Draw one sample for each regimes from the estimated residuals of the model, $\varepsilon^{(i)}_{t,t+s}$ $(i = 1, 2)$.

(b) Generate a baseline simulation $\mathbf{y}^{bs,l}_{t,t+s}$ for the endogenous variables, $d^{bs,l}_{t,t+s}$ for the debt-to-GDP ratio and $z^{bs,l}_{t,t+s}$ for the threshold variable by solving the model equations (2a)-(2c) forward conditional on the initial condition $\mathbf{\Theta}^{(j)}_t$. The superscript bs stands for baseline simulation, while l denotes the number of the current replication. Use the resampled residuals from step 1a for the simulation.

(c) Generate an alternative simulation $\mathbf{y}^{as,l}_{t,t+s}$ for the endogenous variables, $d^{as,l}_{t,t+s}$ for the debt-to-GDP ratio and $z^{as,l}_{t,t+s}$ for the threshold variable by solving the model forward conditional on $\mathbf{\Theta}_t$. The superscript as stands for alternative simulation. Use the resampled residuals from step 1a for the simulation with the first realization perturbed by the shock, i.e. replace $\varepsilon^{(j)}_t$ with $\varepsilon^{(j)}_t + \varepsilon$.

[39] We use $L = 2000$ for all of our simulations throughout the paper.

(d) Compute the impulse response of replication l as the difference between the alternative and the baseline scenarios. Notice that this response is conditional on the particular histories $\varepsilon_{t,t+s}^{(j)}$.

(e) Repeat steps 1a to 1d for $l = 1, \ldots, L$ to average out the effects of histories $\varepsilon_{t,t+s}^{(j})$, which affect both the baseline and the alternative scenarios similarly. Notice that we use the same draws to compute the two simulations, which guarantees that the only source of difference between them is the shock ε.

2. To obtain impulse responses $GI_y^{(j)}(\varepsilon)$ we average $GI_y^{(j)}(\varepsilon, \mathbf{\Theta}_t^{(j)})$ over all histories belonging to regime j.

Table 1. Candidate threshold variables

Related to	**Variable**
Financing the debt	real short-run rate
	real long-run rate
	real cost of financing the debt
Stationarity of debt	real short-run rate over GDP growth rate
	real long-run rate over GDP growth rate
	real cost of financing debt over GDP growth rate
Deficit	quarterly primary deficit-to-GDP ratio
	annual primary deficit-to-GDP ratio
	worse of the last two periods' quarterly deficit-to-GDP ratio
Debt stock	debt-to-GDP ratio
	cyclical component of real per capita debt
	quarterly change in the debt-to-GDP ratio
	year-on-year change in the debt-to-GDP ratio
Ability to repay	net interest payments-to-receipts ratio
	net interest payments-to-GDP ratio
	sensitivity of real short-run rate to debt-to-GDP ratio
	sensitivity of real long-run rate to debt-to-GDP ratio
	sensitivity of real cost of financing debt to debt-to-GDP ratio

Note: The real short term rate is computed using the 3-month T-bill rate. The real long term rate is computed using the 10-year government bond rate. The real cost of financing the debt is computed using the nominal cost of financing the debt. Real rates are computed using the GDP deflator inflation. For the definition of the nominal cost of financing the debt and the GDP deflator see Appendix A.

Table 2. The output and price elasticity values used for identification

	α_{gy}	$\alpha_{g\Delta p}$	α_{gi}	α_{ty}	$\alpha_{t\Delta p}$	α_{ti}
Linear model	0	-0.5	0	1.807	0.996	0
Good times	0	-0.5	0	1.809	0.997	0
Bad times	0	-0.5	0	1.799	0.993	0

Table 3. Value of the information criteria for the different nonlinear models

Threshold Variable	AIC	SC	HQC
Financing the debt			
real short-run rate	-36.16	-34.56	-35.37
real long-run rate	-36.15	-34.62	-35.36
real cost of financing the debt	-36.10	-34.83	-35.48
Stationarity of debt			
real short-run rate over GDP growth rate	-35.89	-34.28	-35.15
real long-run rate over GDP growth rate	-35.74	-34.45	-35.07
real cost of financing debt over GDP growth rate	-36.28	-34.66	-35.58
Deficit			
quarterly primary deficit-to-GDP ratio	-36.17	-34.64	-35.38
annual primary deficit-to-GDP ratio	-36.18	-34.61	-35.43
worse of the last two periods' quarterly deficit-to-GDP ratio	-36.31	-34.66	-35.50
Debt stock			
debt-to-GDP ratio	**-36.53**	**-35.22**	**-35.87**
cyclical component of real per capita debt	-35.88	-34.74	-35.35
quarterly change in the debt-to-GDP ratio	-36.17	-34.57	-35.43
year-on-year change in the debt-to-GDP ratio	-35.98	-34.54	-35.23
Ability to repay			
net interest payments-to-receipts ratio	-36.29	-34.77	-35.50
net interest payments-to-GDP ratio	-36.25	-34.75	-35.44
sensitivity of real short-run rate to debt-to-GDP ratio	-35.92	-34.42	-35.17
sensitivity of real long-run rate to debt-to-GDP ratio	-35.74	-34.28	-34.98
sensitivity of real cost of financing debt to debt-to-GDP ratio	-35.70	-34.38	-34.99

Table 4. Nonlinearity test results

	Variable addition test				**Tsay's predictive residuals test**			
Delay	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 1$	$d = 2$	$d = 3$	$d = 4$
Test statistics	103.5**	108.7**	125.8**	131.4**	146.3**	142.7**	145.7**	151.0**

Note: The test statistics of the variable addition and the predictive residuals tests have both a chi-squared distribution with 35 and 40 degrees of freedom, respectively. Two asterisks denote values significant at 1 percent.

