Online Appendix

Understanding the Size of the Government Spending Multiplier: It's in the Sign

Regis Barnichon, Davide Debortoli and Christian Matthes

This appendix provides (1) additional empirical results supporting our findings from the main text, (2) independent evidence about multiplier asymmetry using the Blanchard and Leigh (2013) approach, (3) details about the estimation of the FAIR model, (4) a discussion on the use of auto-regressive models to capture asymmetric DGPs, (5) a description of the numerical solution method, and (6) a quantitative study of the effect of persistence in the public spending process on the size of the multiplier.

1 Additional empirical results

In this section, we report a number of additional checks: we report (i) the time series for government spending used in the two identification schemes, (ii) the impulse responses and the multiplier estimated using either a (symmetric) FAIR model or a VAR model, (iii) the response of taxes to government spending shocks, (iv) the multiplier for non-defense spending in the recursive identification case, (v) results from a more general model of state dependence (where the persistence of the IRFs also depend on the state of the cycle), and (vi) impulse responses depicting the state dependent results for the 1947-2014 sample.

1.1 Time series of government spending

Figure A1 plots the time series of government spending as a ratio to potential output estimated as in Ramey and Zubairy (2018).

1.2 VAR vs (linear) FAIR

In section 5.2 of the main text, we aimed to assess the magnitude of the bias introduced by the FAIR approximation by contrasting our FAIR estimates from those of Local Projections. In this subsection, we provide an additional robustness check to assess the size of the bias: we contrast the estimates from a (symmetric) FAIR model with those of a standard VAR. As shown in Figure A2, the impulse response point estimates are close and lie comfortably within the uncertainty bands of the VAR estimates.[1]

1.3 The response of taxes

In this section, we report the response of taxes to our government spending shocks. In doing so, we will explore whether asymmetry in the method of financing —taxes vs. deficit— for positive and negative spending shocks could be behind the asymmetry in the multiplier.

Figures A3 and A4 show the responses of government spending, the average tax rate (ATR), the average marginal tax rate (AMTR) and the fiscal deficit in response to a government spending shock identified recursively or narratively. The ATR is computed as the ratio of federal receipts to lagged nominal GDP (as in Ramey and Zubairy, 2018), the AMTR is the income-weighted average of the individual marginal tax rates, taken from Barro and Redlick (2011) and available over 1912-2014, and the deficit is total federal expenditures minus total federal receipts as a ratio to nominal GDP (as in Ramey and Zubairy, 2018).

With the progressivity of the tax system, one would expect the ATR to co-move positively with aggregate income, as the income distribution moves through the tax brackets. Consistent with this mechanism, we find that the ATR co-moves positively with output. The only case where there is no significant response of the ATR is when the output response is small and non-significant (in response to an expansionary shock identified recursively (top row, Figure 2 in the main text)).

[1] Note that the VAR estimates should not be taken as the truth. The VAR is also a biased estimator because the VAR is a finite-order VAR and thus also introduces a bias (see e.g., Barnichon and Matthes, 2018). What is reassuring is that the FAIR and VAR estimates lie close to each other.

Although the ATR mixes personal and corporate taxes, it is instructive to study the response of AMTR which does not suffer from a systematic association with income (unlike the ATR). We find no significant response of AMTR to government spending shocks, regardless of the identification scheme. This shows that we are not systematically confounding government spending shocks with shocks to the personal marginal income tax rates.

In line with the absence of changes in the AMTR, we find that these spending changes are deficit financed to a large extent, although the precise magnitude of the deficit response varies across identification schemes. Note that we do not find any systematic relationship between the sign of the shock and the method of financing. For instance, under the narrative identification contractionary shocks are associated with larger changes in deficits than expansionary shocks, but the opposite happens under the recursive identification scheme. This indicates that asymmetry in the method of financing is not be behind the asymmetry in the multiplier.

1.4 Defense vs. Non-defense spending

Perotti (2014) argues that defense spending and non-defense spending have different multipliers. To verify that the defense/non-defense composition of the shocks is not behind our results, we re-estimated our recursively-identified FAIR model using non-defense government spending as our measure of government spending.

Figure A5 plots the same results as Figure 2 in the main text but using Non-Defense spending as the g variable. The results are very similar to the baseline results, if anything pointing to a larger contractionary multiplier.

1.5 A more general model of state dependence

In the main text we model state dependence by allowing the state of the cycle to affect the scale of the impulse response but not its shape. A natural question is whether this is general enough. In particular, a worry could be that the persistence of the effect of a government spending shock depends on the state of the economy as well. To make progress on this question, recall that

c^+ is the parameter governing persistence when the shock is positive (and similarly with c^-). To assess whether this persistence really depends on the state of the business cycle, we replace c^+ and c^- with the following expressions:

$$c_t^+ = |c_{intercept}^+ + c_{state}^+ * z_t| \tag{1}$$

$$c_t^- = |c_{intercept}^- + c_{state}^- * z_t| \tag{2}$$

z_t is an observed business cycle indicator. We keep the same priors as in our benchmark (with the $c_{intercept}$ parameters inheriting the priors of the original $c^\pm$ parameters except that we now do not impose any positivity constraint), and the prior on $c_{state}^\pm$ is a normal distribution with mean 0 and standard deviation 2.

In Table A1, we show the estimated multipliers as a function of the business cycle (we use the same indicator as in our benchmark analysis). For all c_{state} parameters, 0 is within the 5th-95th percentile posterior band, and the estimated multipliers are very similar to our benchmark case, if anything the state dependence in the contractionary multiplier is slightly stronger.

1.6 State dependence estimated over 1947-2014

To complement Table 3 from the main text, Figures A6 and A7 show how the impulse responses of government spending and output vary with the state of the cycle.

2 Blanchard and Leigh (2013)

Blanchard and Leigh (BL, 2013) analyze the size of the multiplier from a very different angle. Using a panel of EU countries over 2009-2012, BL regress the forecast error of real GDP growth on forecasts of fiscal adjustments. Under rational expectations, and assuming that forecasters use the correct model of forecasting, the coefficient on the fiscal adjustment forecast should be zero. However, BL find that there is a significant relation between fiscal adjustment forecasts and subsequent growth forecast errors, which indicates that the size of the multiplier was under-estimated during the last recession. Moreover, the magnitude of the under-estimation is large: if forecasters had in mind a multiplier of about 0.5, BL's estimates imply that the multiplier was 1.6 during 2009-2012.

BL's approach is an interesting testing ground for our findings. Since we find that only the contractionary multiplier is above one, BL's results should be driven by fiscal consolidations alone, and not by fiscal expansions. As we show below, this is exactly what we find.

Specifically, BL run the regression

$$\text{Forecast error of } \Delta Y_{i,t+1|t} = \alpha + \beta \left(\text{Forecast of } \Delta F_{i,t+1|t}\right) + \varepsilon_{i,t+1} \tag{3}$$

on a cross-section of European countries where $\Delta Y_{i,t+1|t}$ denotes cumulative (year-over-year) growth of real GDP in economy i and the associated forecast error is $\Delta Y_{i,t+1|t} - \widehat{\Delta Y}_{i,t+1|t}$ with $\widehat{\Delta Y}_{i,t+1|t}$ the forecast made with information available at date t, and where $\Delta F_{i,t+1|t}$ denotes the change in the general government structural fiscal balance in percent of potential GDP.

Under the null hypothesis that fiscal multipliers used for forecasting were accurate, the coefficient β should be zero.[2] In contrast, a finding that β is negative indicates that forecasters tended to be optimistic regarding the level of growth associated with a fiscal consolidation, i.e., that they under-estimated the size of the multiplier. Using World Economic Outlook (WEO) forecast data, BL find that $\beta \approx -1.1$ for forecasts over 2009-2012, indicating that the multiplier was substantially under-estimated by forecasters, and implying that the multiplier was 1.6 (0.5+1.1) during the recession.[3]

To test our prediction that BL's findings is driven by fiscal consolidations, we re-estimate BL's baseline specification but allowing for different β coefficients depending on the sign of the

[2] In other words, information known when the forecasts were made should be uncorrelated with subsequent forecast errors.

[3] BL conduct a number of robustness checks to argue that their non-zero β is symptomatic of an under-estimated multiplier and is not due to other confounding factors. In particular, they verify that their results hold after controlling for other factors that could trigger both planned fiscal adjustments and lower than expected growth, or that the forecast error in fiscal adjustment was not correlated with the initial fiscal adjustment forecast (which would bias β).

fiscal adjustment –expansionary or contractionary–. Figure A8 shows the corresponding fitted lines. To avoid confusion, note that a fiscal consolidation in BL's framework shows up as in *increase* in the fiscal balance and thus shows up as positive entries in Figure A8. We can see that BL's results are indeed driven by fiscal consolidations. Table A2 presents the regression results. The β associated with fiscal expansions is not different from zero –suggesting an expansionary multiplier of about 0.5 during the recession–, but the β associated with fiscal consolidations is $\beta_G^- \approx -1.2$ –suggesting a contractionary multiplier of about 1.7–. These results are close to our estimates on the size of the multiplier during recessions.[4]

A caveat in our analysis so far is that fiscal adjustments include not only changes in government purchases but also changes in revenues. To better map BL's results with ours, we follow BL and treat separately changes in spending and changes in revenues by running the regression

$$\text{Forecast error of } \Delta Y_{i,t+1|t} = \alpha + \beta_G \left(\text{Forecast of } \Delta G_{i,t+1|t}\right) + \beta_T \left(\text{Forecast of } \Delta T_{i,t+1|t}\right) + \varepsilon_{i,t+1} \tag{4}$$

where $\Delta G_{i,t+1|t}$ denotes the WEO forecast of the change in structural spending in 2010-11 and $\Delta T_{i,t+1|t}$ denotes the WEO forecast of the change in structural revenue in 2010-11, both in percent of potential GDP.

Column (3) of Table A2 presents the results of regression (4) where we treat separately fiscal consolidations and fiscal expansions. In line with our findings, the only significant coefficient is $\beta_G^- \approx 1.6$,[5] corresponding to a contractionary spending multiplier of about 2 in recessions (again in line with our findings), whereas β_G^+ is not significantly different from zero, consistent with an expansionary spending multiplier of about 0.5 in recessions.

[4]Throughout this exercise, we follow BL in keeping the assumption that forecasters had in mind a multiplier of 0.5.

[5]The coefficient β_G^- is positive because a fiscal consolidation corresponds to a *decrease* in government spending. In contrast, in columns (1) and (2), β^- is negative because a fiscal consolidation corresponds to an *increase* in the fiscal balance.

3 SUR-FAIR models and External Instruments

The section describes a fast and efficient procedure we call SUR-FAIR to estimate impulse responses when the shocks have been previously identified through a narrative approach (possibly with measurement error), that is when external instruments are available.

Denoting Y_t a variable of interest and G_t government spending, the SUR-FAIR model writes

$$\begin{pmatrix} Y_t \\ G_t \end{pmatrix} = \sum_{k=0}^{K} \begin{pmatrix} \varphi_Y(k) \\ \varphi_G(k) \end{pmatrix} \xi_{t-k}^G + \begin{pmatrix} u_t^Y \\ u_t^G \end{pmatrix} \tag{5}$$

with φ_Y and φ_G given by a functional approximation, $\mathbf{u}_t = \begin{pmatrix} u_t^Y \\ u_t^G \end{pmatrix}$ the vector of residuals, and where ξ_t^G denotes a proxy (i.e., an instrument) for the government spending shock ε_t^G. The proxy can contain measurement error and is only correlated with the true shocks, that is we have

$$\xi_t^G = \alpha\varepsilon_t^G + \eta_t$$

with η_t i.i.d with variance σ_η^2. In the language of Instrument Variables, ξ_t^G is correlated with the shock of interest, but is uncorrelated with any other shocks.

To highlight the bias coming from measurement error only, we ignore the FAIR aspect of our method here and consider the simpler model where φ_Y and φ_G are left unrestricted (i.e., no functional parametrization). The model is then a simple Distributed Lags model with a SUR structure.

