Online Appendix

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

Regis Barnichon and Christian Matthes

This appendix describes (i) the SUR-FAIR model and its estimation, (ii) the VMA-FAIR model and its estimation, and then provides additional empirical results supporting our findings from the main text.

1 SUR-FAIR models and External Instruments

The section describes a fast and efficient procedure (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

$$\begin{pmatrix} Y_t \\ G_t \end{pmatrix} = \sum_{k=0}^{K} \begin{pmatrix} \varphi_Y(k) \\ \varphi_G(k) \end{pmatrix} \hat{\varepsilon}^G_{t-k} + \begin{pmatrix} u^Y_t \\ u^G_t \end{pmatrix} \tag{1}$$

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

$$\hat{\varepsilon}^G_t = \alpha\varepsilon^G_t + \eta_t$$

with η_t i.i.d with variance σ^2_η. In the language of Instrument Variables, $\hat{\varepsilon}^G_t$ 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 $\hat{\varepsilon}^G_t$ is only correlated with the true shock ε^G_t, 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

$$\hat{\psi}_Y = \frac{\hat{\varphi}_Y}{\hat{\varphi}_G(h_0)}$$

with h_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 (1). The computational advantage of this SUR-FAIR approach (compared to a VMA-FAIR model as in Appendix A2) 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

$$y_t = \sum_{k=0}^{K} \psi(k)\hat{\varepsilon}^G_{t-k} + u_t \tag{2}$$

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 $\hat{\varepsilon}^G_{t-k}$ (in the non-linear case).

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 $N(0, \sigma^2_\eta)$ with σ_η a parameter to be estimated. We set η_{-1} and η_0 to zero, and from (2), it is straightforward to build the likelihood given a series of previously identified shocks $\left\{\hat{\varepsilon}_t^G\right\}$. For prior elicitation, we use very loose priors with $\sigma_a = 10$, $\sigma_b = K$ and $\sigma_c = K$.

For a SUR-FAIR model like (1), 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 section 2. 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 univariate and 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 (2), 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 (2) 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 ε_{t-k}^G. We then re-arrange (2) as follows:

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

Defining

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

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

$$y_t = \sum_{n=1}^{N} a_n X_{n,t} + \alpha + \beta u_t \tag{3}$$

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

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

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 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 (3) using a simplex algorithm.

2 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 Auerbach and Gorodnichenko (2012). As in the main text, for $\boldsymbol{y}_t$ a vector of stationary macroeconomic variables, the VMA model writes

$$\boldsymbol{y}_t = \sum_{k=0}^{K} \boldsymbol{\Psi}_k(\boldsymbol{\varepsilon}_{t-k}, z_{t-k}) \boldsymbol{\varepsilon}_{t-k} \tag{4}$$

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 $\boldsymbol{\Psi}_k = \boldsymbol{\Psi}_k(\varepsilon_{t-k}, z_{t-k})$.

2.1 Likelihood function

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

$$p(\boldsymbol{y}^T|\boldsymbol{\theta}) = \prod_{t=1}^{T} p(\boldsymbol{y}_t|\boldsymbol{\theta}, \boldsymbol{y}^{t-1}). \tag{5}$$

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,[1] and we note that the density $p(\boldsymbol{y}_t|\boldsymbol{y}^{t-1}, \boldsymbol{\theta})$ can be re-written as $p(\boldsymbol{y}_t|\boldsymbol{\theta}, \boldsymbol{y}^{t-1}) = p(\boldsymbol{\Psi}_0\boldsymbol{\varepsilon}_t|\boldsymbol{\theta}, \boldsymbol{y}^{t-1})$ since

$$\boldsymbol{y}_t = \boldsymbol{\Psi}_0\boldsymbol{\varepsilon}_t + \sum_{k=1}^{K} \boldsymbol{\Psi}_k\boldsymbol{\varepsilon}_{t-k}. \tag{6}$$

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 (6), 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.[2] 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 $\{\varepsilon_j\}_{j=-K}^{0}$ to zero.

In the non-linear case where we have $\boldsymbol{\Psi}_k = \boldsymbol{\Psi}_k(\varepsilon_{t-k}, z_{t-k})$, 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 (6):

$$\boldsymbol{\Psi}_0(\varepsilon_{t-k}, z_{t-k})\boldsymbol{\varepsilon}_t = \mathbf{u}_t \tag{7}$$

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 (7) has a unique solution vector $\boldsymbol{\varepsilon}_t$) in a recursive identification scheme like the one considered in this paper.