Table 5. Estimated feedback coefficients from the debt-to-GDP ratio

Lag	Model	g_t	t_t	y_t	Δp_t	i_t
d_{t-1}	Linear	0.53	-1.60	-0.53	-0.00	-0.26
		(0.94)	(-2.08)	(-2.33)	(-0.03)	(-4.25)
	Non linear, good times	0.39	-1.92	-0.47	0.06	-0.33
		(0.54)	(-1.89)	(-1.62)	(0.58)	(-4.17)
	Non linear, bad times	-0.38	0.49	-0.36	-0.15	-0.05
		(-0.60)	(0.73)	(-1.35)	(-2.71)	(-0.93)
d_{t-2}	Linear	-0.57	1.64	0.53	-0.01	0.26
		(-1.00)	(2.12)	(2.33)	(-0.18)	(4.30)
	Non linear, good times	-0.49	1.92	0.48	-0.08	0.34
		(-0.68)	(1.89)	(1.66)	(-0.81)	(4.23)
	Non linear, bad times	0.33	-0.49	0.37	0.15	0.06
		(0.52)	(-0.70)	(1.34)	(2.66)	(1.03)

Note: The table shows the coefficients from the matrices $\mathbf{\Gamma}$ and $\mathbf{\Gamma}^{(j)}$ in equations (1) and (2), respectively. T-statistics are reported in the brackets.

Table 6. The reduced form equivalents of the structural shocks

Structural shock	e_g	e_t	e_y	$e_{\Delta p}$	e_i
	Expenditure shock				
Good times ($u_g^{(1)}$)	1.00	-0.16	0.03	-0.01	-0.02
Bad times ($u_g^{(2)}$)	0.99	-0.02	0.12	0.02	-0.01
	Revenue shock				
Good times ($u_t^{(1)}$)	0.02	-0.92	0.06	-0.03	-0.00
Bad times ($u_t^{(2)}$)	-0.01	-0.91	0.04	0.02	-0.00

Note: Each row contains the reduced form equivalent of a one standard deviation structural shock used to simulate the impulse responses.

Table 7. The fiscal multiplier as a function of the size and the sign of the shock

Size of the shock	**Linear model**			**Good times**			**Bad times**		
(% of GDP)	**6**	**12**	**20**	**6**	**12**	**20**	**6**	**12**	**20**
	quarters after impact								
Expenditure multipliers									
5.0				1.01	1.33	1.33	0.47	0.51	0.60
1.0				1.13	1.51	1.51	0.49	0.57	0.69
0.2	0.75	1.17	1.43	1.14	1.54	1.54	0.49	0.59	0.72
-0.2				1.14	1.55	1.55	0.50	0.59	0.72
-1.0				1.15	1.56	1.56	0.50	0.59	0.72
-5.0				1.17	1.60	1.60	0.52	0.64	0.80
Revenue multipliers									
-5.0				0.99	1.23	1.23	-0.29	-0.48	-0.26
-1.0				1.01	1.32	1.32	-0.22	-0.21	0.25
-0.2	0.84	1.24	1.54	1.01	1.33	1.33	-0.19	-0.09	0.47
0.2				1.01	1.35	1.35	-0.17	-0.03	0.63
1.0				1.01	1.36	1.36	-0.13	0.08	0.85
5.0				1.00	1.42	1.42	0.06	0.87	1.38

Note: The table shows the expansionary effect of the expenditure and revenues shocks with different size and sign. Expenditure multipliers are computed as the response of output divided by the size of the shock $\left(\frac{y}{g}\frac{\Delta g}{\Delta y}\right)$. Revenue multipliers are computed as the response of output divided by the size of the shock multiplied by minus one $\left(-\frac{y}{t}\frac{\Delta t}{\Delta y}\right)$ in order to express their expansionary effect and to facilitate comparison with expenditure multipliers.

Table 8. The Recovery Act shocks used in our simulation (percent of GDP)

Period	**2009Q2**	**2009Q3**	**2009Q4**	**2010Q1**	**2010Q2**	**2010Q3**	**2010Q4**
Expenditure	0.28	0.42	0.56	0.67	0.73	0.67	0.52
Revenue	-0.36	-0.43	-0.40	-0.32	-0.19	-0.12	-0.12

Table 9. Values of the information criteria for structural break models

Model	**AIC**	**SC**	**HQC**
ET-VAR with debt-to-GDP ratio	**-36.53**	**-35.22**	**-35.87**
SB-VAR	-36.04	-34.71	-35.38
ESB-VAR	-36.06	-34.75	-35.40