Since ξ_t^G is only correlated with the true shock ε_t^G, the maximum-likelihood estimates $\hat{\varphi}_Y$ and $\hat{\varphi}_G$ are biased estimates of the true impulse responses ψ_Y and ψ_G with

$$\begin{aligned} \hat{\varphi}_G &= \nu\psi_G \\ \hat{\varphi}_Y &= \nu\psi_Y \end{aligned}$$

with the bias given by

$$\nu = \frac{\alpha\sigma^2_{\varepsilon^G}}{\alpha^2\sigma^2_{\varepsilon^G} + \alpha\sigma^2_\eta}$$

where $\sigma^2_{\varepsilon^G}$ is the variance of the true government spending shocks ε^G_t.

Exactly like in an IV regression, an unbiased estimate $\hat{\psi}_Y$ of the impulse response function ψ_Y can be recovered by appropriately re-scaling $\hat{\varphi}_Y$ with $\hat{\varphi}_G$, i.e., from

$$\hat{\psi}_Y = \frac{\hat{\varphi}_Y}{\hat{\varphi}_G(k_0)}$$

with k_0 some arbitrary horizon. The re-scaling ensures that ν –the term capturing the measurement-error bias– drops out.

Estimation procedure

Next, we describe the estimation procedure of a SUR-FAIR model like (5). The computational advantage of this SUR-FAIR approach (compared to a VMA-FAIR model as in section 4) is that only the impulse responses of interest are parametrized and estimated, yielding a small parameter space and a very fast estimation procedure.

For ease of exposition, we focus on a univariate model first, since the SUR model is a simple extension of the univariate case. Recall that for a variable y_t we have a model of the form

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

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 ξ^G_{t-k} (in the non-linear case). Unlike the main text where we approximate each impulse response with one Gaussian function, in this appendix, we consider directly the more general case with N Gaussian functions.

Since $\{u_t\}$ is likely serially correlated by construction, in order to improve efficiency, we allow for serial correlation in u_t by positing that u_t follows an $AR(1)$ process. That is, we posit that $u_t = \rho u_{t-1} + \eta_t$ where η_t is Normally distributed $N(0, \sigma_\eta^2)$ with σ_η a parameter to be estimated. We set η_{-1} and η_0 to zero, and from (6), it is straightforward to build the likelihood given a series of previously identified shocks $\left\{\xi_t^G\right\}$. For prior elicitation, we use very loose priors with $\sigma_a = 10$, $\sigma_b = K$ and $\sigma_c = K$.

For a multi-variate model, the estimation proceeds along the same lines as above, except that we take into account that the one-step forecast error u_t is now a vector that follows a VAR(1) process instead of an AR(1) process.

Estimation routine and initial guess

As estimation routine, we use a Metropolis-within-Gibbs algorithm, as described in more details in section 4. Regarding the initial guess, an interesting advantage of a univariate FAIR is that it is possible to compute a good initial guess, even in non-linear models:

Obtaining a non-linear initial guess

To obtain a good (possibly non-linear) initial guess in SUR-FAIR models, we use the following two-step method:

1. Recover the $\{a_n\}$ factors given $\{b_n, c_n\}$

Assume 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, estimating the coefficients $\{a_n\}_{n=1}^N$ in (6), a non-linear problem, can be recast into a linear problem that can estimated by OLS. In other words, compared to a direct non-linear least square of (6) that treats all three sets of parameters a_n, b_n and c_n as free parameters, our two-step approach has the advantage of exploiting the efficiency of OLS to find $\{a_n\}$ given $\{b_n, c_n\}$.

To see that, consider first a linear model where $\psi(k)$ is independent of ξ^G_{t-k}. We then re-arrange (6) as follows:

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

Defining

$$X_{n,t} = \sum_{k=0}^{K}e^{-(\frac{k-b_n}{c_n})^2}\xi^G_{t-k},$$

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(1), we can estimate the $\{a_n\}$ with a NLS procedure.

The method described above is straightforward to apply to a case with asymmetry and state dependence. Consider for instance the case with asymmetry

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

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

$$\begin{cases} X^+_{n,t} = \sum\limits_{k=0}^{K} h_n(k)\xi^G_{t-k}1_{\xi\geq 0} \\ X^-_{n,t} = \sum\limits_{k=0}^{K} h_n(k)\xi^G_{t-k}1_{\xi< 0}\end{cases}$$

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

2. Choose $\{b_n, c_n\}$

To estimate $\{b_n, c_n\}_{n=1}^N$ (and therefore $\{a_n\}_{n=1}^N$ from the OLS regression), we minimize the sum of squared residuals of (7) using a simplex algorithm.

4 VMA-FAIR models

In this section, we describe the implementation and estimation of structural VMA-FAIR models, where government spending shocks are identified from a recursive ordering as in Auerbach and Gorodnichenko (2012). As in the main text, for $\boldsymbol{y}_t$ a vector of stationary macroeconomic variables, the VMA model writes

$$\boldsymbol{y}_t = \sum_{k=0}^{K} \boldsymbol{\Psi}_k(\boldsymbol{\varepsilon}_{t-k}, z_{t-k})\boldsymbol{\varepsilon}_{t-k} \tag{8}$$

with $\boldsymbol{\Psi}$ given by a functional approximation.

The approach is identical to Barnichon and Matthes (2017) in the case of recursively-identified monetary shocks, bar one non-trivial extension: We show how to identify (and estimate) non-linear FAIR models with asymmetric *and* state dependent effects, i.e., where we have $\boldsymbol{\Psi}_k = \boldsymbol{\Psi}_k(\boldsymbol{\varepsilon}_{t-k}, z_{t-k})$.

4.1 Likelihood function

We use the prediction error decomposition to break up the density $p(\boldsymbol{y}^T|\boldsymbol{\theta})$ as follows:

$$p(\boldsymbol{y}^T|\boldsymbol{\theta}) = \prod_{t=1}^{T} p(\boldsymbol{y}_t|\boldsymbol{\theta}, \boldsymbol{y}^{t-1}). \tag{9}$$

To calculate the one-step-ahead conditional likelihood function needed for the prediction error decomposition, we assume that all innovations $\{\boldsymbol{\varepsilon}_t\}$ are Gaussian with mean zero and variance one,[6] and we note that the density $p(\boldsymbol{y}_t|\boldsymbol{y}^{t-1}, \boldsymbol{\theta})$ can be re-written as $p(\boldsymbol{y}_t|\boldsymbol{\theta}, \boldsymbol{y}^{t-1}) =$

[6] The estimation could easily be generalized to allow for non-normal innovations such as t-distributed errors.

$p(\boldsymbol{\Psi}_0\boldsymbol{\varepsilon}_t|\boldsymbol{\theta}, \boldsymbol{y}^{t-1})$ since

$$\boldsymbol{y}_t = \boldsymbol{\Psi}_0\boldsymbol{\varepsilon}_t + \sum_{k=1}^{K}\boldsymbol{\Psi}_k\boldsymbol{\varepsilon}_{t-k}. \tag{10}$$

Since the contemporaneous impact matrix $\boldsymbol{\Psi}_0$ is a constant, $p(\boldsymbol{\Psi}_0\boldsymbol{\varepsilon}_t|\boldsymbol{\theta}, \boldsymbol{y}^{t-1})$ is a straightforward function of the density of $\boldsymbol{\varepsilon}_t$.

To recursively construct $\boldsymbol{\varepsilon}_t$ as a function of $\boldsymbol{\theta}$ and $\boldsymbol{y}^t$, we need to uniquely pin down the value of the components of $\boldsymbol{\varepsilon}_t$ from (10), that is we need that $\boldsymbol{\Psi}_0$ is invertible. We impose this restriction by only keeping parameter draws for which $\boldsymbol{\Psi}_0$ is invertible.[7] It is also at this stage that we impose the identifying restriction that $\boldsymbol{\Psi}_0$ has its first two rows filled with 0 except for the diagonal coefficients. Finally, to initialize the recursion, we set the first K innovations $\{\boldsymbol{\varepsilon}_j\}_{j=-K}^{0}$ to zero.

In the non-linear case where we have $\boldsymbol{\Psi}_k = \boldsymbol{\Psi}_k(\boldsymbol{\varepsilon}_{t-k}, z_{t-k})$, we proceed similarly. However, a complication arises when one allows $\boldsymbol{\Psi}_0$ to depend on the sign of the shock *while also* imposing identifying restrictions on $\boldsymbol{\Psi}_0$. The complication arises, because with asymmetry the system of equations implied by (10):

$$\boldsymbol{\Psi}_0(\boldsymbol{\varepsilon}_{t-k}, z_{t-k})\boldsymbol{\varepsilon}_t = \mathbf{u}_t \tag{11}$$

where $\mathbf{u}_t = \boldsymbol{y}_t - \sum_{k=1}^{K}\boldsymbol{\Psi}_k\boldsymbol{\varepsilon}_{t-k}$ need not have a unique solution vector $\boldsymbol{\varepsilon}_t$, because $\boldsymbol{\Psi}_0(\boldsymbol{\varepsilon}_t)$, the impact matrix, depends on the sign of the shocks, i.e., on the vector $\boldsymbol{\varepsilon}_t$. In section 2.4, we show that this is not a problem (so that (11) has a unique solution vector $\boldsymbol{\varepsilon}_t$) in a recursive identification scheme like the one considered in this paper.

Finally, when constructing the likelihood, to write down the one-step ahead forecast density $p(\boldsymbol{y}_t|\boldsymbol{\theta}, \boldsymbol{y}^{t-1})$ as a function of past observations and model parameters, we use the standard result (see e.g., Casella-Berger, 2002) that for $\boldsymbol{\Psi}_0$ a function of $\boldsymbol{\varepsilon}_t$ and z_t, we have

$$p(\boldsymbol{\Psi}_0(\boldsymbol{\varepsilon}_t, z_t)\boldsymbol{\varepsilon}_t|\boldsymbol{\theta}, \boldsymbol{y}^{t-1}) = J_t p(\boldsymbol{\varepsilon}_t)$$

[7] Parameter restrictions (such as invertibility) are implemented by assigning a minus infinity value to the likelihood whenever the restrictions are not met.

where J_t is the Jacobian of the (one-to-one) mapping from $\boldsymbol{\varepsilon}_t$ to $\boldsymbol{\Psi}_0(\boldsymbol{\varepsilon}_t, z_t)\boldsymbol{\varepsilon}_t$ and where $p(\boldsymbol{\varepsilon}_t)$ is the density of $\boldsymbol{\varepsilon}_t$.[8,9]

4.2 FAIR estimation algorithm

This section describes our FAIR estimation algorithm in more detail. We are interested in estimating the parameter vector θ by combining the likelihood function $p(y^T|\theta)$ with the prior distribution $p(\theta)$. We want to generate N from the posterior by using a multiple-block Metropolis-Hastings algorithm (Robert & Casella 2004) with the blocks given by the different groups of parameters in our model (there is respectively one block for the a parameters, one block for the b parameters, one block for the c parameters and one block for the constant and other parameters). We use N^{tune} draws to tune the proposal distributions, which we update every n^{tune} draws during the tuning process. We split the parameter vector into J non-overlapping blocks $\theta_1, ..., \theta_J$. We denote θ_{-j} the parameters in all blocks but block j.

- estimate a VAR on y^T and compute the implied structural MA representation (imposing a identification scheme that is consistent with the scheme used in the FAIR model). Compute the parameter value θ^{VAR} that minimizes the quadratic distance between the VAR-implied IRFs and the FAIR IRFs.

- starting from θ^{VAR}, use an optimizer to maximize the posterior kernel $p(y^T|\theta)p(\theta)$.[10] Denote the resulting parameter estimate by θ^{start}

- for $j = 1, ..., J$, compute the inverse of the Hessian of the posterior kernel Σ_j at θ_j^{start} (holding all other blocks fixed at θ_{-j}^{start}) and use this as the first guess for the variance of the proposal density in block j

[8]Recall that we assume that the indicator variable z_t is a function of lagged values of $\boldsymbol{y}_t$ (so that z_t is known conditional on $\boldsymbol{y}^{t-1}$) or that z_t is a function of variables exogenous to $\boldsymbol{y}_t$ (and thus taken as given and known).

[9]In our case with asymmetry, this Jacobian is simple to calculate, but the mapping is not differentiable at $\varepsilon = 0$. Since we will never exactly observe $\varepsilon = 0$ in a finite sample, we can implicitly assume that in a small neighborhood around 0, we replace the original mapping with a smooth function.