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

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

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

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

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 $\boldsymbol{\varepsilon}_t$.[3,4]

2.2 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 $p(y^T|\theta)p(\theta)$.[5] 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

[3]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).

[4]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.

[5]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

2.3 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.[6]

2.4 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

$$\boldsymbol{\Psi}_0(\boldsymbol{\varepsilon_t}, \boldsymbol{z_t})\boldsymbol{\varepsilon_t} = \mathbf{u}_t \tag{8}$$

[6] This is easy to see for a and b. For c, if it easy to show that $c\sqrt{\ln 2}$ is the half-life of the effect of a shock. Thus, $c = K$ corresponds to a very persistent impulse response function, since $K\sqrt{\ln 2} = 38$ quarters.

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.[7]

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.[8]

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 *while also* imposing identifying restrictions on $\boldsymbol{\Psi}_0$. The complication arises, because with asymmetry, the system of equations $\boldsymbol{\Psi}_0(\boldsymbol{\varepsilon_t})\boldsymbol{\varepsilon_t} = \mathbf{u}_t$ 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 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.

2.4.1 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 $\boldsymbol{\Gamma}$ of size $L \times L$ and $(i, j) \in \{1, ..., L\}^2$, denote $\boldsymbol{\Gamma}^{<i,<j}$ the $(i-1) \times (j-1)$ submatrix of Γ made of the first $(i-1)$ rows and $(j-1)$ columns. Similarly, we denote

[7] 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 $\mathbf{u}_t$.

[8] 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.

$\mathbf{\Gamma}^{>i,>j}$ 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 $\ell\in\{1,..,L\}$, $\mathbf{\Psi_0}^{<\ell,<\ell}$ and $\mathbf{\Psi_0}^{>\ell,>\ell}$ 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.

2.4.2 Asymmetric impulse response functions

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

$$\begin{aligned} \mathbf{\Psi}_k(\boldsymbol{\varepsilon_{t-k}}) &= \mathbf{\Psi}_k(\varepsilon_{\ell,t-k}) && (9) \\ &= \left[\mathbf{\Psi}_k^{+}1_{\varepsilon_{\ell,t-k}>0} + \mathbf{\Psi}_k^{-}1_{\varepsilon_{\ell,t-k}<0}\right], \;\; \forall k\in\{0,..,K\},\; \forall t\in\{1,..,T\} && (10) \end{aligned}$$

with $\ell\in\{1,..,L\}$, $\varepsilon_{\ell,t}$, 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.

Proof. We first establish the following lemma:

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

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

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* $\mathbf{\Gamma}$ *satisfies*

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

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

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

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

$$\mathbf{u}_t = \mathbf{\Psi_0}(\varepsilon_{\ell,t})\boldsymbol{\varepsilon_t} \tag{11}$$

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

Notice that (11) implies the sub-system with ℓ equations

$$\mathbf{u}_t^{\leq \ell} = \begin{pmatrix} \mathbf{\Psi}_0^{<\ell,<\ell} & \mathbf{0}^{<\ell,1} \\ \mathbf{\Psi}_0^{\ell,<\ell} & \mathbf{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t}) \end{pmatrix} \boldsymbol{\varepsilon}_{\boldsymbol{t}}^{\leq \ell} \tag{12}$$

and notice that the matrix in (12) depends on $\varepsilon_{\ell,t}$ only through the scalar $\mathbf{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t})$. Denoting $\mathbf{A} \equiv \mathbf{\Psi_0}^{<\ell,<\ell}$ a $(\ell-1)\times(\ell-1)$ invertible matrix (from Assumption 1), $\mathbf{C} \equiv \mathbf{\Psi_0}^{\ell,<\ell}$ a $1 \times (\ell - 1)$ matrix, $\mathbf{B} \equiv \mathbf{0}$ of dimension $(\ell - 1) \times 1$, and $D(\varepsilon_{\ell,t}) \equiv \mathbf{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t})$ the (ℓ, ℓ) coefficient of $\mathbf{\Psi}_0$ (a scalar), we can use Lemma 1 to invert the system (12) and obtain

$$\boldsymbol{\varepsilon}_{\boldsymbol{t}}^{\leq \ell} = \frac{1}{D(\varepsilon_{\ell,t})} \begin{pmatrix} D(\varepsilon_{\ell,t})\mathbf{A}^{-1} & \mathbf{0}^{<\ell,1} \\ -\mathbf{C}\mathbf{A}^{-1} & 1 \end{pmatrix} \mathbf{u}_t^{\leq \ell}. \tag{13}$$

