DISCUSSION PAPER SERIES

No. 10584

STIMULUS VERSUS AUSTERITY: THE ASYMMETRIC GOVERNMENT SPENDING MULTIPLIER

Régis Barnichon and Christian Matthes

INTERNATIONAL MACROECONOMICS and PUBLIC ECONOMICS

Centre for Economic Policy Research

ISSN 0265-8003

STIMULUS VERSUS AUSTERITY: THE ASYMMETRIC GOVERNMENT SPENDING MULTIPLIER

Régis Barnichon and Christian Matthes

Discussion Paper No. 10584
May 2015
Submitted 27 April 2015

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801
www.cepr.org

STIMULUS VERSUS AUSTERITY: THE ASYMMETRIC GOVERNMENT SPENDING MULTIPLIER[†]

Abstract

Despite intense scrutiny estimates of the government spending multiplier remain highly uncertain with values ranging from 0.5 to 2. While a fiscal consolidation is generally assumed to have the same (mirror-image) effect as a fiscal expansion, we show that relaxing this assumption is crucial to understanding the effects of fiscal policy. The government spending multiplier is substantially below 1 for fiscal expansions, but the multiplier is substantially above 1 for fiscal consolidations.

JEL Classification: C32 and E62
Keywords: fiscal policy and Gaussian mixture approximation

Régis Barnichon   rbarnichon@crei.cat
CREI, Universitat Pompeu Fabra and CEPR

Christian Matthes   Christian.Matthes@rich.frb.org
Federal Reserve Bank of Richmond

[†] We would like to thank Jordi Gali, Peter Karadi, Silvana Tenreyro, Helene Rey and seminar participants for helpful comments. Barnichon acknowledges financial support from the Spanish Ministerio de Economia y Competitividad (grant ECO2011-23188) and the Barcelona GSE Research Network. The views expressed here do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. Any errors are our own.

1 Introduction

Understanding the impact of expansionary fiscal policy on output is a central part of fiscal policy analysis, and the question received a lot of attention as fiscal stimuli were passed in many OECD countries during the early stages of the 2008-2009 recession.[1] As government debt levels rose rapidly, a swift shift to fiscal austerity followed (particularly in continental Europe) and spurred work on the mirror-image question –the effect of contractionary fiscal policy–.[2] However despite intense scrutiny, the range of estimates of the government spending multiplier remains wide with estimates lying between 0.5 and 2.

Perhaps surprisingly, the literature has so far treated the effects of expansionary and contractionary policy symmetrically; standard techniques are linear with respect to the sign of the policy intervention, and a contractionary policy is assumed to have the same (mirror-image) effect as an expansionary policy. In this paper, we relax this assumption, and we find that treating expansionary and contractionary fiscal interventions separately is crucial to understanding the effects of fiscal policy.

We allow the effect of fiscal policy to depend on the sign of the intervention, and we find that the government spending multiplier is substantially below 1 for fiscal expansions, but that the multiplier is substantially above 1 for fiscal consolidations. In other words, while the "austerity multiplier" is significantly above 1, the "stimulus multiplier" is significantly below 1.

To allow the effects of fiscal policy to depend on the nature of the policy intervention, we build on and extend the approach proposed by Barnichon and Matthes (BM, 2014) and we parametrize the impulse response functions to fiscal policy shocks using mixtures of Gaussian functions. We extend the BM method by showing that, given a series of fiscal policy shocks, the problem of identifying the non-linear effects of a fiscal shock on a variable of interest can be recast into a linear one. As a result, the standard linear regression toolkit can be used and estimation is straightforward and instantaneous.

Treating expansionary and contractionary fiscal shocks separately also allows us to reconcile seemingly contradictory findings in the literature on government multipliers: While estimates from shocks to defense spending (Ramey, 2011, Barro and Redlick 2011) are generally significantly less than 1, estimates from shocks recovered with a recursive identification scheme (Blanchard and Perotti, 2002, Auerbach and Gorodnichenko (AG), 2012) are close to and slightly above 1. More recently, Jorda and Taylor (2013) find even larger multipliers with

[1] See e.g., Blanchard and Perotti (2002), Hall (2009), Ramey (2009), Mertens and Ravn (2010), Barro and Redlick (2011), Parker (2011), Ramey (2011), Owyang, Ramey and Zubairy (2013) and Ramey and Zubairy (2014).

[2] See e.g., Alesina and Ardagna (2010), Guajardo, Leigh, and Pescatori (2011) and Jorda and Taylor (2013).

values significantly above 1 and closer to 2.

We argue that these different results stem from variations in the sample of fiscal shocks used for identification in each method and more specifically from variations in the relative importance of expansionary and contractionary shocks.[3] Results from the narrative approach on news to defense spending (e.g., Ramey, 2011, Barro and Redlick 2011) are driven primarily by positive shocks –unexpected *increases* in government spending–, because the Ramey news shock series contain much fewer negative shocks than positive shocks. The estimated multiplier is small, because the multiplier associated with positive shocks is small. In contrast, the spending shocks identified by AG are much more evenly distributed between positive and negative values. As a result, AG find a larger multiplier, lying in between our estimates of the expansionary multiplier and the contractionary multiplier. Finally, Jorda and Taylor (2013)'s estimates are much larger than the other lines of work simply because they strictly focus on fiscal consolidations. Their estimated multiplier is thus that of a contractionary multiplier, which we also find to be significantly larger than one.

Our "Gaussian Mixture Approximation" approach builds on two premises: (i) any mean-reverting impulse response function can be approximated to any degree of accuracy by a mixture of Gaussian functions, and (ii), in practice, only a very small number of Gaussian functions are needed to approximate a typical impulse response function. Intuitively, the impulse response functions of stationary variables are often found to be monotonic or hump-shaped. In such cases, a single Gaussian function can already provide an excellent approximation of the impulse response function.

Thanks to the small number of free parameters allowed by our Gaussian mixture approximation, it is possible to directly estimate the impulse response functions from the data using maximum likelihood or Bayesian methods.[4] The parsimony of the approach in turn allows us to estimate more general non-linear models.

Our use of Gaussian functions to approximate (and parametrize) impulse response functions builds on Barnichon and Matthes (2014) and relates to a large literature that relies on radial basis functions (of which Gaussian functions are one example) to approximate arbitrary multivariate functions (e.g., Buhmann, 2003) or to approximate arbitrary distributions using

[3]While Ramey (2011) compiles a series of shocks to anticipated future future defense expenditure using newspaper sources, Auerbach and Gorodnichenko (2012) rely on professsional forecasts of government spending to identify unexpected shocks.

[4]Another advantage of using Gaussian basis functions is that, in the context of Bayesian estimations, prior elicitation is much easier than with Bayesian estimations of standard VARs, because the coefficients to be estimated are directly interpretable as features of impulse responses. For instance, in the context of a one Gaussian function approximation, an impulse response function is characterized by 3 parameters; one capturing the peak effect of the shock, another capturing the timing of that peak effect, and a final one capturing the persistence of the effect of the shock.

a mixture of Gaussian distributions (Alspach and Sorenson 1971, 1972, McLachlan and Peel, 2000). In economics, our parametrization of impulse responses relates to an older literature on distributed lag models and in particular on the Almon (1965) lag specification, in which the successive weights, i.e., the impulse response function in our context, are given by a polynomial function.

Our estimation approach relates to the local projection method pioneered by Jorda (2005) to estimate impulse response functions. Local projection can also easily accommodate non-linearities in the response function.[5] However, local projection does not impose any underlying dynamic system and while this model-free approach is attractive, this can come at an efficiency cost (Ramey, 2012), which can make inferences on a rich set of non-linearities difficult. By positing that the response function can be approximated by a few Gaussian functions, our approach imposes strong dynamic restrictions between the parameters of the impulse response function, which in turn allow us to estimate a rich set of non-linearities.