[10]We initialize the parameters capturing asymmetry and state dependence at zero (i.e., no non-linearity). This approach is consistent with the starting point (the null) of this paper: structural shocks have linear effects on the economy, and we are testing this null against the alternative that shocks have some non-linear effects.

- for $n = 1$ to $\frac{N^{tune}}{n^{tune}}$
  - for $j = 1, ..., J$, compute n^{tune} draws for block j using the Metropolis-Hastings, holding all other parameters fixed at the latest draws for the respective blocks
  - if the acceptance probability is smaller than some threshold (say 0.15), multiply the variance of the proposal density by a positive constant smaller than 1
  - if the acceptance probability is larger than some threshold (say 0.5), multiply the variance of the proposal density by a positive constant larger than 1
- for $m = 1$ to N
  - for $j = 1, ..., J$ generate a draw of θ_j (conditioning on θ_{-j}) using the Metropolis-Hastings algorithm

4.3 Prior elicitation

We use (loose) Normal priors centered around the impulse response functions obtained from the benchmark (linear) VAR. Specifically, we put priors on the a, b and c coefficients that are centered on the values for a, b and c obtained by matching the impulse responses obtained from the VAR, as described in the previous paragraph. Specifically, denote $a^0_{ij,n}$, $b^0_{ij,n}$ and $c^0_{ij,n}$, $n \in \{1, N\}$ the values implied by fitting a FAIR model to the VAR-based impulse response of variable i to shock j. The priors for $a_{ij,n}$, $b_{ij,n}$ and $c_{ij,n}$ are centered on $a^0_{ij,n}$, $b^0_{ij,n}$ and $c^0_{ij,n}$, and the standard-deviations are set as follows $\sigma_{ij,a} = 10$, $\sigma_{ij,b} = K$ and $\sigma_{ij,c} = K$ (K is the maximum horizon of the impulse response function). While there is clearly some arbitrariness in choosing the tightness of our priors, it is important to note that they are very loose.

4.4 Identifying restrictions in non-linear VMA models

We now detail how to impose the recursive identifying restriction used in the paper, and we show that the structural shocks can be identified even with asymmetric and/or state dependent effects of shocks, i.e., when $\boldsymbol{y_t} = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k(\boldsymbol{\varepsilon_{t-k}}, \boldsymbol{z_{t-k}})\boldsymbol{\varepsilon_{t-k}}$.

As described above, to recursively construct the likelihood at time t, one must ensure that the shock vector $\boldsymbol{\varepsilon_t}$ is uniquely determined given a set of model parameters and the history of variables up to time t. Specifically, the system of equations

$$\boldsymbol{\Psi}_0(\boldsymbol{\varepsilon_t}, \boldsymbol{z_t})\boldsymbol{\varepsilon_t} = \mathbf{u}_t \tag{12}$$

need to have a unique solution vector $\boldsymbol{\varepsilon_t}$ given $\mathbf{u}_t = \boldsymbol{y_t} - \sum_{k=0}^{K} \boldsymbol{\Psi}_k(\boldsymbol{\varepsilon_{t-k}}, \boldsymbol{z_{t-k}})\boldsymbol{\varepsilon_{t-1-k}}$. That is, we must ensure that there is a one-to-one mapping from $\boldsymbol{\varepsilon_t}$ to $\boldsymbol{\Psi}_0(\boldsymbol{\varepsilon_t}, \boldsymbol{z_t})\boldsymbol{\varepsilon_t}$. In the linear case, this means that we must ensure $\boldsymbol{\Psi}_0$ is invertible. In the non-linear case, ensuring that the shock vector $\boldsymbol{\varepsilon_t}$ is uniquely determined becomes more complicated, *when* we allow $\boldsymbol{\Psi_0}$ to depend on the sign of the shock or on some state variable.[11]

Consider first the consequences of allowing for state dependence, i.e., when $\boldsymbol{\Psi}_k$ depends on the value of the indicator vector $\boldsymbol{z_t}$, so that the likelihood also depends on the value of the indicator vector $\boldsymbol{z_t}$. Technically, constructing the likelihood of this specification is a straightforward extension of the linear case, when $\boldsymbol{z_t}$ is a function of lagged values of $\boldsymbol{y}_t$. To see that, note that we use the prediction-error decomposition to construct the likelihood function. We build a sequence of densities for $\boldsymbol{y_t}$ that conditions on past values of $\boldsymbol{y_t}$. Thus, conditional on past values of $\boldsymbol{y}_t$, $\boldsymbol{z_t}$ is known, and as long as $\boldsymbol{\Psi}_0(\boldsymbol{z_t})$ is invertible, there is (one-to-one) mapping from $\boldsymbol{\varepsilon_t}$ to $\boldsymbol{\Psi}_0\boldsymbol{\varepsilon_t}$, and the likelihood can be recursively constructed.[12]

Consider now the consequences of allowing for asymmetry, i.e., when $\boldsymbol{\Psi}_k$ depends on the sign of $\boldsymbol{\varepsilon_t}$. A complication arises when one allows $\boldsymbol{\Psi_0}$ to depend on the sign of the shock

[11]Note that if the impact matrix $\boldsymbol{\Psi}_0$ is a constant and does not depend on $\boldsymbol{\varepsilon_t}$ or $\boldsymbol{z_t}$ (so that $\boldsymbol{\Psi}_k$ depends on $\boldsymbol{\varepsilon_t}$ or $\boldsymbol{z_t}$ only for $k > 0$), then one can construct the likelihood just as in the linear case, because as long as $\boldsymbol{\Psi}_0$ is invertible, there is (one-to-one) mapping from $\boldsymbol{\varepsilon_t}$ to $\boldsymbol{\Psi}_0\boldsymbol{\varepsilon_t}$, and $\boldsymbol{\varepsilon_t}$ is uniquely defined from $\mathbf{u}_t$.

[12]If we wanted to use an indicator function that was not a function of the history of endogenous variables $\boldsymbol{y}^{t-1}$, this would also be possible by using a quasi-likelihood approach. That is, we would build a likelihood function that not only conditions on the parameters, but also the sequence of indicators $\boldsymbol{z_t}$. This would in general not be efficient because the joint density of $\boldsymbol{z_t}$ and $\boldsymbol{y}_t$ could carry more information about the parameters in our model than the conditional density we advocate using. As long as $\boldsymbol{z_t}$ is highly correlated with elements of (functions of) $\boldsymbol{y}_t$, this loss in efficiency will likely be small.

while also imposing identifying restrictions on $\mathbf{\Psi}_0$. The complication arises, because with asymmetry, the system of equations $\mathbf{\Psi}_0(\boldsymbol{\varepsilon_t})\boldsymbol{\varepsilon_t} = \mathbf{u}_t$ need not have a unique solution vector $\boldsymbol{\varepsilon_t}$, because $\mathbf{\Psi}_0(\boldsymbol{\varepsilon_t})$, the impact matrix, depends on the sign of the shocks, i.e., on the vector $\boldsymbol{\varepsilon_t}$.

In this appendix, we show how to address the issue when we allow the identified shocks to have asymmetric and state dependent effects on the impulse response functions.

4.4.1 Recursive identification scheme

It will be convenient to adopt the following conventions for notation:

- Denote $y_{\ell,t}$ the ℓth variable of vector $\boldsymbol{y_t}$ and denote $\boldsymbol{y_t}^{<\ell} = (y_{1,t}, ..., y_{\ell-1,t})'$ the vector of variables ordered before variable $y_{\ell,t}$ in $\boldsymbol{y_t}$. Similarly, we can define $\boldsymbol{y_t}^{\leq\ell}$ or $\boldsymbol{y_t}^{>\ell}$.

- For a matrix $\mathbf{\Gamma}$ of size $L \times L$ and $(i,j) \in \{1,...,L\}^2$, denote $\mathbf{\Gamma}^{<i,<j}$ the $(i-1)\times(j-1)$ submatrix of Γ made of the first $(i-1)$ rows and $(j-1)$ columns. Similarly, we denote $\mathbf{\Gamma}^{>i,>j}$ the $(L-i)\times(L-j)$ submatrix of $\mathbf{\Gamma}$ made of the last $(L-i)$ rows and $(L-j)$ columns. In the same spirit, we denote $\mathbf{\Gamma}^{i,<j}$ the submatrix of $\mathbf{\Gamma}$ made of the ith row and the first $(j-1)$ columns. $\mathbf{\Gamma}^{i,<j}$ is in fact a row vector. A combination of these notations allows us to denote any submatrix of $\mathbf{\Gamma}$. Finally, denote $\mathbf{\Gamma}_{ij}$ the ith row jth column element of $\mathbf{\Gamma}$.

With these notations, we can now state the recursive identifying assumption

Assumption 1 (Partial recursive identification)**.** *The contemporaneous impact matrix* $\mathbf{\Psi_0}$ *of dimension* $L \times L$ *is of the form*

$$\mathbf{\Psi_0} = \begin{bmatrix} \underset{(\ell-1)\times(\ell-1)}{\mathbf{\Psi_0}^{<\ell,<\ell}} & \underset{(\ell-1)\times 1}{\mathbf{0}^{<\ell,\ell}} & \underset{(\ell-1)\times(L-\ell)}{\mathbf{0}^{<\ell,>\ell}} \\ \underset{1\times(\ell-1)}{\mathbf{\Psi_0}^{\ell,<\ell}} & \underset{1\times 1}{\mathbf{\Psi}_{0,\ell\ell}} & \underset{1\times(L-\ell)}{\mathbf{0}^{\ell,>\ell}} \\ \underset{(L-\ell)\times(\ell-1)}{\mathbf{\Psi_0}^{>\ell,<\ell}} & \underset{(L-\ell)\times 1}{\mathbf{\Psi_0}^{>\ell,\ell}} & \underset{(L-\ell)\times(L-\ell)}{\mathbf{\Psi_0}^{>\ell,>\ell}} \end{bmatrix}.$$

with $\ell \in \{1,..,L\}$, $\boldsymbol{\Psi}_{\mathbf{0}}^{<\ell,<\ell}$ *and* $\boldsymbol{\Psi}_{\mathbf{0}}^{>\ell,>\ell}$ *matrices of full rank and* $\mathbf{0}$ *denoting a conformable matrix of zeros.*

Assumption 1 states that the shock of interest $\varepsilon_{\ell,t}$, ordered in ℓth position in $\boldsymbol{\varepsilon_t}$, affects the variables ordered from 1 to $\ell - 1$ with a one period lag, and that the first ℓ variables in $\boldsymbol{y_t}$ do not react contemporaneously to shocks ordered after $\varepsilon_{\ell,t}$ in $\boldsymbol{\varepsilon_t}$.

We first consider a model with only asymmetry and then a model with asymmetry and state dependence.

4.4.2 Asymmetric impulse response functions

Proposition 1. *Consider the non-linear moving average model*

$$\begin{aligned} \boldsymbol{\Psi}_k(\boldsymbol{\varepsilon_{t-k}}) &= \boldsymbol{\Psi}_k(\varepsilon_{\ell,t-k}) && (13) \\ &= \left[\boldsymbol{\Psi}_k^{+} 1_{\varepsilon_{\ell,t-k}>0} + \boldsymbol{\Psi}_k^{-} 1_{\varepsilon_{\ell,t-k}<0}\right], \ \forall k \in \{0,..,K\}, \ \forall t \in \{1,..,T\} && (14) \end{aligned}$$

with $\ell \in \{1,..,L\}$, $\varepsilon_{\ell,t}$, *the* ℓ*th structural shock in* $\boldsymbol{\varepsilon_t}$ *and with* $\boldsymbol{\Psi}_0$ *satisfying Assumption 1. Then, given* $\{\boldsymbol{y_t}\}_{t=1}^{T}$, *given the model parameters and given* K *initial values of the shocks* $\{\boldsymbol{\varepsilon}_{-K}...\boldsymbol{\varepsilon}_0\}$, *the series of shocks* $\{\boldsymbol{\varepsilon_t}\}_{t=1}^{T}$ *is uniquely determined.*

Proof. We first establish the following lemma:

Lemma 1. *Consider a matrix* $\boldsymbol{\Gamma}$ *that can be written as*

$$\boldsymbol{\Gamma} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C}$ *and* $\mathbf{D}$ *are matrix sub-blocks of arbitrary size, with* A *a non-singular squared* *matrix and* $\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ *nonsingular. Then, the inverse of* $\boldsymbol{\Gamma}$ *satisfies*

$$\boldsymbol{\Gamma}^{-1}=\begin{pmatrix} \mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}\mathbf{F}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{F}^{-1} \\ -\mathbf{F}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{F}^{-1} \end{pmatrix}$$

with $\mathbf{F}=\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}$.