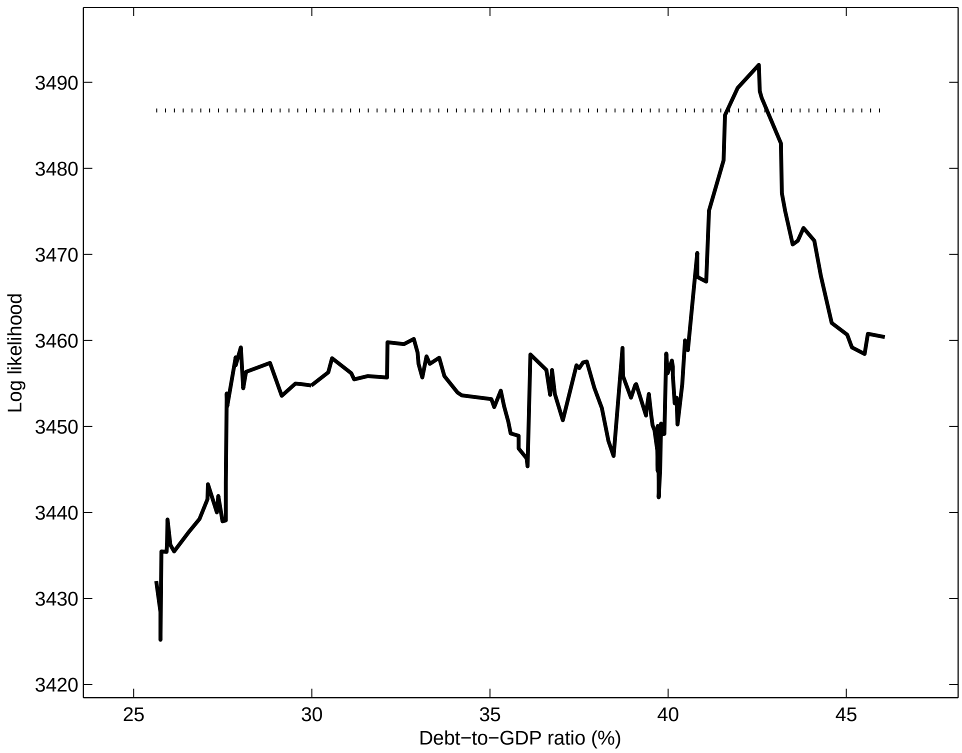

FIGURE 1. The log likelihood as a function of the threshold value. The intersections of the dotted horizontal line with the log likelihood function form the 1 percent confidence interval around the estimated threshold value based on the LR-statistics approach of Hansen (2000).

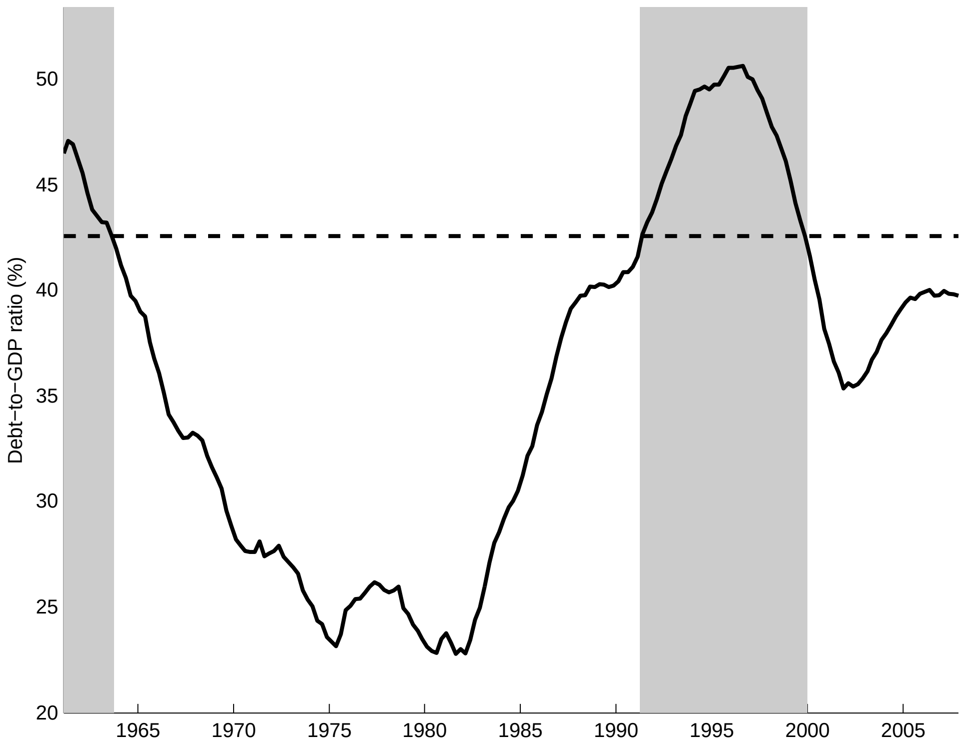

FIGURE 2. The history of regimes. The dashed horizontal line represents the estimated threshold value. The shaded periods belong to the regime with high debt-to-GDP ratio.