Section 2 presents the empirical model, Section 3 presents the results; and Section 4 concludes.

2 Empirical model

In this section, we present our strategy to identify the (possibly non-linear) effects of government spending shocks on the economy. To isolate fiscal policy shocks, we use two alternative approaches: (i) a recursive identification scheme as in Blanchard-Perotti (BP, 2002) but augmented with expected changes in government spending as in Auerbach and Gorodnichenko (AG, 2012, 2013), and (ii) a narrative approach to identify unanticipated changes in future defense spending as in Ramey (2011). In each case, we show how to generalize the identification scheme to a non-linear setting by building on Barnichon and Matthes (2014) and by using Gaussian Mixture Approximations (GMA) of the impulse response functions.

Considering separately the two main identification schemes –a recursive identification scheme on the one hand a narrative approach on the other– has a number of advantages. First, it allows us assess the robustness of our findings across identification schemes but also allows us to inform the current debate on the size of the multiplier since the two different identification schemes lead to different conclusions.

Second, if allows us to show how our Gaussian Mixture Approximation method can be used in the two popular cases of impulse response estimation: (i) both the shocks and the impulse response functions are simultaneously recovered using a short-run identifying restriction (as in

[5] For instance, Auerbach and Gorodnichenko (2013) use local projection to estimate the extent of state dependence in the effect of monetary policy.

BP and AG), or (ii) a series of shocks have been previously identified and only the impulse response functions need to be recovered.

2.1 Gaussian Mixture Approximations (GMA) and short-run identifying restrictions

We start by describing our approach in the case where both the shocks and the impulse response functions are simultaneously recovered.

Our starting point is a general description of the economy, in which the behavior of a system of macroeconomic variables is dictated by its response to past and present structural shocks, i.e., we start from the vector moving average representation of the economy. Then, we approximate the vector moving average representation of the system with mixtures of Gaussian functions.

2.1.1 A Vector Moving Average representation

Denoting $\mathbf{Y}_t$ a vector of macroeconomic variables (e.g., output growth, inflation, policy stance), the of macroeconomic variables is dictated by its response to past and present structural shocks, or

$$\mathbf{Y}_t = \sum_{k=0}^{\infty} \mathbf{\Psi}_k(\boldsymbol{\varepsilon}_{t-k})\boldsymbol{\varepsilon}_{t-k} \tag{1}$$

where boldface letters are used to indicate vectors or matrices of variables or coefficients. $\boldsymbol{\varepsilon}_t$ is the vector of structural innovations with $E\boldsymbol{\varepsilon}_t = \mathbf{0}$ and $E\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' = \mathbf{I}$, $\mathbf{\Psi}_k(\boldsymbol{\varepsilon}_{t-k})$ is the matrix of lag coefficients.

Equation (1) can be seen as a non-linear vector moving average representation of the economy, in which the matrices $\{\mathbf{\Psi}_k\}$ –the coefficients of the impulse response functions to shocks– can depend on the values of the structural shocks in $\boldsymbol{\varepsilon}_{t-k}$. With $\mathbf{\Psi}_k$ a function of $\boldsymbol{\varepsilon}_{t-k}$, the impulse response functions to a given structural shock depend on the value (sign or size) of that shock. For instance, a positive shock may trigger a different impulse response than a negative shock.

2.1.2 Identifying restriction

To be able to identify structural shocks, we must impose an additional restriction on the system. As in the fiscal policy literature and Blanchard and Perotti (2002), we use a short-run restriction. Specifically, we use a recursive identification restriction, and fiscal policy is assumed to react with a lag to shocks affecting macro variables, so that government spending

G_t is ordered first in $\mathbf{Y}_t$ and $\boldsymbol{\Psi}_0$ is lower-triangular. Moreover, to address the issue of anticipation of fiscal policy changes (Ramey, 2011), we follow Auerbach and Gorodnichenko (2012) by augmenting the vector $\mathbf{Y}_t$ with professional forecasts of the growth rate of government spending $\Delta G^F_{t|t-1}$. This forecast is ordered first in $\mathbf{Y}_t$ and is meant to soak up the forecastable components of shocks to government spending G_t.

2.1.3 Using Gaussian functions to approximate impulse responses

If (1) admitted a VAR representation, the model could be estimated by first estimating a VAR for $\mathbf{Y}_t$ and then inverting it to get the moving average representation (1). However, this standard approach is only possible if (1) admits a VAR representation, which imposes that the model is linear, i.e., $\boldsymbol{\Psi}_k(\boldsymbol{\varepsilon}_{t-k}, \mathbf{Z}_{t-k}) = \boldsymbol{\Psi}_k$,[6] and prevents the study of possible non-linear effects of fiscal policy.

To address this issue, we follow Barnichon and Matthes (2014) and we directly estimate (1) using a mixture of Gaussian functions to parametrize the impulse response functions. Specifically, for each impulse response function $\psi_{ij}(k)$ (with $\psi_{ij}(k)$ the row-i column-j coefficient of $\boldsymbol{\Psi}_k$, i.e., the impulse response function of variable i to shock j), we will estimate a model of the form (omitting the ij subscripts for clarity of exposition):

$$\psi(k) \simeq \sum_{n=1}^{N} a_n e^{-(\frac{k-b_n}{c_n})^2}$$

with a_n, b_n, and c_n the parameters to be estimated.

To allow for asymmetry, we simply generalize the model to

$$\mathbf{Y}_t = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k^+ \mathbf{1}_{\boldsymbol{\varepsilon} \geq \mathbf{0}} \boldsymbol{\varepsilon}_{t-k} + \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k^- \mathbf{1}_{\boldsymbol{\varepsilon} < \mathbf{0}} \boldsymbol{\varepsilon}_{t-k} \tag{2}$$

with

$$\psi_{ij}^+(k) = a_{ij}^+ e^{-\left(\frac{k-b_{ij}^+}{c_{ij}^+}\right)^2}, \ \forall k > 0 \tag{3}$$

with a_{ij}^+, b_{ij}^+, c_{ij}^+ some constants to be estimated. A similar expression would hold for $\psi_{ij}^-(k)$.

In practice, to allow for more flexibility, we will let (3) only hold for $k > 0$, and we will leave the contemporaneous coefficient matrix $\boldsymbol{\Psi}_0$ unconstrained except that the elements above the main diagonal are zero (the structural identifying restriction) and that the diagonal elements

[6] In addition, the existence of a VAR representation imposes some restrictions on the $\boldsymbol{\Psi}_k$ coefficients. The matrices $\boldsymbol{\Sigma}_k \equiv \boldsymbol{\Psi}_0^{-1}\boldsymbol{\Psi}_k$ must be such that there exists $p > 0$ and matrices $\Phi_1, ..., \Phi_p$ such that $\boldsymbol{\Sigma}_k$ can be written as $\boldsymbol{\Sigma}_k = \Phi_k \boldsymbol{\Sigma}_{k-1} + ... + \Phi_p \boldsymbol{\Sigma}_{k-p}$, $\forall k > 0$.

are positive (a normalization). We also allow for intercepts in each equation.

Finally, to select N, the number of Gaussian functions used in the mixture, we will consider models with increasing N and select the lowest N for which a likelihood-ratio test cannot reject a model with N Gaussians in favor of a model with $N+1$ Gaussians.

2.1.4 Intuition and Motivation

The interest of our approach, and its use for studying the (possibly non-linear) effects of policy, rests on the fact that, in practice, only a very small number of Gaussian basis functions are needed to approximate a typical impulse response function.