Proof. Verify that $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{-1}=\mathbf{I}$. ■

We prove Proposition 1 by induction, so that given past shocks $\{\boldsymbol{\varepsilon_{t-1-K}},...,\boldsymbol{\varepsilon_{t-1}}\}$ (and given model parameters $\{\boldsymbol{\Psi}_k\}_{k=0}^{K}$), we will prove that the system

$$\mathbf{u}_t=\boldsymbol{\Psi}_{\mathbf{0}}(\varepsilon_{\ell,t})\boldsymbol{\varepsilon_t} \tag{15}$$

with $\mathbf{u}_t=\boldsymbol{y_t}-\sum\limits_{k=0}^{K}\boldsymbol{\Psi}_k(\varepsilon_{\ell,t-1-k})\boldsymbol{\varepsilon_{t-1-k}}$, has a unique solution vector $\boldsymbol{\varepsilon_t}$.

Notice that (15) implies the sub-system with ℓ equations

$$\mathbf{u}_t^{\leq\ell}=\begin{pmatrix} \boldsymbol{\Psi}_0^{<\ell,<\ell} & \mathbf{0}^{<\ell,1} \\ \boldsymbol{\Psi}_0^{\ell,<\ell} & \boldsymbol{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t}) \end{pmatrix}\boldsymbol{\varepsilon_t}^{\leq\ell} \tag{16}$$

and notice that the matrix in (16) depends on $\varepsilon_{\ell,t}$ only through the scalar $\boldsymbol{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t})$. Denoting $\mathbf{A}\equiv\boldsymbol{\Psi}_{\mathbf{0}}^{<\ell,<\ell}$ a $(\ell-1)\times(\ell-1)$ invertible matrix (from Assumption 1), $\mathbf{C}\equiv\boldsymbol{\Psi}_{\mathbf{0}}^{\ell,<\ell}$ a $1\times(\ell-1)$ matrix, $\mathbf{B}\equiv\mathbf{0}$ of dimension $(\ell-1)\times 1$, and $D(\varepsilon_{\ell,t})\equiv\boldsymbol{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t})$ the (ℓ,ℓ) coefficient of $\boldsymbol{\Psi}_0$ (a scalar), we can use Lemma 1 to invert the system (16) and obtain

$$\boldsymbol{\varepsilon_t}^{\leq\ell}=\frac{1}{D(\varepsilon_{\ell,t})}\begin{pmatrix} D(\varepsilon_{\ell,t})\mathbf{A}^{-1} & \mathbf{0}^{<\ell,1} \\ -\mathbf{C}\mathbf{A}^{-1} & 1 \end{pmatrix}\mathbf{u}_t^{\leq\ell}. \tag{17}$$

The last row of (17) provides the equation $\varepsilon_{\ell,t}=\frac{1}{D(\varepsilon_{\ell,t})}\begin{pmatrix} -\mathbf{C}\mathbf{A}^{-1} & 1 \end{pmatrix}\mathbf{u}_t$, which defines $\varepsilon_{\ell,t}$. Since the right hand side of that equation only depends on $\varepsilon_{\ell,t}$ through $D(\varepsilon_{\ell,t})$, the sign of the right hand side depends on $\varepsilon_{\ell,t}$ only through the sign of $D(\varepsilon_{\ell,t})=\boldsymbol{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t})$. But since

$\boldsymbol{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t})$, the sign of the contemporaneous effect of the shock $\varepsilon_{\ell,t}$ on variable $y_{l,t}$, is posited to be positive as a normalization, the sign (and the value) of $\varepsilon_{\ell,t}$ is uniquely determined from the last row of (17). Then, with $\boldsymbol{\Psi}_{\mathbf{0}}^{<\ell,<\ell}$ and $\boldsymbol{\Psi}_{\mathbf{0}}^{>\ell,>\ell}$ invertible, (15) has a unique solution vector $\boldsymbol{\varepsilon_t}$. ■

Proposition 1 ensures that the system (11) has a unique solution vector, even when the shock $\varepsilon_{\ell,t}$, identified from a recursive ordering, triggers asymmetric impulse response functions.

With Proposition 1, we can then construct the likelihood recursively. To write down the one-step ahead forecast density $p(\boldsymbol{y_t}|\boldsymbol{\theta}, \boldsymbol{y^{t-1}})$ as a function of past observations and model parameters, we use the standard result (see e.g., Casella-Berger, 2002) that for $\boldsymbol{\Psi}_0$ a function of $\boldsymbol{\varepsilon_t}$, we have

$$p(\boldsymbol{\Psi}_0(\varepsilon_{\ell,t})\varepsilon_{\ell,t}|\boldsymbol{\theta}, \boldsymbol{y^{t-1}}) = J_t p(\boldsymbol{\varepsilon_t})$$

where J_t is the Jacobian of the (one-to-one) mapping from $\boldsymbol{\varepsilon_t}$ to $\boldsymbol{\Psi}_0(\boldsymbol{\varepsilon_t})\boldsymbol{\varepsilon_t}$ and where $p(\boldsymbol{\varepsilon_t})$ is the density of $\boldsymbol{\varepsilon_t}$.[13]

Finally, note that while we considered the case of a partially identified model, we can proceed similarly for a fully identified model with $\boldsymbol{\Psi}_0$ lower triangular and show that the shock vector $\boldsymbol{\varepsilon}_t$ is uniquely determined by (11) even when all shocks have asymmetric effects.

4.4.3 Asymmetric and state-dependent impulse response functions

We now consider a model with asymmetry and state dependence. For clarity of exposition, we consider the simpler case of a univariate state variable $z_t \in [\underline{z}, \overline{z}]$ with $\underline{z} = \min_{t\in[1,T]}(z_t)$ and $\overline{z} = \max_{t\in[1,T]}(z_t)$. The following proposition establishes the condition under which system (11) has a unique solution even when the identified shock $\varepsilon_{\ell,t}$ has asymmetric and state dependent effects.

[13]In our case with asymmetry, this Jacobian is simple to calculate, but the mapping is not differentiable at $\varepsilon_{\ell,t} = 0$. Since we will never exactly observe $\varepsilon_{\ell,t} = 0$ in a finite sample, we can implicitly assume that in a small neighborhood around 0, we replace the original mapping with a smooth function.

Proposition 2. *Consider the non-linear moving average model*

$$\mathbf{\Psi}_k(\boldsymbol{\varepsilon_{t-k}}, z_{t-k}) = \left[\mathbf{\Psi}_k^+(z_{t-k})1_{\varepsilon_{\ell,t-k}>0} + \mathbf{\Psi}_k^-(z_{t-k})1_{\varepsilon_{\ell,t-k}<0}\right], \ \ \forall k \in \{0,..,K\}, \forall t \in \{1,..,T\} \tag{18}$$

with $z_t \in [\underline{z}, \overline{z}]$, $\ell \in \{1,..,L\}$, $\varepsilon_{\ell,t}$, the ℓth structural shock in $\boldsymbol{\varepsilon_t}$, and with $\mathbf{\Psi}_0$ satisfying Assumption 1. Then, given $\{\boldsymbol{y_t}\}_{t=1}^T$, given the model parameters and given K initial values of the shocks $\{\boldsymbol{\varepsilon}_{-K}...\boldsymbol{\varepsilon}_0\}$, the series of shocks $\{\boldsymbol{\varepsilon_t}\}_{t=1}^T$ is uniquely determined provided that $sgn\left(\mathbf{\Psi}_{0,\ell\ell}^+(z_t)\right) = sgn\left(\mathbf{\Psi}_{0,\ell\ell}^-(z_t)\right) > 0$, $\forall z_t \in [\underline{z}, \overline{z}]$.

Proof. The proof proceeds exactly as with Proposition 1 and consists in showing that the system $\mathbf{u}_t = \mathbf{\Psi_0}(\varepsilon_{\ell,t}, z_t)\boldsymbol{\varepsilon_t}$ determines a unique solution vector $\boldsymbol{\varepsilon_t}$. As with Proposition 1, this is the case as long as $\mathbf{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t}, z_t) > 0$ regardless of the value of z_t. ■

Note that the restriction implied by Proposition 2 is very mild, in that it is in fact an existence condition for the moving average model, since the diagonal coefficients of $\mathbf{\Psi}_k$ are posited to be positive as a normalization. For instance, in our empirical application, it means that that the coefficient of the impact response of G to a G shock is always positive, regardless of the state of the cycle.

With Proposition 2 in hand, we can then construct the likelihood recursively as described in the previous section.

5 Auto-regressive models and asymmetric DGPs

Baseline VARs are linear models, but a popular way to introduce non-linearities in VARs is by means of regime-switching models, notably threshold VARs (e.g., Hubrich and Teräsvirta, 2013) and Markov-switching VARs (Hamilton, 1989). However, while regime-switching VARs can capture certain types of non-linearities, notably state dependence (whereby the value of some state variable affects the impulse response functions), regime-switching VARs face two major issues when the underlying data-generating process (DGP) features asymmetric effects of shocks (whereby the impulse response to a structural shock depends on the sign of that shock): (i) shock identification within the VAR is not possible, (ii) even if the underlying shocks were known, there is no parsimonious regime-switching auto-regressive representation of an asymmetric DGP.

To make our point, we consider an asymmetric DGP like the one considered in the main text

$$y_t = \sum_{k=0}^{K} \psi_k^{\pm}(\varepsilon_{t-k})\varepsilon_{t-k} \tag{19}$$

where

$$\psi_k^{\pm}(\varepsilon_{t-k}) = \psi_k^{+}\mathbf{I}(\varepsilon_{t-k} > 0) + \psi_k^{-}\varepsilon_{t-k}\mathbf{I}(\varepsilon_{t-k} \leq 0).$$

ψ^+ is the impulse response triggered by a positive shock and ψ^- the impulse response triggered by a negative shock.

5.1 Shock identification

Intuitively, the identification problem with a regime-switching VAR boils down to the fact that the parameters of that model are, as of period t, independent of the shock ε_t, the variable responsible for the non-linearity (i.e., triggering the regime change).

More specifically, suppose a researcher had access to the true one-step ahead forecast errors u_t (so that misspecification of the conditional mean is for now not an issue, see the next section). With a VMA representation like the one we used in the paper, identifying the structural shocks of interest ε_t boils down to solving the equation

$$u_t = \psi_0^{\pm}(\varepsilon_t)\varepsilon_t \tag{20}$$

where $\psi_0^{\pm}(\varepsilon_t)$ is the impact coefficient that depends on the sign of ε_t. By contrast, somebody working with regime switching models would try to solve an equation of the form

$$u_t = \tilde{\psi}(s_t)\tilde{\varepsilon}_t \tag{21}$$

where s_t is the state variable of the regime switching model. It can be either a function of *lagged* endogenous variables (as in a threshold model) or a random variable governed by a discrete state Markov chain (as in a Markov-switching model). For the latter, the transition probabilities of the discrete Markov state are either assumed to be exogenous (and independent of the other shocks in the model, in particular independent of ε_t) or dependent *only on lagged endogenous variables* (e.g., Sims et al., 2008). Importantly, for either case —a threshold model or a Markov-switching model—, s_t ends up independent of ε_t and hence identification of the true value of ε_t, i.e., solving (21), is not possible (unless we are back to the symmetric (linear) case where $\psi_0^{\pm}(\varepsilon_t)$ does not depend on ε_t): the researcher can not find a $\tilde{\varepsilon}_t$ such that $\tilde{\varepsilon}_t = \varepsilon_t$. Thus, when shocks have asymmetric effects, it is not possible to use a regime-switching VAR and identify stuctural shocks within the model, e.g., by using a recursive ordering as in Auerbach and Gorodnichenko (2012).

5.2 Approximating asymmetric DGPs with AR models

When the shocks have been independently identified, for instance through a narrative approach, shock identification is not an issue. However, even in that case a regime-switching model cannot parsimoniously capture a DGP with asymmetric impulse responses.