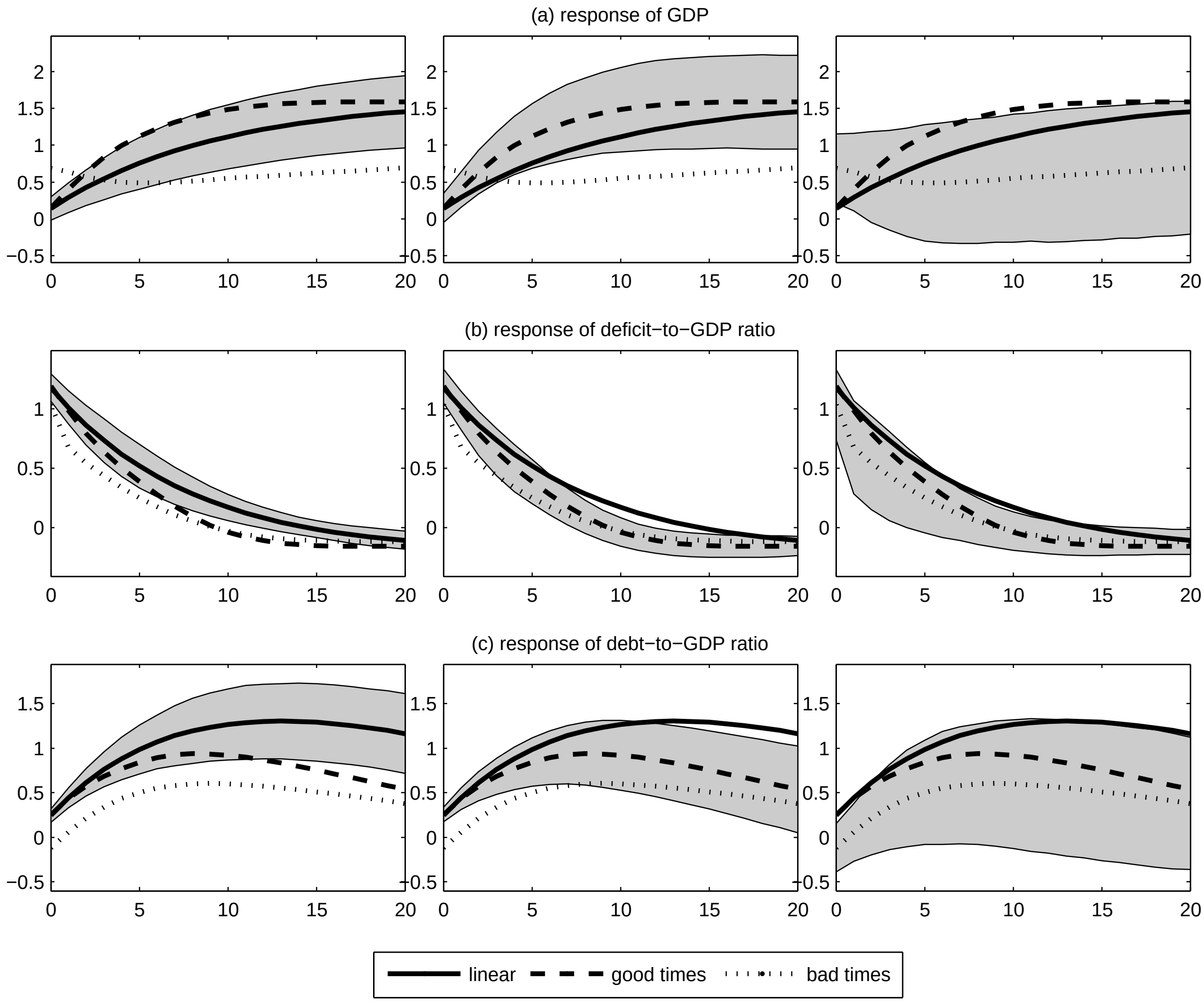

FIGURE 3. The impulse responses to an expenditure shock. We use a structural shock equivalent to 1 percent of GDP during our simulations. The shaded area represents a one standard deviation confidence band for the impulse responses of the linear model, good times and bad times in the first, second and third columns respectively.