Intuitively, impulse response functions of stationary variables are often found to be monotonic or hump-shaped.[7] And in such cases, one or two Gaussian functions can already provide a good approximate description of the impulse response. The small number of free parameters (only three per impulse response function in the one-Gaussian case), has two important advantages. First, it allows us to directly estimate the impulse response functions from the MA representation (1).[8] Second, it allows us to introduce (and estimate) asymmetric or non-linear effects of monetary policy.

The advantages are most striking in the case of one Gaussian function, i.e., using the approximation

$$\psi(k) \simeq a e^{-\frac{(k-b)^2}{c^2}}. \tag{4}$$

In that case, the a, b and c coefficients can be very easily interpreted in light of the impulse-response function. With a one-Gaussian parametrization, the impulse response function is summarized by three parameters –peak effect, timing of peak effect, and persistence of peak effect–, which are generally considered the most relevant characteristics of an impulse response function. As illustrated in Figure 1, parameter a is the height of the impulse-response, which corresponds to the maximum effect that a *unit* shock has on the variable of interest. In other words, a is the maximum marginal effect of a shock. Parameter b is the timing of this maximum effect. Parameter c captures how persistent the effect of the shock is. Specifically, for the latter term c, it is easy to show that the amount of time τ required for the effect of a shock to be 50% of its maximum value is given by

$$\tau = c\sqrt{\ln 2}.$$

[7]This is also the case in theoretical models, e.g., New-Keynesian models, in which the impulse response functions are generally monotonic or hump-shaped (see e.g., Walsh 2010).

[8]For instance, with 4 variables, we only have one $9 * 4 = 36$ parameters to estimate to capture the whole set of impulse response functions $\{\mathbf{\Psi}_k\}_{k=1}^{\infty}$.

2.2 Gaussian Mixture Approximations (GMA) and independently identified shocks

We now consider the case in which a series have been previously identified (for instance through the narrative approach as in Ramey, 2011). In that case, it is no longer necessary to specify a self-contained model capturing the relevant features of the economy (and thus a *vector* moving average model), and one only needs to estimate the impulse response functions to the shocks of interest.

Again, we start from a moving average representation, but this time we only need to consider the *scalar* moving average representation of the variable of interest y_t, which can be written as

$$y_t = \sum_{k=0}^{\infty} \psi(k)\varepsilon_{t-k}^{G} + \sum_{j}\sum_{k=0}^{\infty} \varphi_j(k)\varepsilon_{j,t-k} \tag{5}$$

with $\left\{\varepsilon_t^G\right\}$ the time series of government spending shocks and $\{\varepsilon_{j,t}\}$ time series of all the other j shocks affecting the economy. In words, the value of output at time t is the sum over history of all the past shocks that affected output.

As previously, the model may be non-linear, and $\psi(k)$ can depend on the value of the fiscal shock at $t-k$, i.e., $\psi(k) = \psi(k;\varepsilon_{t-k}^G)$ (and similarly for $\varphi_j(k)$). In such a non-linear world, and as in the previous section, there is no AR representation, because we cannot invert (5).

Re-writing (5), truncating the MA at $k = K$ and using a mixture of Gaussian functions to approximate $\psi(k)$ gives

$$y_t = \sum_{k=0}^{K} \psi(k)\varepsilon_{t-k}^{G} + \alpha + u_t \tag{6}$$

with

$$\psi(k) = \sum_{n=1}^{N} a_n e^{-(\frac{k-b_n}{c_n})^2}$$

where a_n, b_n and c_n can be functions of ε_{t-k}, and where the residual is $u_t \equiv \sum\limits_{j}\sum\limits_{k=0}^{K} \varphi_j(k)\varepsilon_{j,t-k}$. Since $\{u_t\}$ is serially correlated, in order to improve efficiency, we can allow for serial correlation in u_t by positing that u_t follows an $AR(p)$ process.

2.2.1 Recovering the $\{a_n\}$ factors given $\{b_n, c_n\}$

When the series of shocks has been independently identified (as is the case with the narrative approach), we now show that estimating the coefficients $\{a_n\}$ in (6), a non-linear problem, can be recast into a linear problem that can estimated by OLS.

Assume for now that the parameters of the Gaussian kernels $-\{b_n, c_n\}_{n=1}^{N}-$ are known, so that we have a "dictionary" of basis functions to decompose our impulse response. Then, recovering the $\{a_n\}_{n=1}^N$ is straightforward, even in the case with non-linearities.

Linear case For simplicity, we start with the linear case where $\psi(k)$ is independent of ε_{t-k}^G. We then re-arrange (6) as follows:

$$\begin{aligned}\sum_{k=0}^{K}\psi(k)\varepsilon_{t-k}^G &= \sum_{k=0}^{K}\sum_{n=1}^{N}a_n e^{-(\frac{k-b_n}{c_n})^2}\varepsilon_{t-k}^G \\ &= \sum_{n=1}^{N}a_n\sum_{k=0}^{K}e^{-(\frac{k-b_n}{c_n})^2}\varepsilon_{t-k}^G \\ &= \sum_{n=1}^{N}a_n\sum_{k=0}^{K}h_n(k)\varepsilon_{t-k}^G.\end{aligned}$$

Defining

$$X_{n,t} = \sum_{k=0}^{K}h_n(k)\varepsilon_{t-k}^G,$$

our estimation problem becomes a linear problem (conditional on knowing $\{b_n, c_n\}_{n=1}^N$):

$$y_t = \sum_{n=1}^{N}a_n X_{n,t} + \alpha + \beta u_t \tag{7}$$

where the $\{a_n\}$ parameters can be recovered instantaneously by OLS. Assuming that u_t follows an AR(p), we can estimate the $\{a_n\}$ with a simple NLS procedure.

Non-linear case The method described above in the linear case is straightforward to apply to a case with non-linearities. Consider the case with asymmetry

$$a_n(\varepsilon_{t-k}^G) = a_n^+ 1_{\varepsilon^G \geq 0} + a_n^- 1_{\varepsilon^G < 0}.$$

Then, we can proceed as in the previous section and define the following right-hand side variables

$$\begin{cases} X_{n,t}^+ = \sum\limits_{k=0}^{K}h_n(k)\varepsilon_{t-k}^G 1_{\varepsilon\geq 0} \\ X_{n,t}^- = \sum\limits_{k=0}^{K}h_n(k)\varepsilon_{t-k}^G 1_{\varepsilon< 0}\end{cases}$$

and use OLS to recover a_n^+ and b_n^-.

2.3 Choosing $\{b_n, c_n\}$

So far, we assumed that the parameters of the Gaussian kernels $-\{b_n, c_n\}_{n=1}^{N}-$ were given and taken from a dictionary. To choose $\{b_n, c_n\}_{n=1}^{N}$, we minimize the SSR of (7), and we select N (the number of Gaussian basis functions) using BIC and AIC criteria (taking N as the smallest number suggested by the two criteria).

2.4 Taking stock

There are two approaches to identify the non-linear effects of structural shocks: "Vector-GMA" described in section 2.1, and "Scalar-GMA" described in section 2.2. We think that both methods are complementary: They each have different pros and cons, and combining them helps address the main limitations of each method.

Vector-GMA has the important advantage of being self-contained: we estimate a non-linear model of the economy and recover simultaneously *both* the structural shocks (in this case from the recursive identification assumption) and the corresponding impulse-response functions. Its main limitation is that it suffers from the curse of dimensionality inherent to vector models such as VARs. The number of free parameters (and associated computation time) increases rapidly with the number of endogenous variables and the number of Gaussian functions.

In contrast, scalar-GMA is fast and does not suffer from the curse of dimensionality.[9],[10] As a result, one can approximate the moving average process with a larger number of Gaussian basis functions and thereby capture a much richer set of impulse responses. The disadvantage of scalar-GMA is that a series of structural shocks must have been previously recovered, either from a narrative approach or from the vector method. Another limitation is that the shocks much be serially uncorrelated, as otherwise, the estimates will not be consistent.