To make this point, we consider a simple asymmetric Moving-Average (MA) model and show that there is no parsimonious regime-switching Auto-Regressive (AR) representation of that DGP. Specifically, consider as DGP a univariate asymmetric model[14]

$$y_t = \sum_{k=1}^{H} \left(\rho_{\pm}(\varepsilon_{t-k})\right)^k \varepsilon_{t-k} \tag{22}$$

[14] We consider a univariate model for clarity of exposition, but the argument would be identical with a multivariate model.

with $\rho_{\pm}(\varepsilon_{t-k}) = \rho_{+}\mathbf{I}(\varepsilon_{t-k} > 0) + \rho_{-}\mathbf{I}(\varepsilon_{t-k} \leq 0)$ and $\rho_{+} < 1$ and $\rho_{-} < 1$ the model parameters. Model (22) is like a generalized AR(1), in that the effect of a shock ε_t on y decays geometrically, but the decay rate depends on the sign of the shock ε_t. The model reduces to a standard (symmetric) AR(1) when $\rho_{+} = \rho_{-}$.

Clearly it is not possible to invert (22), i.e., it is not possible to represent (22) with an AR model, unless the MA is symmetric ($\rho_{+} = \rho_{-}$). The more interesting question however is whether it is possible to represent (22) as a regime-switching AR model of the form

$$y_t = \alpha(s_t)y_{t-1} + \beta(s_t)\varepsilon_t \tag{23}$$

where s_t denotes the state at time t and where ε_t is known at time t (that is assuming that the shocks have been independently identified, to avoid the problem highlighted in the previous section). This Regime-Switching AR (RSAR) representation allows for a new set of AR coefficients in each state. The question is then whether it is possible to represent (22) with a reasonable number of states.

Under standard regularity conditions, (23) will have a moving average representation of the form

$$y_t = \beta(s_t)\varepsilon_t + \sum_{k=1}^{\infty}\left[\prod_{j=1}^{k}\alpha(s_{t-j})\right]\beta(s_{t-k})\varepsilon_{t-k} \tag{24}$$

which implies that the RSAR model will capture (22) if the coefficients $\{\alpha(s_{t-k}),\beta(s_{t-k})\}$ satisfy the relation

$$\begin{cases} \beta(s_{t-k}) = \rho_{\pm}(\varepsilon_{t-k}) \\ \alpha(s_{t-k}) = \frac{(\rho_{\pm}(\varepsilon_{t-k}))^k}{\prod_{j=1}^{k-1}\alpha(s_{t-j})} \end{cases}. \tag{25}$$

This recursive definition of $\{\alpha(s_{t-k}),\beta(s_{t-k})\}$ implies that, in order for the RSAR model to capture the asymmetric MA process (22), a new state variable is needed for each new shock realization. Given $k-1$ state variables $\{s_t, ..., s_{t-k+1}\}$, the effect of the $t-k$ shock ε_{t-k} on y_t cannot be captured by these $k-1$ state variables. Instead, a new state variable s_{t-k} needs to be introduced at time $t-k$ with $\alpha(s_{t-k})$ and $\beta(s_{t-k})$ given by (25).

Since $\rho_+ < 1$ and $\rho_- < 1$, shocks only have a transitory effect on y, and only a finite number of state variables are needed to *approximate* (22) with a RSAR model. However, except for highly transitory processeses (and unlike the processes driving government spending and output), a substantial number of states will be needed, far larger than the 2 to 3 number of regimes encountered in typical applications (e.g., Hubrich and Terasvirta, 2013). To give an order of magnitude, with a persistence parameter of .8 (roughly representative for quarterly macro data) the effect of a shock is only 10 percent of its initial value after 10 periods, implying that about 2^{10} different regimes would be needed to capture a simple asymmetric AR(1) model like (22), a prohibitively large parameter space.

6 Solution Method of the Theoretical Model

As explained in the main text, the model can be summarized by the following three equations:

$$\begin{aligned}
&C_t^e Y_t + C^u (1 - Y_t) = Y_t - G_t \\
&(C_t^e)^{-\sigma} = \beta \bar{R} \Pi_t^{\phi_\pi} \exp\{z_t\} \mathbb{E}_t \left\{ \left(C_{t+1}^e\right)^{-\sigma} \Pi_{t+1}^{-1} \left[(1-\delta) + \delta \left(\frac{C^u}{C_{t+1}^e} \right)^{-\sigma} \right] \right\} \\
&\left(Y_t - \bar{Y}\right) \left[\Pi_t - \gamma (1 - Y_t)\right] = 0
\end{aligned}$$

To solve the model, we approximate the expectation term on the RHS of the consumption Euler equation through a Chebyshev polynomial (of order 7) for each of the two state variable (z_t, G_t), i.e. we approximate the function

$$X(z_t, G_t) \equiv \mathbb{E}_t \left\{ \left(C_{t+1}^e\right)^{-\sigma} \Pi_{t+1}^{-1} \left[(1-\delta) + \delta \left(\frac{C^u}{C_{t+1}^e} \right)^{-\sigma} \right] \right\}.$$

Approximating the expectation term rather than the policy functions is convenient in the presence of occasionally binding constraint, since the expectation term is a smooth function, while the policy functions would display some kinks at points where the constraint becomes binding.

For both state variables (the exogenous shocks) we build a grid of 7 Chebyshev nodes, so that the polynomial basis function are orthogonal at the gridpoints between -3 and +3 standard deviations of the shocks.

The numerical algorithm is the following:

1. Guess of the coefficient of the polynomial function and calculate $X(\cdot)$ at all grid points.

2. Given the values of $X(\cdot)$, calculate the policy functions (i.e. the values at all gridpoints) of output, inflation, and consumption of employed households using the three equations of the model.

3. Using the policy functions for consumption and inflation, calculate the implied expectation term in the consumption Euler equation, integrating over 100 Gaussian quadrature nodes for innovations of the two shock processes.

4. If the difference between the guessed functions and the expectations calculated in point 3) is below a tolerance level, stop. Otherwise, iterate on 1) - 4) until convergence.

In practice, the solution is found using a quasi-Newton rootfinding method, iterating on the coefficients of the polynomial until the residual calculated in point 4) is below the tolerance level at all gridpoints.

As an accuracy test, we calculated the Euler residuals at 500 equally spaced nodes (not on the original grid) in the domain of the state variables. As illustrated in Figure A9, the Euler residuals are smaller than 10^{-6} throughout the entire domain, with a mean value below 10^{-7}.

7 Model impulse responses

Figure A10 plots the model impulse responses to a government spending shock, showing how government spending can have asymmetric effects on output.[15]. Following an expansionary

[15]The same reasoning also applies for state-dependence in the multiplier

shock to government spending (blue line), output goes up and unemployment goes down. In this case, total private consumption is crowded out by public spending (top-right panel, dashed-blue line). As discussed in the main text, government spending stimulates total (private) consumption through the "consumption gap" effect. However, with convexity in the (AS) curve inflation goes up strongly following the increase in aggregate demand (left-bottom panel, dashed-blue line), which leads to a strong interest rate response by the central bank and thereby to an ultimate contraction in consumption. In contrast, following a contractionary spending shock, output declines and unemployment increases, which lowers total private consumption through the "consumption gap" effect (for comparability, the impulse responses to contractionary shocks are multiplied by -1). However, this time the reaction of the central bank is not enough to overcome the "consumption gap" effect: with convexity in the (AS) curve the response of inflation is milder (bottom-left panel, plain-red line), and thus the associated interest rate response is not enough to avoid a decline in consumption.

8 Government Spending Multiplier and Persistence

As is well known, one factor that may affect the size of the government spending multiplier is whether shocks are expected to be temporary or persistent.

For instance, standard New-Keynesian models (see e.g., Galí et al., 2007; Christiano et al., 2011; Woodford, 2011) predict that more persistent government spending shocks are associated with smaller fiscal multipliers.[16] This is because consumption at any point in time depends solely on the path of current and future real interest rates (as indicated by the consumption Euler equation). For example, if the central bank follows a standard Taylor rule, an increase in government spending leads to higher inflation and interest rates (both nominal and real), and such an increase in real interest rates leads to a fall in consumption, thus partially counteracting

[16]Standard neo-classical models predict that increased persistence of government purchases leads to a larger impact effects on output (see e.g. Barro (1989) and Baxter and King 1993), mainly due to stronger wealth effects.

the effects of the government spending increase. When the shock is temporary, only the current real rate increases, leading to a small reduction in current consumption. Instead, if the shock is persistent, both the current and the future real rates would increase, thus leading a larger drop in current consumption. For these reasons, the fiscal multiplier is larger for temporary shocks than for persistent shocks.

In that respect, one potential explanation for our empirical findings could be that expansionary shocks have more persistent effects on government spending than contractionary ones, and for that reason the corresponding multiplier is smaller. Our empirical analysis is not conclusive about this aspect. Expansionary shocks seem to have less persistent effects on government spending than contractionary shocks under the recursive identification scheme, but the opposite happens under the narrative identification scheme (Figs 2 and 3 in the main text).

Nonetheless, we can use our theoretical model to assess quantitatively the relationship between persistence and multipliers. To that end, Figure A11 plots the spending multiplier (sum over 20 quarters) for different degrees of ρ_g —the parameter controlling the persistence of government spending—, both for expansionary and contractionary shocks, assuming the economy is initially at the average level of unemployment. The quantitative effects of persistence can be gauged by looking at the size of the contractionary multipliers.[17] For example, as the persistence increases from 0.5 to 0.9 —a wide range of plausible values for ρ_g— the multiplier declines from a value of about 1.4 to about 1.2. This suggests that the difference in the degree of persistence could contribute, but would not suffice, to explain the large asymmetries between contractionary and expansionary multipliers found in the data.

The comparison between the contractionary and expansionary multiplier in Figure A11 also suggests that the higher is ρ_g, the larger is the asymmetry between the multipliers. This

[17]Starting from a situation where the economy is at (or below) full-employment, a contractionary spending shock will push (or continue to push) the economy below full employment, and nominal rigidities will start binding —eq. (14) in the main text—. Thus, the effects of ρ_g on the multiplier is not interacting with the *asymmetric* nature of downward wage rigidities. In contrast, after an expansionary shock, the nominal rigidities may or may not be binding, because a positive shock may or may not push the economy to full employment, at which point the multiplier is zero. This mechanism will blur the effect of persistence on the multiplier. For instance, the more persistent is the expansionary shock, the more likely it is that the economy operates at full employment (where the multiplier is zero), before the shock completely vanishes.

is due to the interaction between the effects of persistence in government spending and wage rigidities. In response to a positive spending shock, the more persistent is the shock the more time the economy operates at high employment level, where wage rigidities are less binding, crowding-out is larger and the multiplier is relatively small. Vice versa, in response to a persistent negative shock, the economy stays longer at higher level unemployment, where wage rigidities are more severe and crowding-out weaker, and the multiplier is relatively large.

References

Auerbach, A. J. and Y. Gorodnichenko (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy 4*(2), 1–27.

Barro, R. J. (1989). The ricardian approach to budget deficits. *Journal of Economic perspectives 3*(2), 37–54.

Barro, R. J. and C. J. Redlick (2011). Macroeconomic effects from government purchases and taxes. *The Quarterly Journal of Economics 126*(1), 51–102.

Baxter, M. and R. G. King (1993). Fiscal policy in general equilibrium. *The American Economic Review*, 315–334.

Blanchard, O. J. and D. Leigh (2013). Growth forecast errors and fiscal multipliers. *American Economic Review 103*(3), 117–20.

Christiano, L., M. Eichenbaum, and S. Rebelo (2011). When is the government spending multiplier large? *Journal of Political Economy 119*(1), 78–121.

Galí, J., J. D. López-Salido, and J. Vallés (2007). Understanding the effects of government spending on consumption. *Journal of the european economic association 5*(1), 227–270.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 357–384.

Hubrich, K. and T. Teräsvirta (2013). Thresholds and smooth transitions in vector autoregressive models. *VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims*, 273–326.

Perotti, R. (2014). Defense government spending is contractionary, civilian government spending is expansionary. Technical report, National Bureau of Economic Research.

Ramey, V. A. and S. Zubairy (2018). Government spending multipliers in good times and in bad: evidence from us historical data. *Journal of Political Economy 126*(2), 850–901.

Sims, C. A., D. F. Waggoner, and T. Zha (2008). Methods for inference in large multiple-equation markov-switching models. *Journal of Econometrics 146*(2), 255–274.