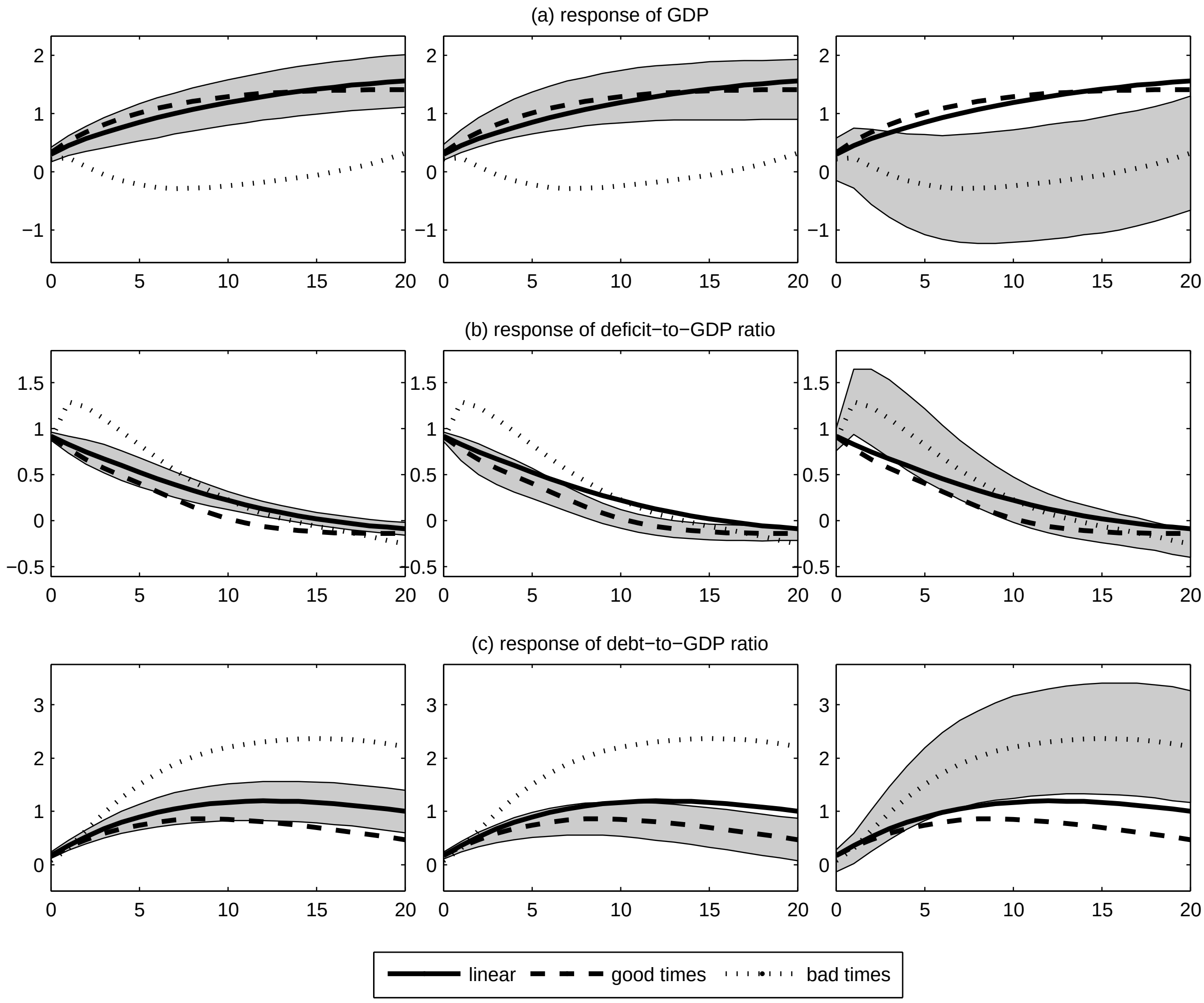

FIGURE 4. The impulse responses to a revenue shock. We use a structural shock equivalent to 1 percent of GDP during our simulations. The shaded area represents a one standard deviation confidence band for the impulse responses of the linear model, good times and bad times in the first, second and third columns respectively.

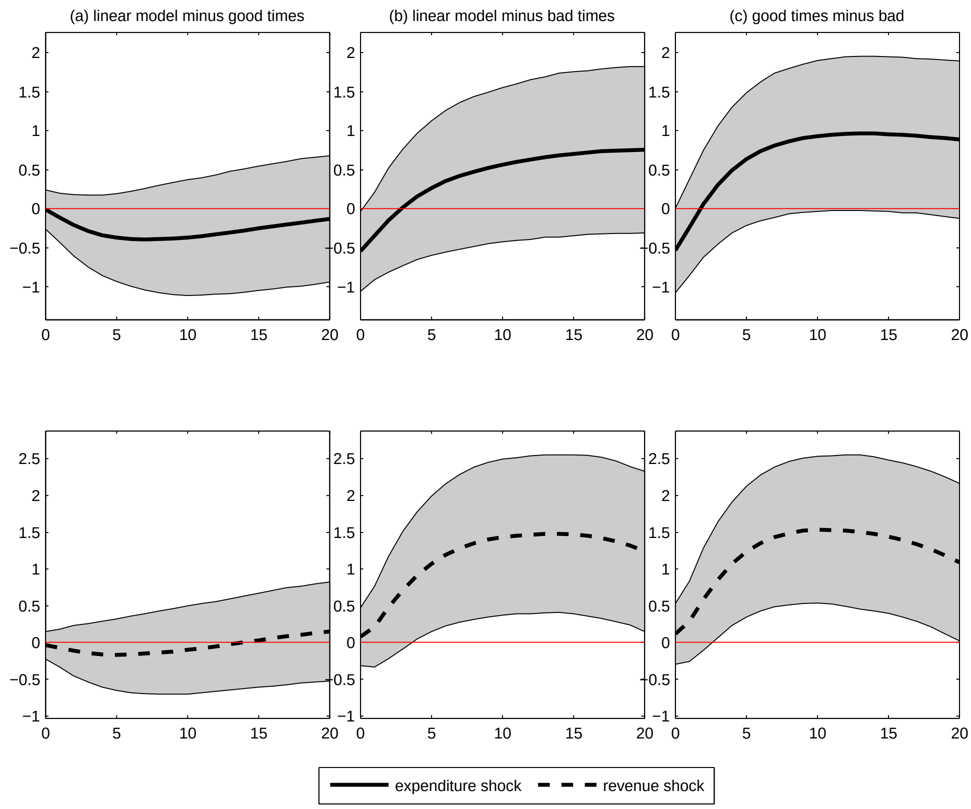

FIGURE 5. Pairwise differences of the output multipliers between models and regimes. The shaded area represents a one standard deviation confidence band.