The combination of the two methods will be useful with respect to two objectives: (i) robustness checks, and (ii) computing additional impulse responses. First, once a series of structural shocks have been retrieved from the vector method, the scalar method can be used to quickly assess whether the number of Gaussian basis functions used in the vector method is large enough. Second, the scalar method allows for quick computation of the impulse responses of additional variables not included in the vector-GMA model.

[9] For instance, for the vector method with sign and state dependence and only one Gaussian basis function, taking 100,000 draws in the MCMC Metropolis-Hasting algorithm takes about 12 hours. With the scalar method, the same exercise takes only 10 minutes.

[10] Note the similarity with Jorda's (2005) local projection method. Jorda's method is model free and does not suffer from any curse of dimensionality, but it requires a series of previously identified shocks.

3 Estimation

Once parametrized, the GMA model (1) is estimated with Bayesian methods. In this section, we first describe how we construct the likelihood function by exploiting the prediction-error decomposition, and then we discuss the estimation routine.

3.1 Constructing the conditional likelihood function

We want to build the likelihood function for a sample of size T and parameter vector θ: $p(\mathbf{Y}_1, ..., \mathbf{Y}_T|\theta)$. To do this, we decompose the likelihood using:[11]

$$p(\mathbf{Y}_1, ..., \mathbf{Y}_T|\theta) = p(\mathbf{Y}_T|\mathbf{Y}_1, ..., \mathbf{Y}_{T-1}, \theta) \dots p(\mathbf{Y}_2|\mathbf{Y}_1, \theta)p(\mathbf{Y}_1|\theta) \tag{8}$$

To calculate the one-step-ahead conditional likelihood function needed for the prediction error decomposition, we proceed as follows. First, we assume that all innovations $\{\boldsymbol{\varepsilon}_t\}$ are Gaussian with mean zero and variance one.[12] We then use the MA representation (1) and the recursive identification scheme to uniquely identify the current period value of the innovations as a function of current period data and past innovations.

As an illustration, we discuss the construction of the likelihood in the case of asymmetry, i.e., when we consider the model (2), since the approach is identical in the more general non-linear models.

We first truncate the MA(∞) representation at some level K, and we set the first K values for $\boldsymbol{\varepsilon}$ to the zero vector.[13] Conditional on knowing the first $[t-K, t]$ innovations, the $t+1$ innovation is obtained from:

$$\boldsymbol{\Psi}_0^{\pm}\left(\boldsymbol{\varepsilon}_{t+1}\right)\boldsymbol{\varepsilon}_{t+1} = \mathbf{Y}_{t+1} - \sum_{k=1}^{K} \boldsymbol{\Psi}_k^{+} \mathbf{1}_{\boldsymbol{\varepsilon}\geq \mathbf{0}} \boldsymbol{\varepsilon}_{t+1-k} + \sum_{k=1}^{K} \boldsymbol{\Psi}_k^{-} \mathbf{1}_{\boldsymbol{\varepsilon}< \mathbf{0}} \boldsymbol{\varepsilon}_{t+1-k}$$

with $\boldsymbol{\Psi}_0^{\pm}\left(\boldsymbol{\varepsilon}_{t+1}\right)$ a function of $\boldsymbol{\varepsilon}_{t+1}$ since the elements of $\boldsymbol{\Psi}_0^{\pm}$, the contemporaneous lag parameter matrix, depends on the sign of the elements of the innovation vector $\boldsymbol{\varepsilon}_{t+1}$.

To calculate $\boldsymbol{\varepsilon}_{t+1}$ as a function of the parameters, one needs to uniquely pin down the value and sign of the components of $\boldsymbol{\varepsilon}_{t+1}$ as a function of the parameters. Fortunately, this

[11] To derive the conditional densities in decomposition (8), our parameter vector θ thus implicitly also includes initial values for the lagged innovations. We will keep those fixed throughout the estimation and discuss alternative initializations below.

[12] The estimation could easily be generalized to allow for non-normal innovations such as t-distributed errors. Note that even with Gaussian structural innovations, our non-linear model implies that the reduced-form residuals need not be Gaussian.

[13] Alternatively, we could use the first K values of the shocks recovered from a structural VAR.

is straightforward with a recursive identifying restriction scheme, when $\Psi_0^{\pm}(\boldsymbol{\varepsilon}_{t+1})$ is lower triangular (as is the case with a recursive identifying restriction.

Denoting $\psi_{0,ij}^{+}$ the element of Ψ_0^{+} and using the bivariate case as an illustration, we have

$$\Psi_0^{\pm}(\boldsymbol{\varepsilon}_{t+1}) = \begin{pmatrix} \psi_{0,11}^{\pm}(\varepsilon_{1,t+1}) & 0 \\ \psi_{0,21}^{\pm}(\varepsilon_{1,t+1}) & \psi_{0,22}^{\pm}(\varepsilon_{2,t+1}) \end{pmatrix} \tag{9}$$

with $\psi_{0,11}^{\pm}(\varepsilon_{1,t+1})$ indicating that the value of $\psi_{0,11}^{\pm}$ depends on $\varepsilon_{1,t+1}$, $\psi_{0,22}^{\pm}(\varepsilon_{2,t+1})$ depends on $\varepsilon_{2,t+1}$, etc... A similar form for would hold in a higher-dimensional case.

The fact that $\Psi_0^{\pm}(\boldsymbol{\varepsilon}_{t+1})$ is lower triangular with positive entries on the diagonal allows us to uniquely identify the value (and sign) of $\varepsilon_{1,t+1}$, the first component of ε_{t+1}, from the first equation with

$$\varepsilon_{1,t+1} = \frac{\mathbf{Y}_{1,t+1}}{\psi_{0,11}^{\pm}(\varepsilon_{1,t+1})}$$

and then uniquely identify the sign and value of each element of ε_t by iterating on (2). For instance, for $\varepsilon_{2,t+1}$, we have

$$\varepsilon_{2,t+1} = \frac{\mathbf{Y}_{2,t+1} - \psi_{0,21}^{\pm}(\varepsilon_{1,t+1})\varepsilon_{1,t+1}}{\psi_{0,22}^{\pm}(\varepsilon_{2,t+1})}. \tag{10}$$

Looking at (10), once the sign and value of $\varepsilon_{1,t+1}$ is known, the first column of $\Psi_0^{\pm}$ is pinned down (and thus $\psi_{0,21}^{\pm}(\varepsilon_{1,t+1})$ is known). Since the only unknown left in (10) is $\psi_{0,22}^{\pm}$, which is positive, we know the sign of $\varepsilon_{2,t+1}$ from the sign of $\mathbf{Y}_{2,t+1} - \psi_{0,21}^{\pm}(\varepsilon_{1,t+1})\varepsilon_{1,t+1}$. We then get the value of $\varepsilon_{2,t+1}$ from (10).

In a higher-dimensional case, we can proceed identically to uniquely identify the sign and value of each element of ε_{t+1}. With the series of $\{\varepsilon_t\}$ in hand, we can then construct the conditional log likelihood function from equation (2).

3.2 Estimation routine

We now discuss the estimation routine. To estimate both models –vector-GMA and scalar-GMA–, we use the Metropolis-Hastings algorithm (see Robert & Casella 2004, Haario et al., 2001). To be specific, we use a Metropolis-within-Gibbs algorithm with the blocks given by the different groups of parameters in our model (one block being composed of all a parameters, another composed of all b parameters and so on). As described above, we build the likelihood assuming a Normal distribution for the innovations, and we set the first K innovations to zero. We use improper uniform priors so that our results are interpretable as maximum-likelihood estimates.