Woodford, M. (2011). Simple analytics of the government expenditure multiplier. *American Economic Journal: Macroeconomics 3*(1), 1–35.

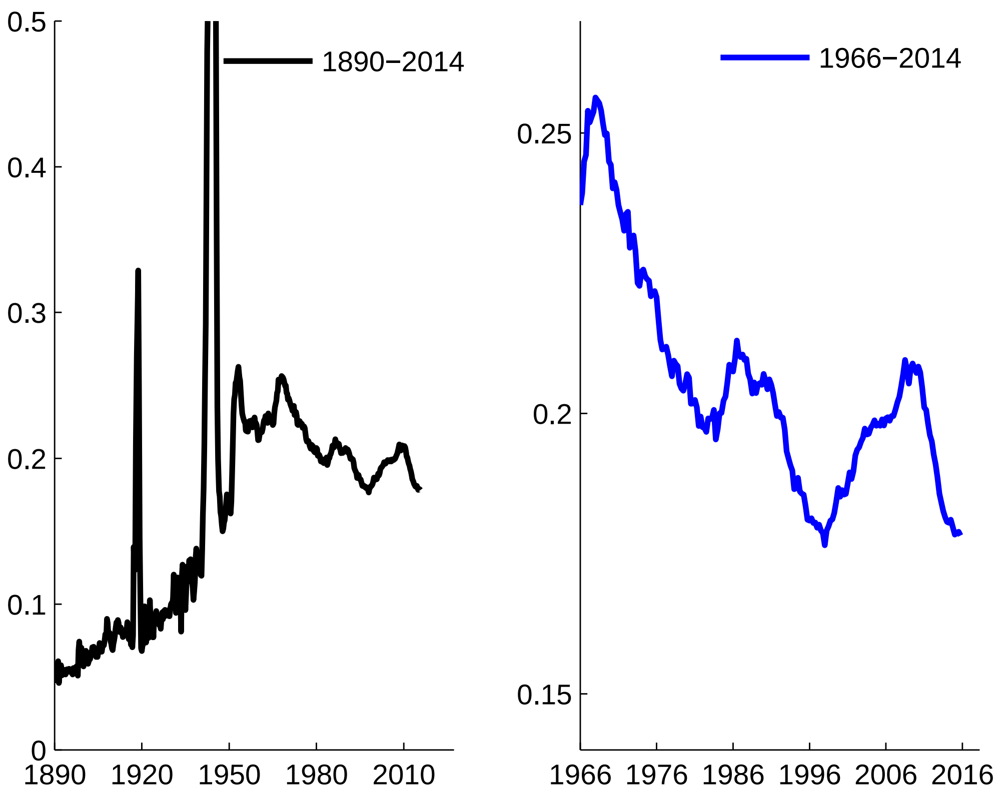

Figure A1: Government spending to potential output over 1890-2014 (left-panel) and 1966-2014 (right-panel)

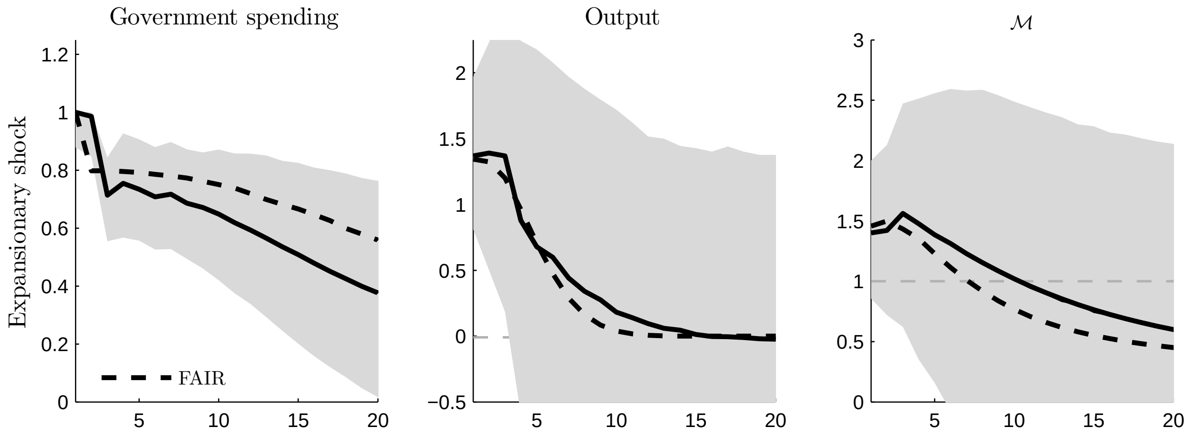

Figure A2: **Recursive identification scheme — VAR, 1966-2014**. Impulse response functions (in percent) of government spending and output to a government spending shock along with the corresponding multiplier $\mathcal{M}$. Recursive identification as in Auerbach and Gorodnichenko (2012). Estimates from a standard VAR (plain-line) or from a linear FAIR with one Gaussian basis function (dashed line). The shaded areas cover 90% of the posterior probability of the VAR estimates.

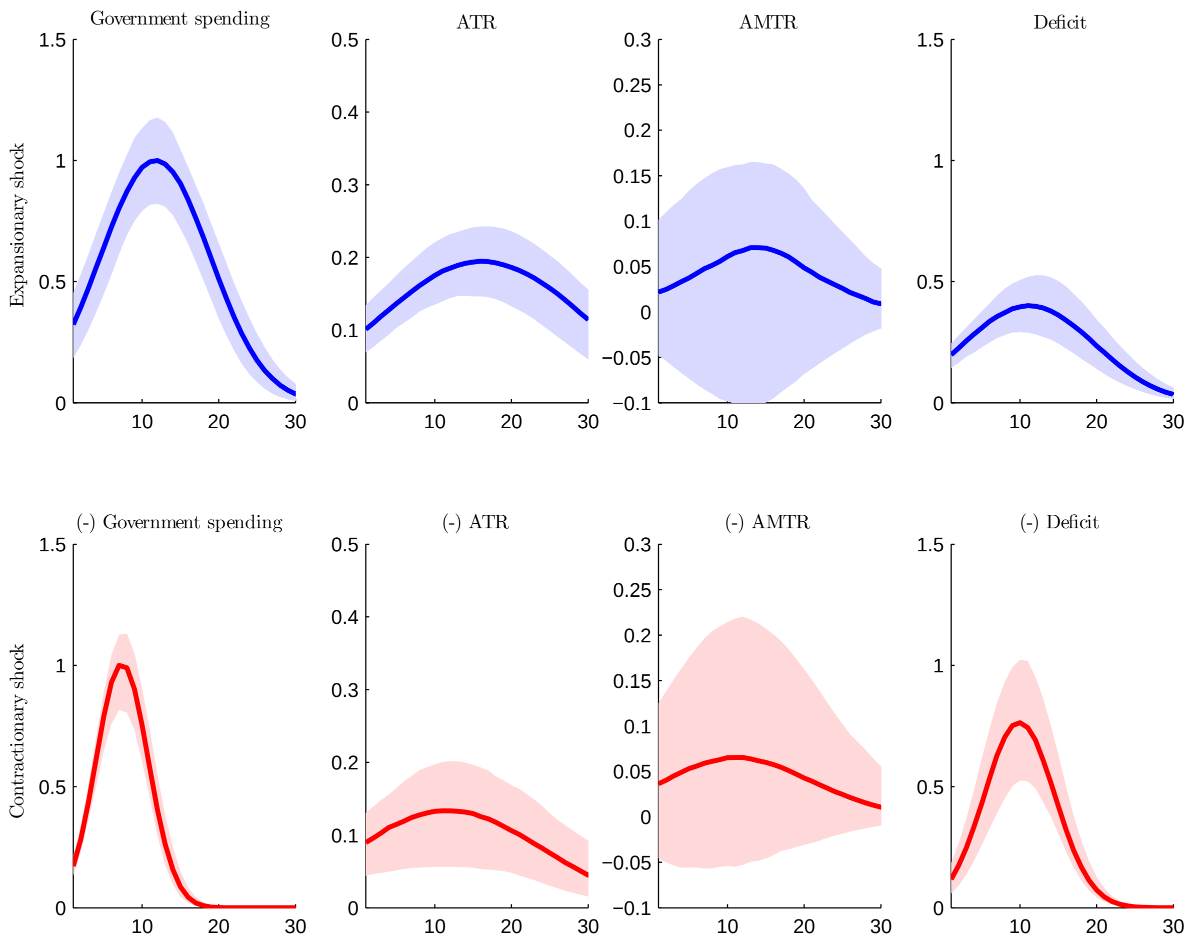

Figure A3: **Narrative identification scheme, FAIR, 1912-2014**. Impulse response functions (in percent) of government spending, average tax rate ("ATR"), the average marginal tax rate ("AMTR") and the fiscal deficit "Deficit") to a government spending shock. The shaded areas cover 90% of the posterior probability. For ease of comparison between the top and bottom panels, the responses to a contractionary shock are multiplied by -1 in the bottom panels.

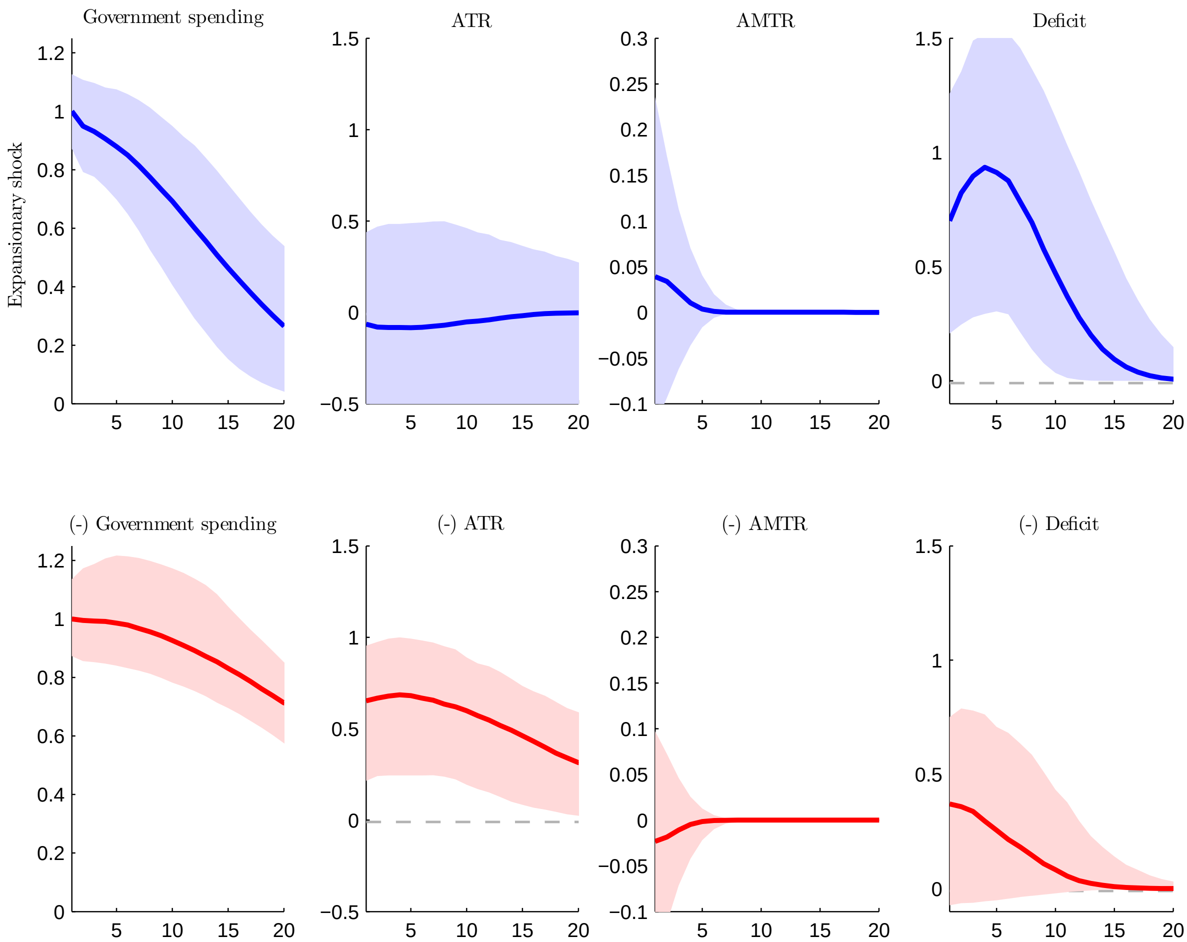

Figure A4: **Recursive identification scheme, FAIR, 1966-2014**. Impulse response functions (in percent) of government spending, average tax rate ("ATR"), the average marginal tax rate ("AMTR") and the fiscal deficit "Deficit") to a government spending shock. The shaded areas cover 90% of the posterior probability. For ease of comparison between the top and bottom panels, the responses to a contractionary shock are multiplied by -1 in the bottom panels.