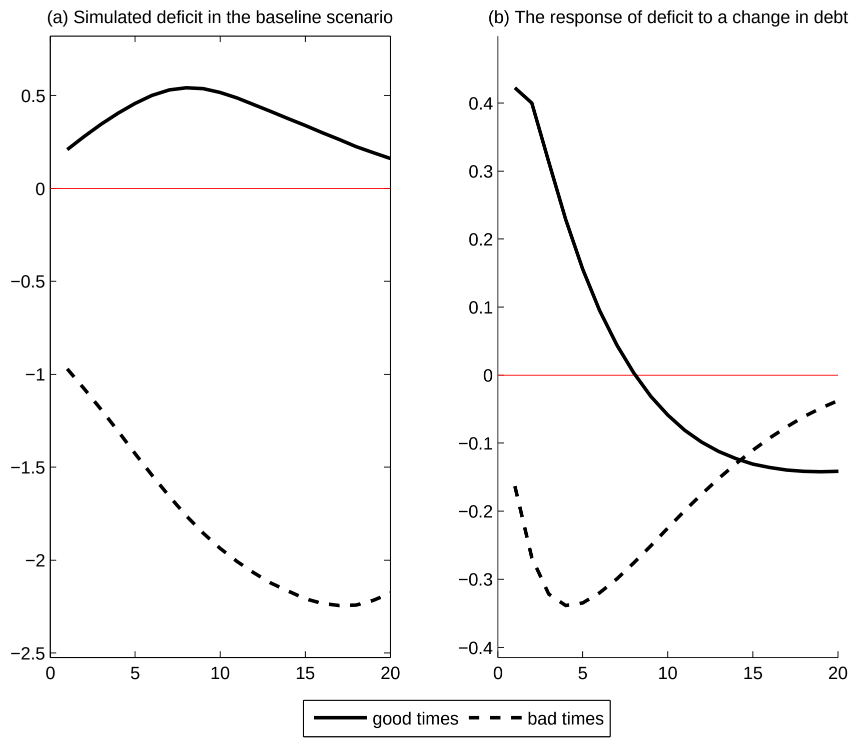

FIGURE 6. The behavior of deficit in the nonlinear model. The left panel depicts the mean simulated deficit in the baseline scenarios in the two regimes. The right panel depicts the mean response of deficit to changes in the debt-to-GDP ratio in each regime.

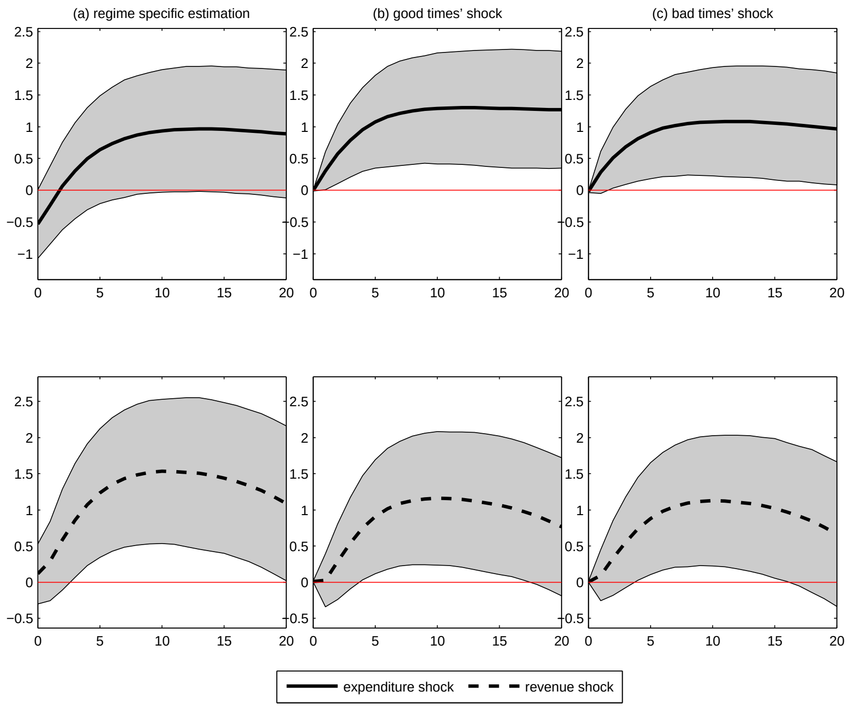

FIGURE 7. Pairwise differences of the output multipliers between regimes. The shaded area represents a one standard deviation confidence band. The panels in the left column are based on simulations that use the identified shocks from our regime specific identification. The second and third columns are based on simulations where we use the same identified shocks in both regimes: the identified good and bad times' shocks in the second and third columns, respectively.

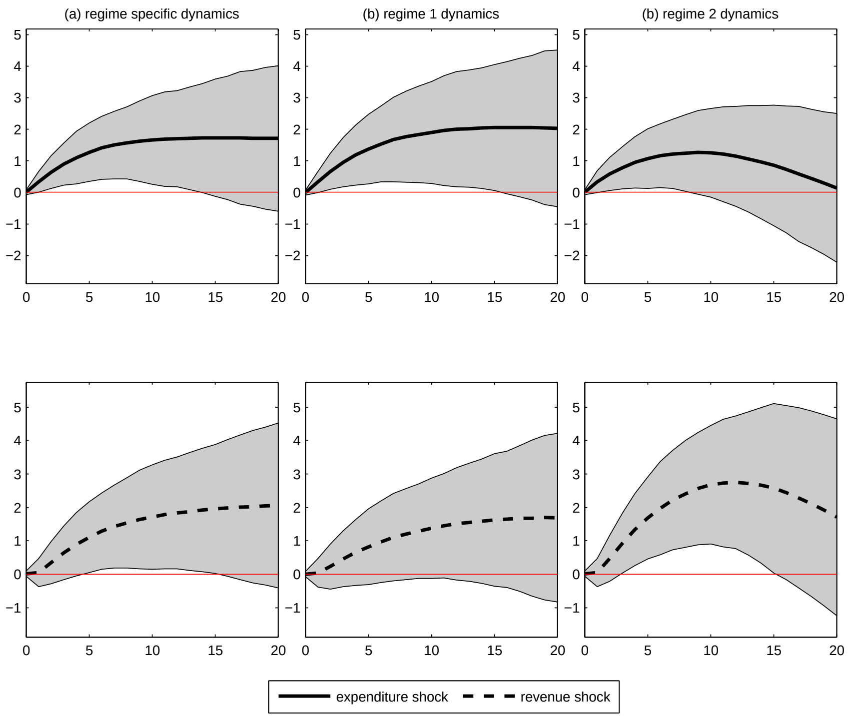

Figure 8. Pairwise differences of the output multipliers between regimes of the counterfactual impulse responses. The shaded area represents a one standard deviation confidence band. The panels in the left column are based on simulations from the benchmark model with identical structural shocks in both regimes. The second and third columns are based on simulations from a counterfactual exercise where we use the same fiscal rules in both regimes; the estimated fiscal rules from good and bad times' in the second and third columns, respectively.