3.2.1 Vector-GMA

In the vector-GMA case, we initialize the Metropolis-Hastings algorithm using a two-step procedure: first, we estimate a standard VAR using OLS on our data set, calculate the MA representation, and we use the impulse response functions implied by the VAR as our starting point. Specifically, we first fit our parametrized impulse response functions to the VAR-based impulse response functions.[14] We then use these coefficients as a starting point for our maximization routines[15] that then give us a starting value for the Metropolis-Hastings algorithm.

Finally, in order to reduce the parameter space, in our benchmark estimations, we only allow for asymmetry or non-linearity in the impulse response functions to monetary shocks. The impulse response functions to the other shocks are set at their values estimated from the corresponding SVAR.[16] However, as a robustness check, we verify that our findings are not driven by this restriction, by estimating a model with non-linearity in response to all shocks.

3.2.2 Scalar-GMA

In the scalar case, we initialize the Metropolis-Hastings algorithm as follows. Given a number N of Gaussian functions, we start with the Gaussian functions with randomly chosen $\{b_n\}$ and $\{c_n\}$ parameters (recall that the $\{a_n\}$ coefficients are then obtained straightforwardly by OLS). We then use Matlab's fminsearch to choose $\{b_n\}$ and $\{c_n\}$ to maximize the likelihood function and get a good initial guess. We then feed this initial guess into the Metropolis-Hastings algorithm and then obtain a distribution of the parameter estimates from the posterior distribution.

4 Results

We now present our results based on the two methods to identify government spending shocks: (i) the Blanchard-Perotti (2002) recursive identification scheme augmented with expected changes in government spending as in Auerbach and Gorodnichenko (2012, 2013), and (ii) the Ramey narrative approach based on news sources to identify unanticipated changes in future defense spending.

[14] Specifically, we set the parameters of our model (for instance, a, b and c in the benchmark model with one Gaussian function) to minimize the discrepancy (sum of squared residuals) between the two sets of impulse responses.

[15] We use both a simplex algorithm and Chris Sims' csminwel routine.

[16] We did a preliminary Metropolis-Hastings MCMC run in order to explore the likelihood around the VAR-based impulse response functions to non-monetary shocks, and we found that these impulse response functions are in an area of high posterior probability.

4.1 Shock identification from a recursive ordering

To identify innovations to government spending, we follow Auerbach and Gorodnichenko (AG, 2012), and we consider the vector $\left(\Delta G^F_{t|t-1}, G_t, T_t, Y_t\right)'$, where the variables are defined following AG: G is log real government (federal, state, and local) purchases (consumption and investment), T is log real government receipts of direct and indirect taxes net of transfers to businesses and individuals, and Y is log real gross domestic product (GDP) in chained 2000 dollars, and $\Delta G^F_{t|t-1}$ is the growth rate of government spending G at time t forecasted at time $t-1$. As described in Section 2, we include the anticipated growth rate of government spending in order to soak up any forecastable changes in government spending. We used quarterly forecasts data compiled by AG combining Greenbook and SPF forecasts over 1966-2008. We can thus estimate the model using quarterly data over 1966-2008.

To calculate the government spending multiplier, we follow the literature and consider two measures; the "max" multiplier $m_{\max} = \frac{\max\limits_{k=[0,20]}(Y_k)}{\max\limits_{k=[0,20]}(G_k)}$, and the "sum" multiplier $m_{\text{sum}} = \sum\limits_{k=0}^{20} Y_k / \sum\limits_{k=0}^{20} G_k$.

4.1.1 Results with one Gaussian function

We first present the results obtained with only one Gaussian function, i.e., $N = 1$. Figure 2 plots the estimated impulse responses following a shock to G_t with the responses scaled so that the peak response equals one. The solid black line depicts the symmetric IRF as recovered from a VAR. A shock to government spending generates a persistent but transitory increase in G_t. The thick blue line depicts the response of G_t to a positive spending shock (an expansionary shock), while the dashed red line depicts the shock to a negative spending shock (a contractionary shock). The response to a positive shock is stronger and much more persistent than the response to a negative shock (with the symmetric IRF lying in-between the two asymmetric IRFs).[17]

We now turn to the responses of output, which were scaled by the average Y/G ratio so that the peak of the impulse response can be read as the "max" government spending multiplier. We can see that the response of output is strong following a contractionary G shock but is not significant following an expansionary G shock. As shown in Table 1, the "max" multiplier is about 1.5 for a contractionary shock but is only 0.5 (and non-significantly different from zero) for an expansionary shock. The difference between positive and negative G shocks is even more striking when we take into account the fact that the impulse response of government spending is

[17] For ease of comparison between positive and negative shocks, the response to the negative shock has been multiplied by -1.

much less persistent following a contractionary spending shock than following an expansionary shock. The "sum" multiplier, depicted in Table 1, shows that the "sum" government multiplier to a spending contraction is 1.67, significantly higher than the linear multiplier (at 1.01) and the expansionary multiplier (at 0.54).

Figure 2 shows that the response of taxes is not behind the asymmetric size of the multiplier. While the response of taxes is not different from zero following an expansionary shock, taxes declined markedly following a contractionary shock (recall that the impulse responses to contractionary shocks are multiplied by -1 in Figure 2). Thus, the tax response should make the adverse effect of contractionary fiscal policy on output smaller, not larger.

Finally, we explore the systematic response of monetary policy and argue that monetary policy cannot explain the asymmetric size of the multiplier. To do so, we extend our 4-variable model to 5 variables by adding the federal funds rate that we order last. Figure 3 plots the impulse responses and shows that the response of monetary policy is roughly symmetric, indicating that monetary policy is unlikely to be behind the spending multiplier asymmetry.

4.1.2 Results with two Gaussian functions

Figure 4 present the results with two Gaussian basis functions. The results are little changed, and the same conclusion holds: the spending multiplier is strongly asymmetric.[18]

4.2 Shock identification from the narrative approach

We now turn to Ramey's identification scheme, and we explore the asymmetry of the multiplier following unexpected changes in anticipated future expenditures.

To identify unexpected changes in anticipated future expenditures, Ramey (2011) measures expectations by using news sources, primarily articles in Business Week, to estimate the present discounted value of expected changes in defense spending during quarters of each year. We estimate impulse response functions from a scalar-GMA model with 3 Gaussian basis functions using quarterly data over 1939-2008.[19]

Figure 5 plots the impulse responses of government spending and output to news shocks, and Table 1 reports the corresponding sizes of the multiplier.

Consistent with Ramey (2011), the multiplier estimated from a linear model is lower than 1, with a point estimate of .75 for the "max multiplier" and .53 for the "integral multiplier".

However, the size of the multiplier differs markedly between expansionary and contractionary shocks.[20] The impulse responses to a positive (expansionary) news shock are similar

[18] In fact, a likelihood ratio test cannot reject the null in favor of the model with 2 Gaussian basis functions.

[19] The number of Gaussian basis functions was determined by BIC-AIC criteria.

[20] Note also that a negative news shock generates a much larger response of both government spending and

to those recovered by the linear model, and the multiplier is significantly lower than 1. In contrast, the multiplier for a contractionary shock is significantly larger than the multiplier for an expansionary shock and is above 1 for the "sum multiplier" (and close to 1 for the "max multiplier").

Thus, these results confirm our previous results based on a recursive identification scheme: the government multiplier is markedly different for increases and decreases in public spending, and the multiplier is below one when government spending increases, but is above one when government spending decreases.[21]

4.3 Contrasting Auerbach-Gorodnichenko and Ramey

The existence of asymmetry in the size of the multiplier helps reconcile seemingly contradictory findings in the literature on government multipliers: While estimates from shocks to defense spending (Ramey, 2011) were significantly less than 1, estimates from shocks recovered with a recursive identification scheme (AG) were close to and slightly above 1.

We argue that these different results stem from variations in the sample of fiscal shocks used for identification in each method, and more specifically from variations in the relative importance of expansionary and contractionary shocks.