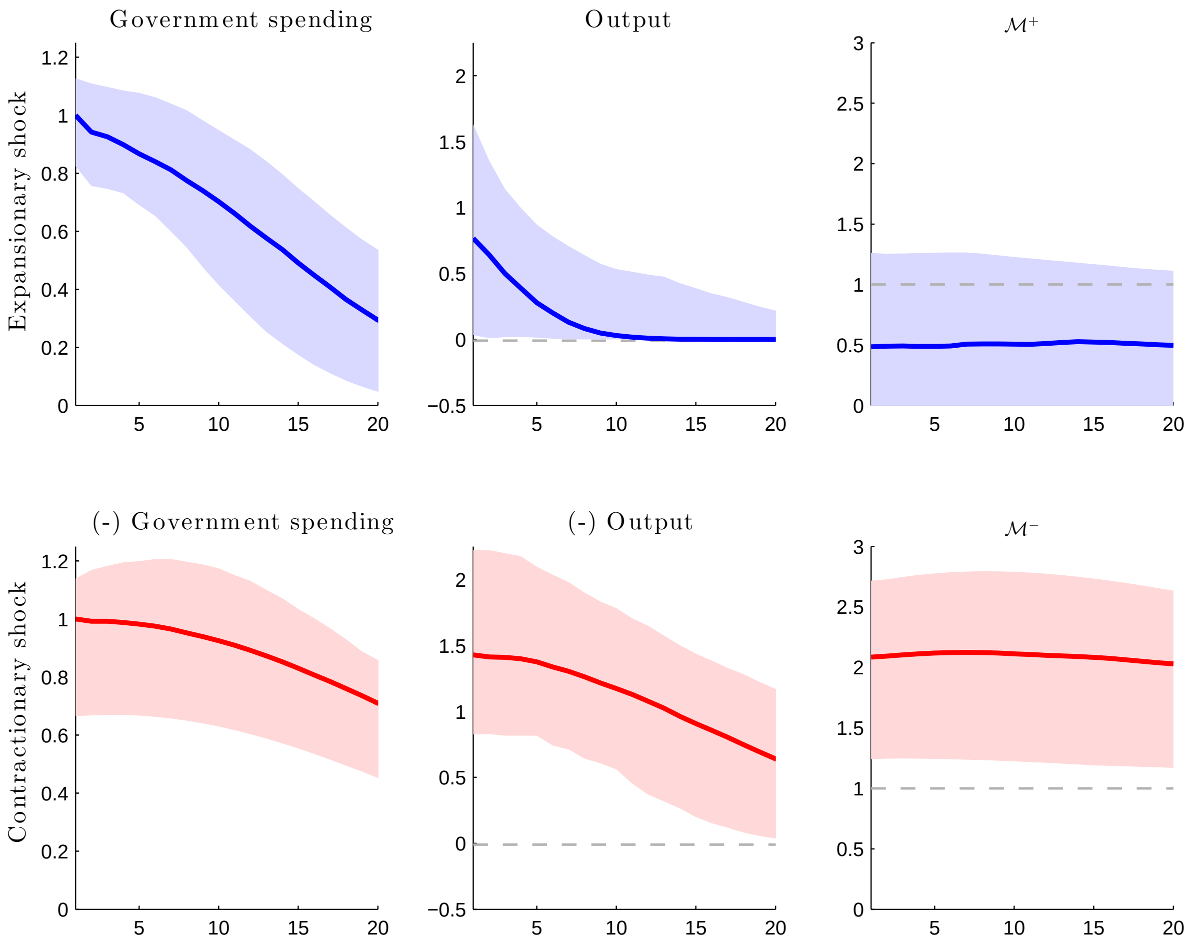

Figure A5: **Recursive identification scheme, 1966-2014**. Impulse response functions (in percent) of non-defense government spending and output to a government spending shock identified from a timing restriction. Estimation from a FAIR model (plain line). The shaded areas cover 90% of the posterior probability. For ease of comparison between the top and bottom panels, the responses to a contractionary shock are multiplied by -1 in the bottom panels.

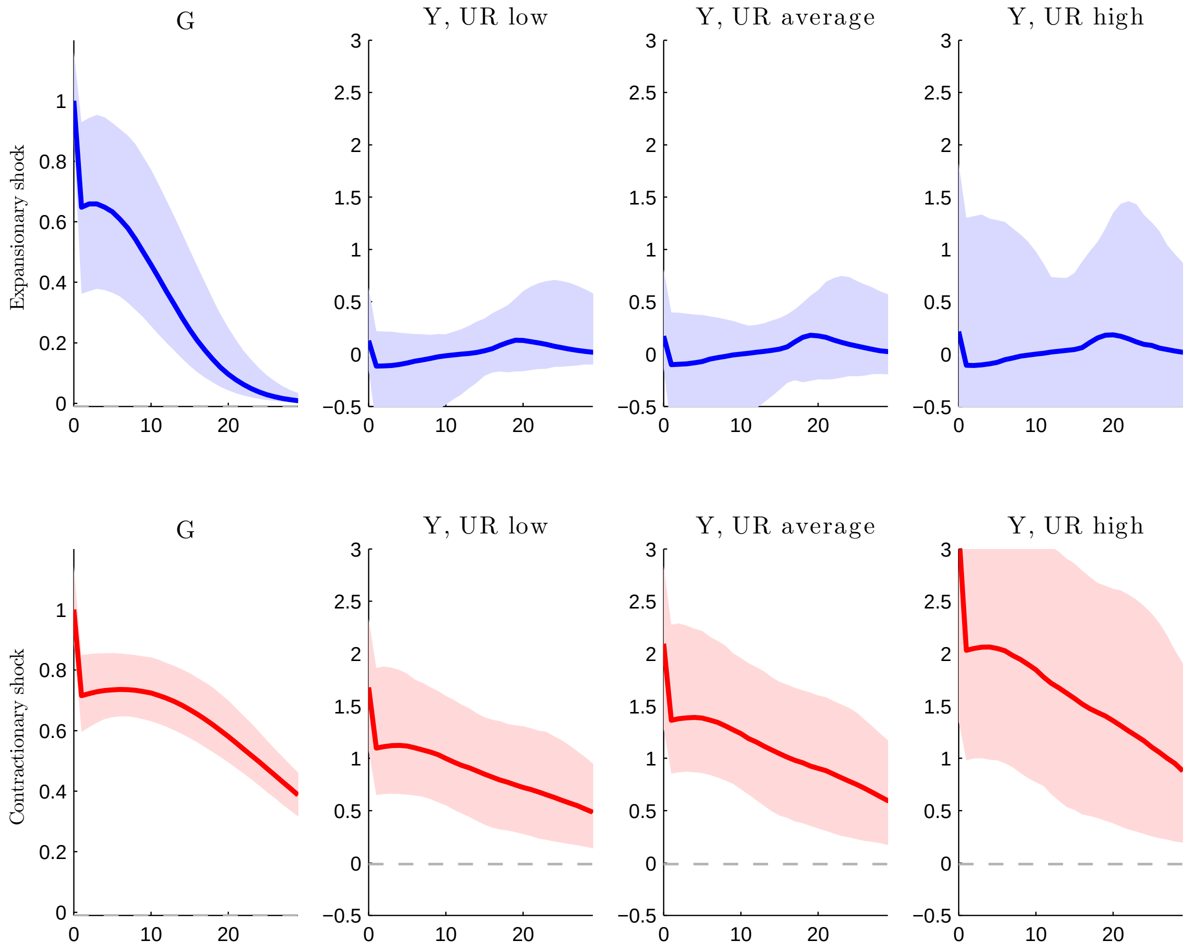

Figure A6: **Recursive identification scheme, FAIR, 1947-2014**. Effect of (detrended) unemployment (UR) on the impulse responses of government spending and output to a government spending shock. “UR high” , “UR average” and “UR low” respectively denotes values of the detrended UR of $+2$, 0 and -1. With our modeling of state dependence (whereby the level of slack only changes the amplitude of the impulse response), the impulse response of government spending, *normalized* to peak at one, is constant. The thin lines cover 90% of the posterior probability. For ease of comparison between the top and bottom panels, the responses to a contractionary shock are multiplied by -1 in the bottom panels.

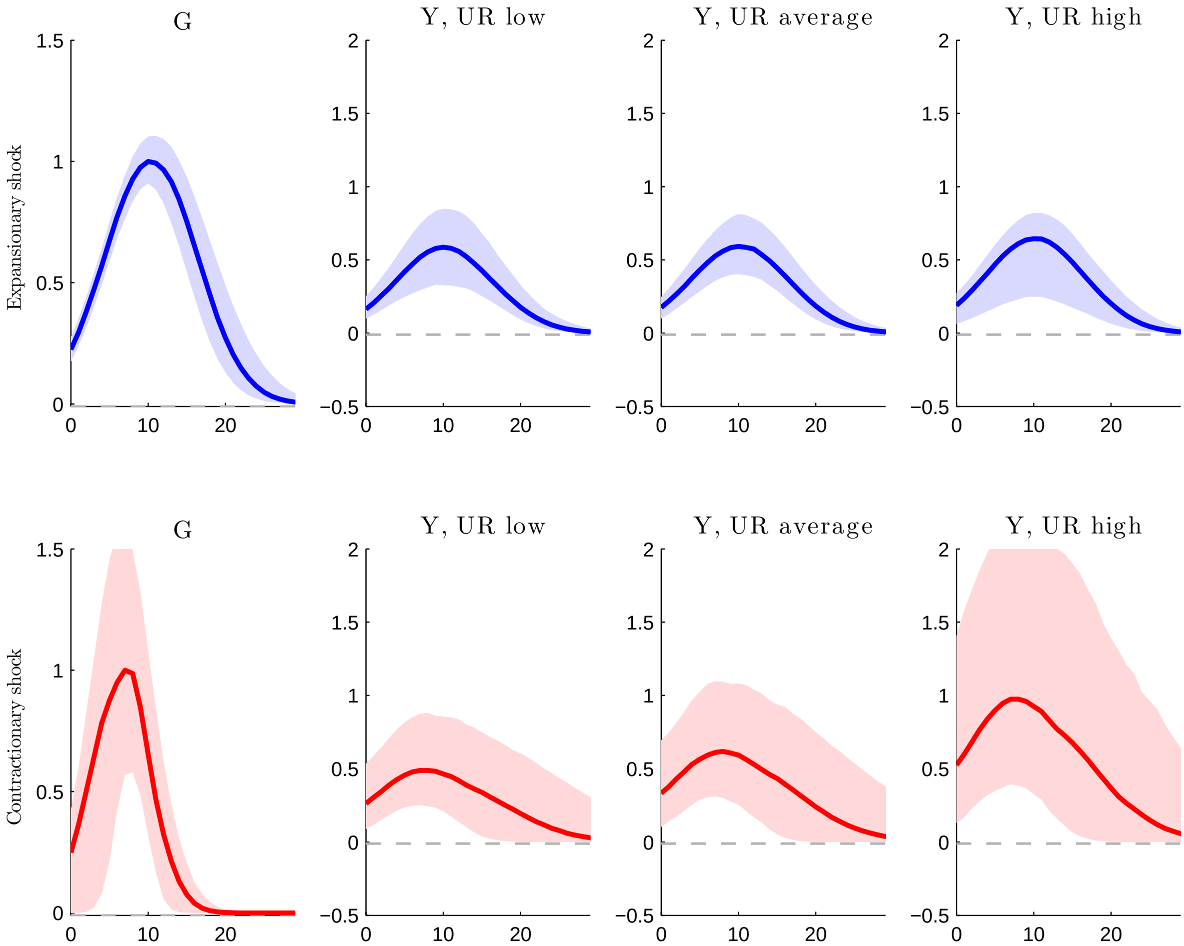

Figure A7: **Narrative identification scheme, FAIR, 1947-2014**. Effect of (detrended) unemployment (UR) on the impulse responses of government spending and output to a government spending shock. "UR high" , "UR average" and "UR low" respectively denotes values of the detrended UR of +2, 0 and −1. With our modeling of state dependence (whereby the level of slack only changes the amplitude of the impulse response), the impulse response of government spending, *normalized* to peak at one, is constant. The thin lines cover 90% of the posterior probability. For ease of comparison between the top and bottom panels, the responses to a contractionary shock are multiplied by -1 in the bottom panels.