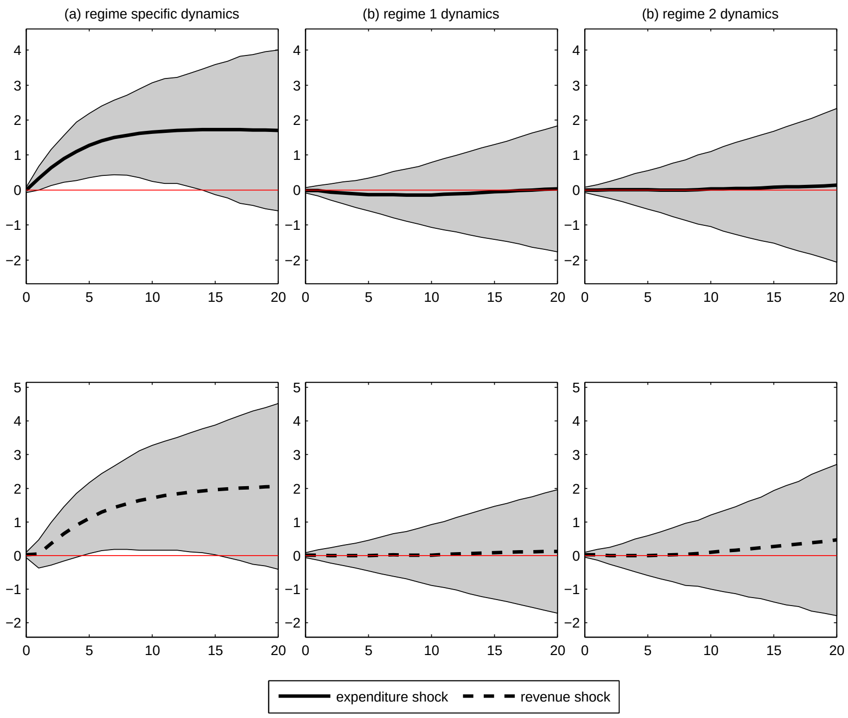

FIGURE 9. Pairwise differences of the output multipliers between regimes of the counterfactual impulse responses. The shaded area represents a one standard deviation confidence band. The panels in the left column are based on simulations from the benchmark model with identical structural shocks in both regimes. The second and third columns are based on simulations from a counterfactual exercise where we use the same estimated coefficients in all but the interest rate equation in both regimes; the estimated coefficients from good and bad times' in the second and third columns, respectively.

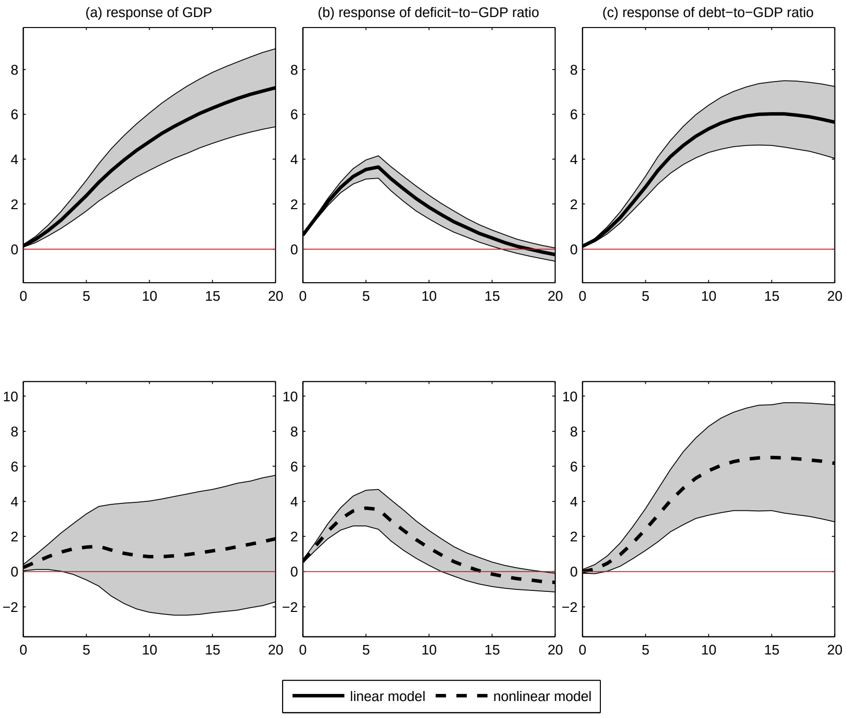

FIGURE 10. The simulated effects of The Recovery Act in the linear and the nonlinear models. The shaded area represents a one standard deviation confidence band for the impulse responses.