Figure 6 plots the distribution of Ramey news shocks along with the distribution of AG shocks. Unlike with AG shocks whose distribution is roughly evenly distributed between positive/negative shocks and large/small shocks, a few very large positive shocks dominate the sample of Ramey news shocks.[22]

The results from the narrative approach on news to defense spending (Ramey, 2011, Barro and Redlick 2011) are driven primarily by positive shocks –unexpected *increases* in government spending–, because the Ramey news shock series contain much fewer negative shocks than positive shocks (Figure 6). Indeed, the (linear) multiplier estimated by Ramey is close to the multiplier of a *positive* shock, which is consistent with the fact that sample of news shocks is dominated by positive shocks. And the estimated multiplier is small, because the multiplier associated with positive shocks is small.

In contrast, the spending shocks identified by AG are much more evenly distributed between positive and negative values, and the (linear) VAR-based impulse responses were in between the responses to negative and positive shocks. As a result, AG find a larger multiplier, lying

output.

[21] To use a sample period similar to the AG exercise, we restricted the sample size to the post-Korean war period. While the error-bands become very large in that case, the point estimates consistently show that the multiplier is substantially larger for contractionary shocks than expansionary shocks.

[22] Indeed, it has been argued (e.g., AG) that the results of Ramey (2011) were driven by the large shocks of WWII and the Korean war.

in between our estimates of the expansionary multiplier and the contractionary multiplier.

5 Conclusion

This paper estimates the asymmetric effect of fiscal policy by using Gaussian mixtures to approximate the the impulse response functions to fiscal policy shocks. We find that the government spending multiplier is substantially below 1 for fiscal expansions, but that the multiplier is substantially above 1 for fiscal consolidations.

These results have two important policy implications. First, they strongly weaken the case for fiscal packages to stimulate the economy. Second, they caution that austerity measures (e.g., as pursued in continental Europe) may have a much higher output cost than suggested by linear estimates.

Treating expansionary and contractionary fiscal shocks separately also allows us to reconcile seemingly contradictory findings in the literature on government multipliers: While estimates from shocks to defense spending (Ramey, 2011, Barro and Redlick 2011) are generally significantly less than 1, estimates from shocks recovered with a recursive identification scheme (Blanchard and Perotti, 2002, Auerbach and Gorodnichenko, 2012) are close to and slightly above 1. We argue that these different results stem from variations in the sample of fiscal shocks used for identification in each method and more specifically from variations in the relative importance of expansionary and contractionary shocks.

References

[1] Alesina, A., and S. Ardagna. "Large Changes in Fiscal Policy: Taxes versus Spending," In Tax Policy and the Economy, edited by Jeffrey R. Brown, vol. 24. Chicago: University of Chicago Press, pp. 35–68, 2010

[2] Almon, S. "The distributed lag between capital appropriations and expenditures," Econometrica, 33, January, 178-196, 1965

[3] Alspach D. and H. Sorenson H. "Recursive Bayesian Estimation Using Gaussian Sums," Automatica, Vol 7, pp 465-479, 1971

[4] Alspach D. and H. Sorenson H. "Nonlinear Bayesian Estimation Using Gaussian Sum Approximations," IEEE Transactions on Automatic Control, Vol 17-4, August 1972

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

[6] Auerbach A and Y Gorodnichenko. "Fiscal Multipliers in Recession and Expansion." In Fiscal Policy After the Financial Crisis, pp. 63–98, University of Chicago Press, 2013

[7] Barnichon R. and C. Matthes. "Measuring the Non-Linear Effects of Monetary Policy," Working Paper, 2014

[8] Barro R. and C. Redlick. "Macroeconomic Effects From Government Purchases and Taxes," Quarterly Journal of Economics (2011) 126 (1): 51-102, 2011

[9] Blanchard, O., and R. Perotti. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." Quarterly Journal of Economics 117 (4): 1329–68, 2002

[10] Buhmann, M., *Radial Basis Functions: Theory and Implementations*, Cambridge University Press, 2003

[11] Guajardo, Jaime, Daniel Leigh, and Andrea Pescatori. "Expansionary Austerity: New International Evidence" IMFWorking Paper 11/158, 2014

[12] Haario, H., E. Saksman, and J. Tamminen, "An adaptive Metropolis algorithm," Bernoulli 7, no. 2, 223–242, 2001

[13] Hall, R. "By HowMuch Does GDP Rise If the Government Buys More Output?" Brookings Papers on Economic Activity, 2, 183–231, 2009

[14] Jorda O., "Estimation and Inference of Impulse Responses by Local Projections," American Economic Review, pages 161-182, March 2005

[15] Jorda O. and A. Taylor. "The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy" Working Paper 2013-25 September 2013

[16] McLachlan, G. and D. Peel. *Finite Mixture Models*. Wiley Series in Probability and Statistics, 2000

[17] Mertens K. and M. Ravn. "Measuring the Impact of Fiscal Policy in the Face of Anticipation: A Structural VAR Approach," Economic Journal, Volume 120, Issue 544, pages 393–413, May 2010.

[18] Owyang, M., V. Ramey and S. Zubairy. "Are Government Spending Multipliers Greater During Periods of Slack? Evidence from 20th Century Historical Data" American Economic Review Papers and Proceedings, May 2013

[19] Parker, J. "On Measuring the Effects of Fiscal Policy in Recessions," Journal of Economic Literature, American Economic Association, vol. 49(3), pages 703-18, September, 2011

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

[21] Ramey V. "Comment on "Roads to Prosperity or Bridges to Nowhere? Theory and Evidence on the Impact of Public Infrastructure Investment", NBER Macroeconomics Annual, Volume 27, pages 147-153, 2012

[22] Ramey V. and S. Zubairy. "Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data ," Working Paper, 2014

[23] Robert, C. and G. Casella *Monte Carlo Statistical Methods* Springer, 2004

[24] Walsh C. *Monetary Theory and Policy*, 3nd. ed., The MIT Press, 2010

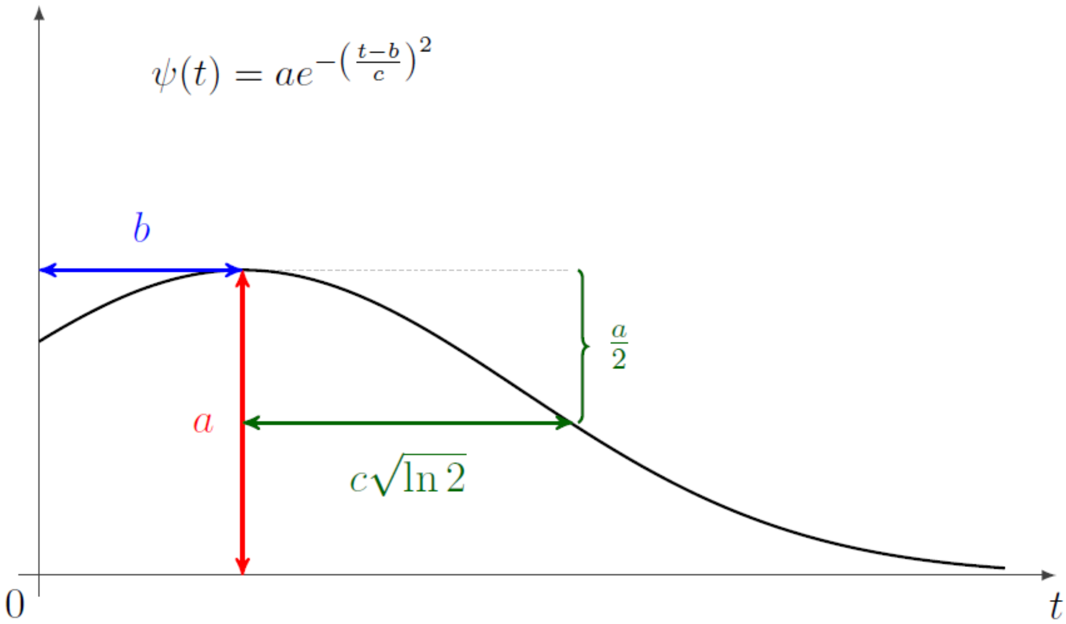

Figure 1: Interpreting an impulse response function with a Gaussian kernel.