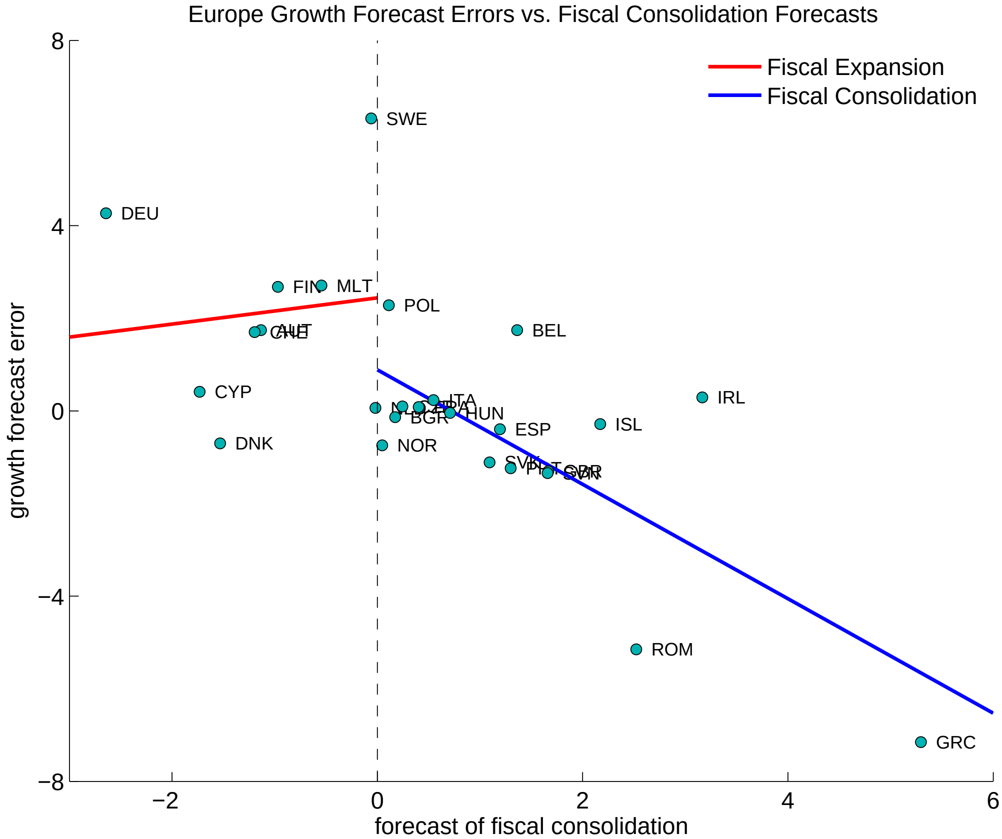

Figure A8: **Blanchard and Leigh (2013) approach**: Regression of forecast error for real GDP growth in 2010 and 2011 relative to forecasts made in the spring of 2010 on forecasts of fiscal consolidation for 2010 and 2011 made in spring of year 2010. The lines depict the regression lines for respectively fiscal consolidation (increase in budget surplus, blue line) and fiscal expansion (decrease in budget surplus, red line). Note that a fiscal consolidation corresponds to an *increase* in the fiscal balance and thus enters as a positive entry on the x-axis.

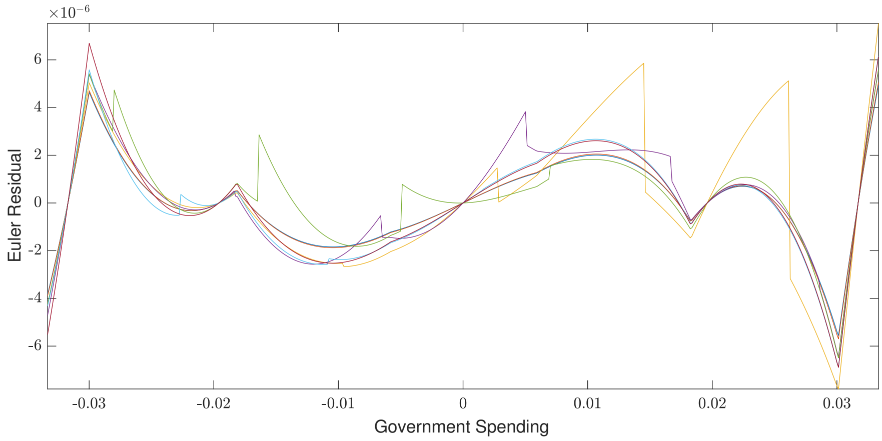

Figure A9: **Euler Residuals**. The figure plots the Euler residuals resulting from our solution algorithm on 500 points in the domain of the government spending shocks. Each line corresponds to a different realization of the discount factor shock

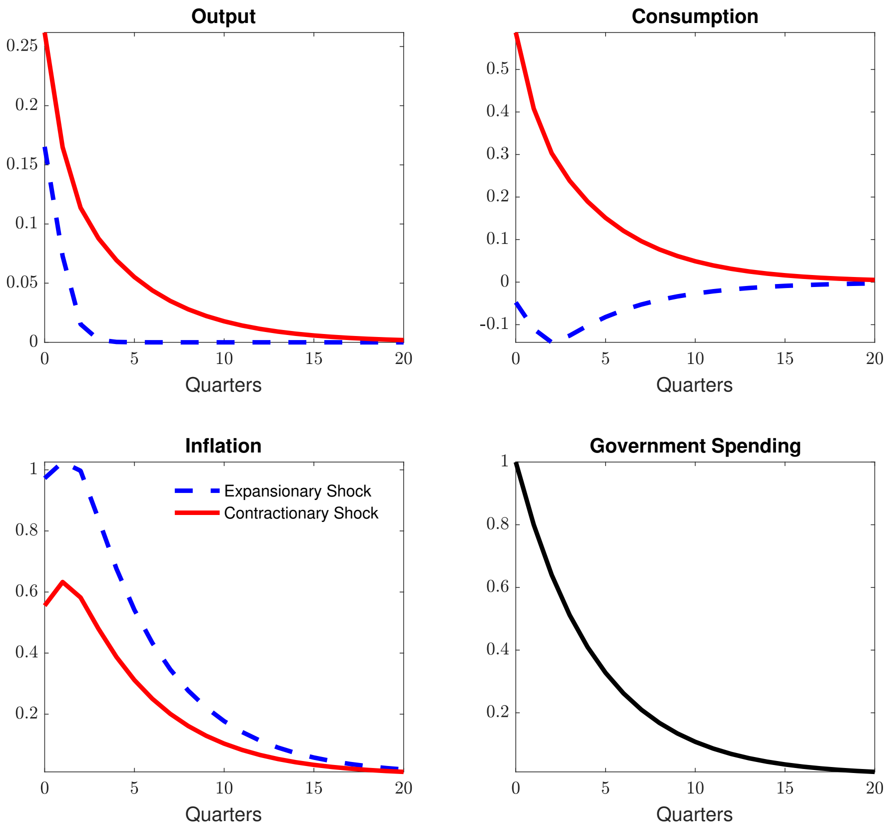

Figure A10: **Impulse responses to a Government Spending Shock**. The figure plots the impulse responses to a government spending shock for expansionary (blue dashed line) and contractionary (red solid line). Each line represents the average impulse response across realizations of the discount factor shock, which determine the state of the economy when the government shock hits. For ease of comparison, the responses to a contractionary shock are multiplied by -1.

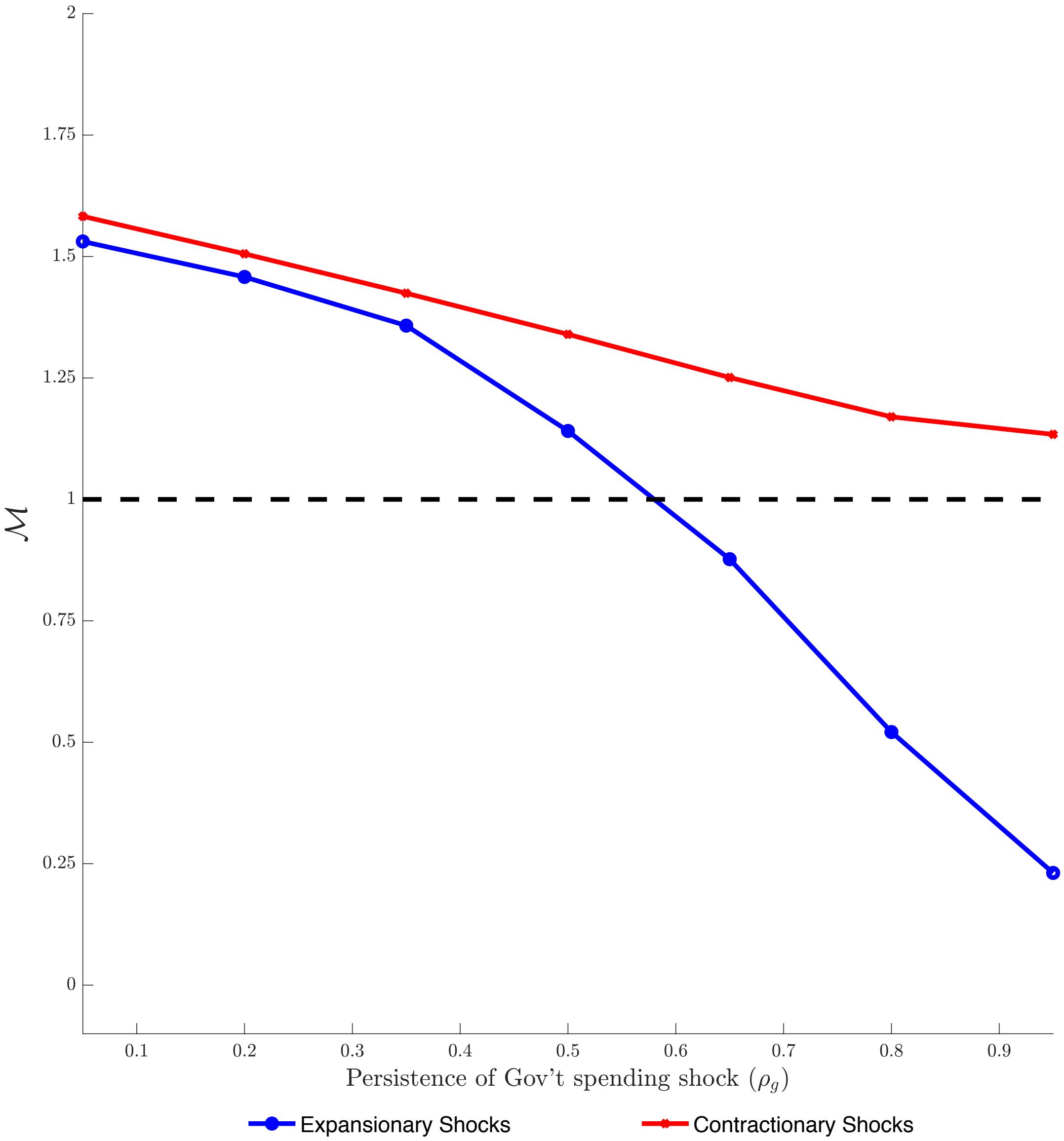

Figure A11: **Multipliers and Persistence of Government Spending Process in Theoretical Model**. The figure plots the government spending multiplier (sum of 20 quarters) for expansionary (blue line with circles) and contractionary (red line with crosses) as a function of the autocorrelation of the government spending process ρ_g. In all cases, the multiplier is calculated assuming that the unemployment rate is initially at its average value (as implied by the initial value of the discount factor shock).