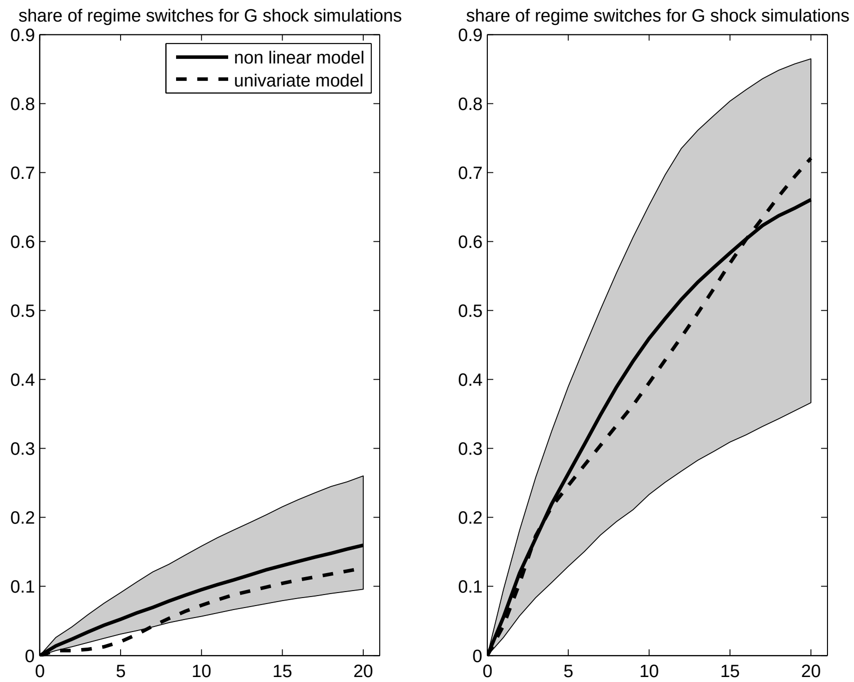

FIGURE 11. Regime switching probabilities in our benchmark nonlinear model and in a univariate model of the threshold variable. The shaded area represents a one standard deviation confidence band. The left panel plots the regime switching probabilities at each horizon for simulations starting in good times approximated by the fraction of the simulations being in bad times. The right panel plots the probabilities for bad times computed in a similar way.

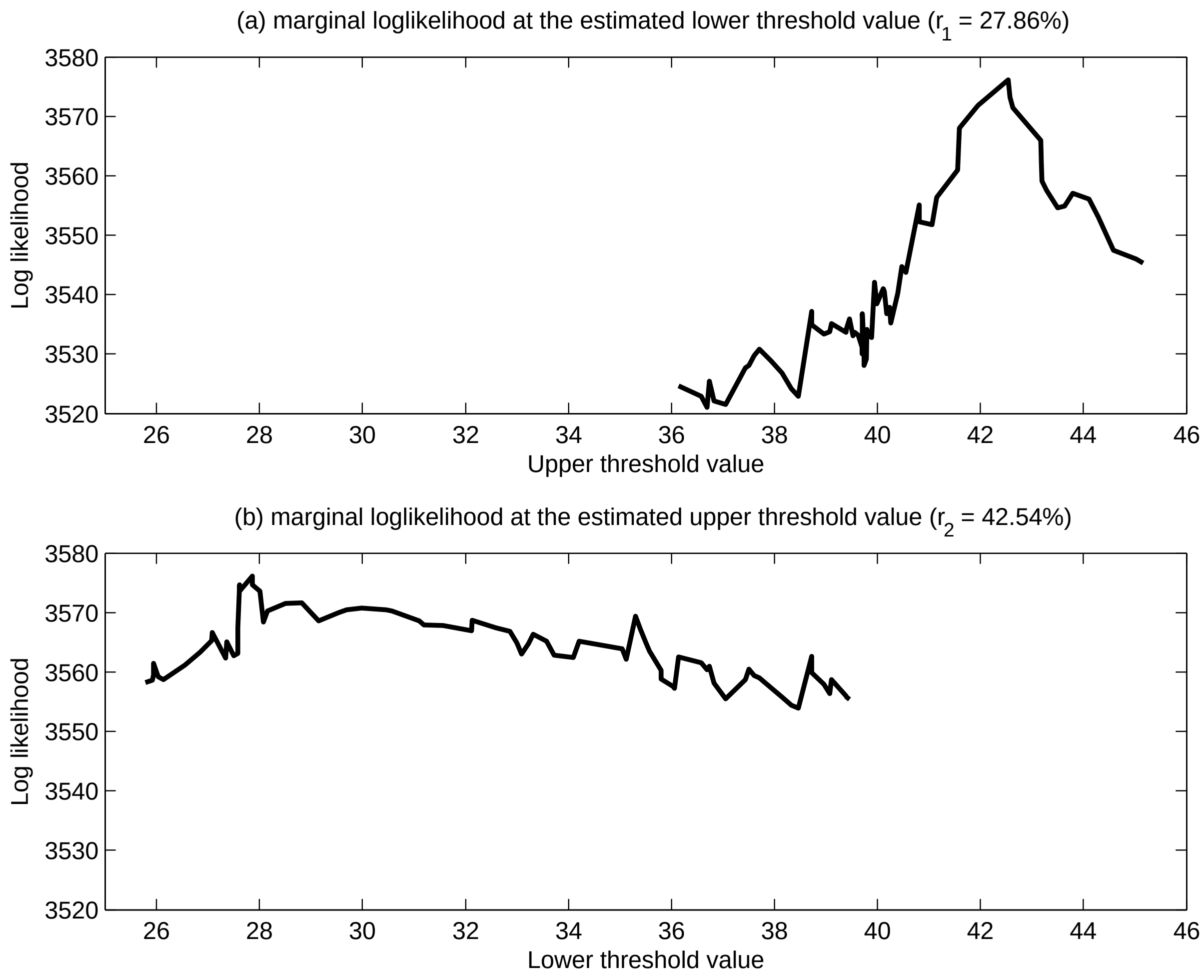

FIGURE 12. The marginal log likelihood as a function of the threshold values. The top panel plots the cross-section of the log likelihood function $\mathscr{L}(r_1, r_2)$ with the lower threshold value (r_1) fixed at its point estimate. The bottom panel plots the cross-section of the log likelihood function with the upper threshold value (r_2) fixed at its point estimate.