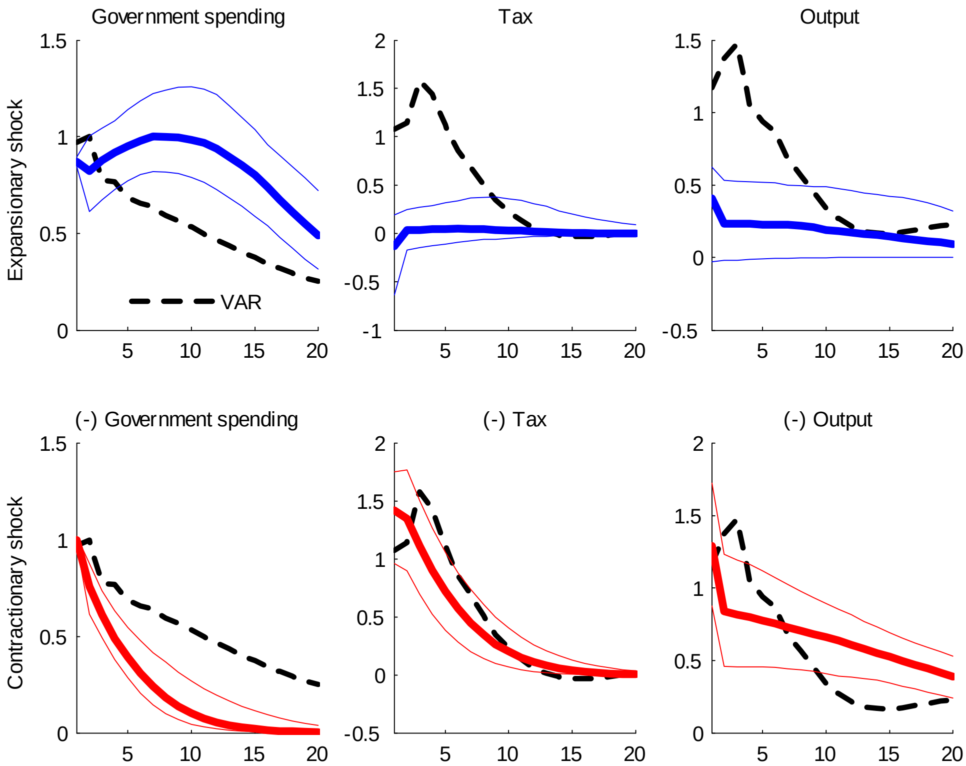

Figure 2: Impulse response functions (in ppt) of government spending, government revenue ("Tax") and output to a one standard-deviation government spending shock. Estimation from a standard VAR (dashed-line) or from a parametric (one Gaussian) specification of the impulse-responses (plain line). The response of output is scaled by the average Y/G ratio (equal to 5 over the sample period). The thin lines cover 90% of the posterior probability. Estimation using quarterly data covering 1966-2008.

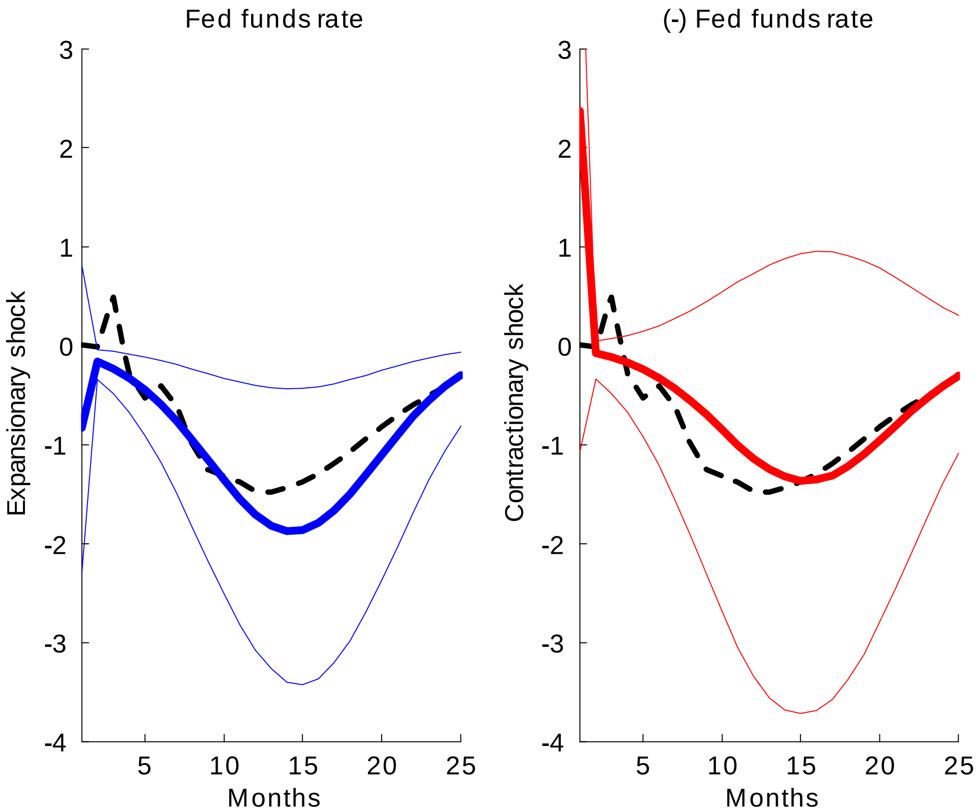

Figure 3: Impulse response functions (in basis points) of the Fed funds rate following an expansionary government spending shock (left panel) and a contractionary shock (right panel). Estimation from a standard VAR (dashed-line) or from a parametric (one Gaussian) specification of the impulse-responses (plain line). The thin lines cover 90% of the posterior probability. Estimation using quarterly data covering 1966-2008.

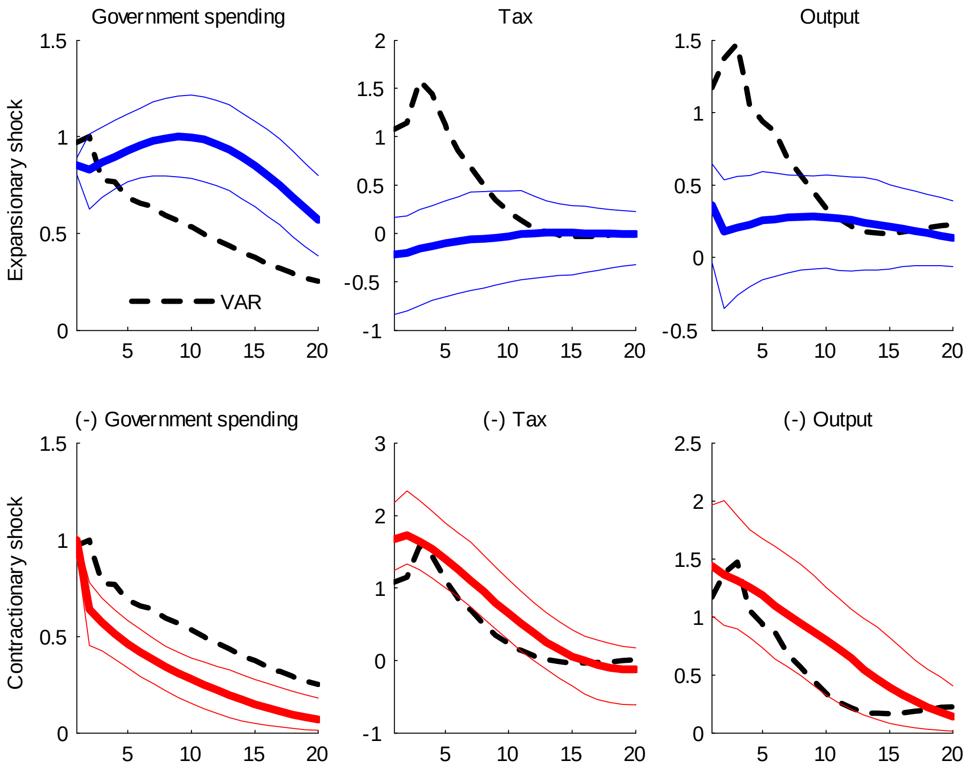

Figure 4: Impulse response functions (in ppt) of government spending, government revenue ("Tax") and output to a one standard-deviation government spending shock. Estimation from a standard VAR (dashed-line) or from a parametric (two Gaussians) specification of the impulse-responses (plain line). The response of output is scaled by the average Y/G ratio (equal to 5 over the sample period). The thin lines cover 90% of the posterior probability. Estimation using quarterly data covering 1966-2008.

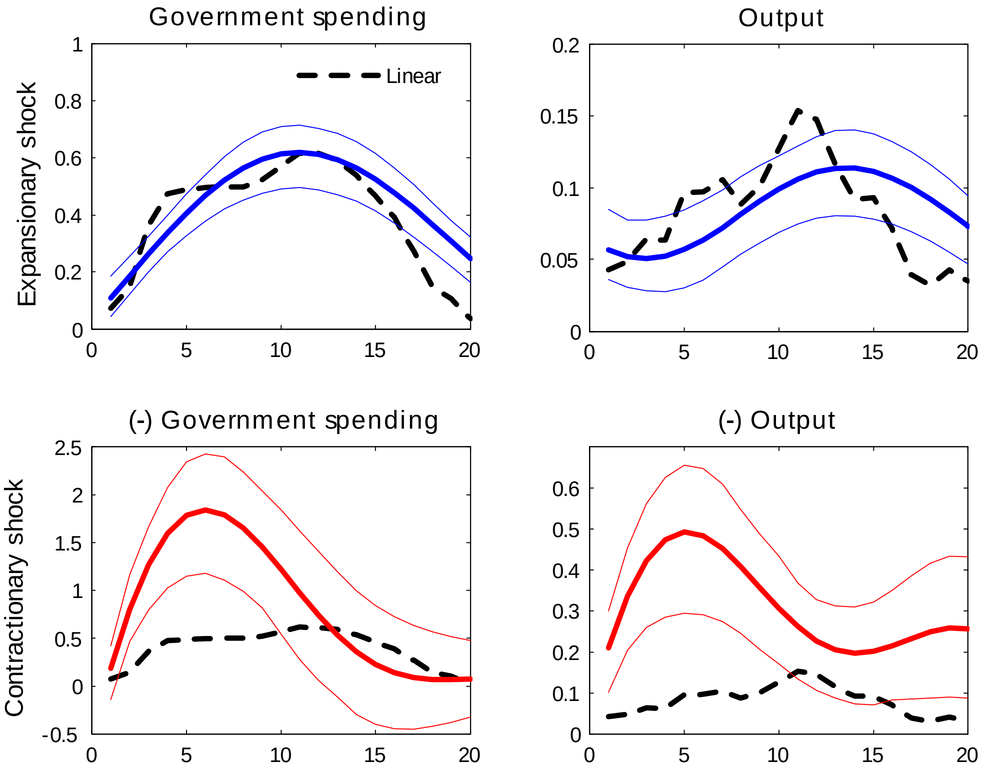

Figure 5: Impulse response functions (in ppt) of government spending and output to a one standard-deviation Ramey news shock. Estimation from a distributed-lags model (dashed-line) or from a parametric (three Gaussians) specification of the impulse-responses (plain line). The thin lines cover 90% of the posterior probability. Estimation using quarterly data covering 1939-2008.

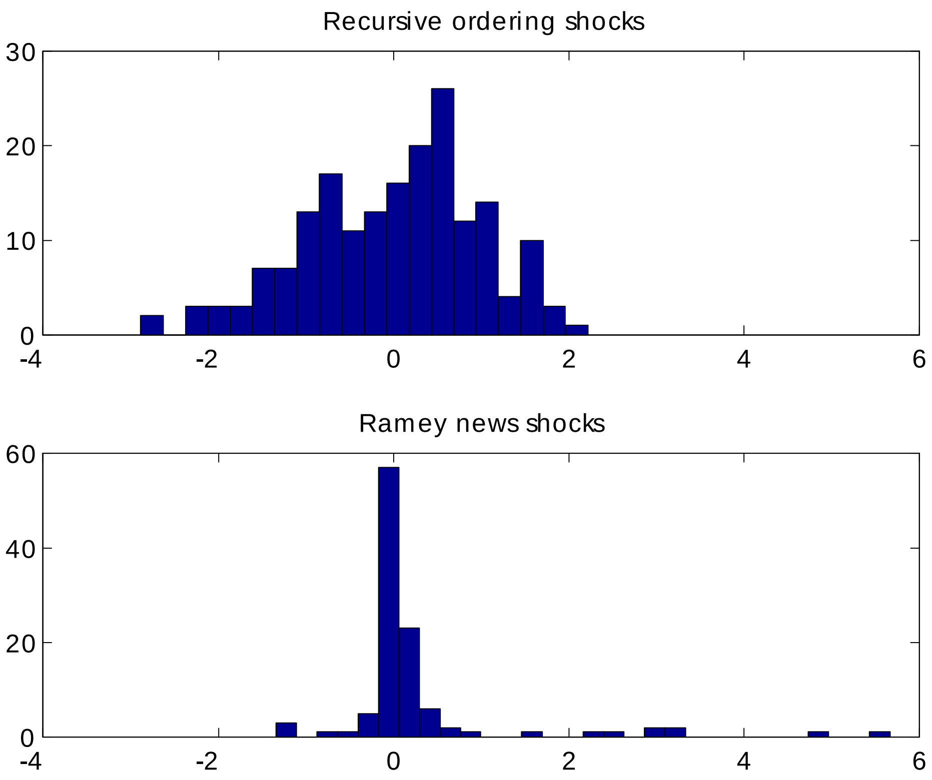

Figure 6: Distribution of government spending shocks (rescaled by their standard-deviation). The upper-panel plots the distribution of shocks recovered from a recursive ordering (1966-2008), the bottom-panel plots the distribution of Ramey news shocks (1939-2008).

Table 1: Estimates of government spending multipliers from Gaussian Mixture Approximation (GMA) models

	Max			Sum		
	Linear	**Expansionary shock**	**Contractionary shock**	**Linear**	**Expansionary shock**	**Contractionary shock**
GMA AG shocks 1966-2008	1.47	0.28 (.03--.60)	1.48 (.76--2.56)	1.01	0.54 (.10--1.06)	1.67 (.94--2.51)
GMA Ramey News shocks 1939-2008	0.75 (.52--.99)	0.29 (.10--.46)	0.89 (.66--1.21)	0.53 (.26--.79)	0.14 (-.08--.41)	1.34 (.98--1.97)

Note: Numbers in parenthesis cover 90% of the posterior probability. Multiplier calculated by taking an average Y/G ratio of 5. AG shocks refer to shocks obtained as in Auerbach and Gorodnichenlo (2012) from a Blanchard-Perotti recursive identification scheme augmented with professional forecasts of government spending. Ramey news shocks are the unexpected changes in anticipated future expenditures constructed by Ramey (2011).