The last row of (13) 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 $D(\varepsilon_{\ell,t}) = \mathbf{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t})$. But since $\mathbf{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t})$, 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 (13). Then, with $\mathbf{\Psi_0}^{<\ell,<\ell}$ and $\mathbf{\Psi_0}^{>\ell,>\ell}$ invertible, (11) has a unique solution vector

$\boldsymbol{\varepsilon_t}$. ■

Proposition 1 ensures that the system (7) 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 $\boldsymbol{\varepsilon_t}$.[9]

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 (7) even when all shocks have asymmetric effects.

2.4.3 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 $\underline{z} = \min\limits_{t\in[1,T]}(z_t)$ and $\overline{z} = \max\limits_{t\in[1,T]}(z_t)$. The following proposition establishes the condition under which system (7) has a unique solution even when the identified shock $\varepsilon_{\ell,t}$ has asymmetric and state dependent effects.

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

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

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 $\boldsymbol{\Psi}_0$ 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 provided that $sgn\left(\boldsymbol{\Psi}_{0,\ell\ell}^+(z_t)\right) = sgn\left(\boldsymbol{\Psi}_{0,\ell\ell}^-(z_t)\right) > 0$, $\forall z_t \in [\underline{z}, \overline{z}]$.

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

[9] 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.

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.

3 Additional empirical results

3.1 Government spending over time

Figure A1 plots the time series of government spending to potential output, calculated following Gordon and Krenn (2010) and as described in the main text. We can see that the series displays non-trivial secular trends, not only over historical times (left panel) but also over the more recent period (right panel).

3.2 Baseline results: the cumulative ("sum") multipliers at different horizons

In the main text, we only report the cumulative multiplier for $h = 20$ quarters (or 5 years) in order to capture the overall effect of a government spending shock. To trace the size of the multiplier from the short-run to the longer-run, Table A1 reports our multiplier estimates from $h = 1$ to 20 for the recursive identification scheme (Auerbach and Gorodnichenko, 2012, "AG shocks") over 1966-2014 (corresponding to figure 2 in the main text) and for the narrative identification scheme (Ramey, 2011, "Ramey shocks") over 1939-2014 (corresponding to figure 3 in the main text). We can see that the contractionary multiplier (m^-) is substantially larger than the expansionary multiplier (m^+) at all horizons.

3.3 Using Local Projections over 1947-2014, no time trend

We extended the robustness section 4.1 of the main text, and ran the same LP-IV as in equation (10) of the main text except that we do not include any polynomial trend as control. Figures A2 and A3 report the corresponding impulse responses of government spending (G), private spending ($Y - G$), investment (I) and consumption (C) to government spending shocks, identified either recursively or narratively. The impulse responses are very similar to the ones presented in the main text. Private spending co-moves positively with government spending following a contractionary G shock (i.e., the multiplier is above one), but co-moves negatives

Table A1: Cumulative ("sum") multiplier at horizon h

	AG shocks		Ramey News shocks	
horizon	m^+	m^-	m^+	m^-
1	0.77	1.46	0.48	1.42
2	0.71	1.46	0.48	1.30
3	0.65	1.46	0.47	1.21
4	0.60	1.45	0.47	1.14
5	0.55	1.42	0.46	1.09
6	0.51	1.42	0.46	1.05
7	0.47	1.40	0.46	1.03
8	0.43	1.40	0.46	1.02
9	0.40	1.39	0.46	1.03
10	0.37	1.38	0.46	1.04
11	0.35	1.37	0.46	1.07
12	0.33	1.34	0.46	1.11
13	0.32	1.33	0.47	1.16
14	0.31	1.32	0.47	1.21
15	0.30	1.31	0.47	1.27
16	0.29	1.30	0.47	1.33
17	0.29	1.28	0.47	1.39
18	0.28	1.27	0.48	1.44
19	0.27	1.26	0.48	1.50
20	0.27	1.24	0.48	1.54

Note: m^- refers to the contractionary multiplier and m^+ refers to the expansionary multiplier. 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, 1966-2014. Ramey news shocks are the unexpected changes in anticipated future expenditures constructed by Ramey (2011), 1939-2014.

with government spending following a contractionary G shock (i.e., the multiplier is below one).

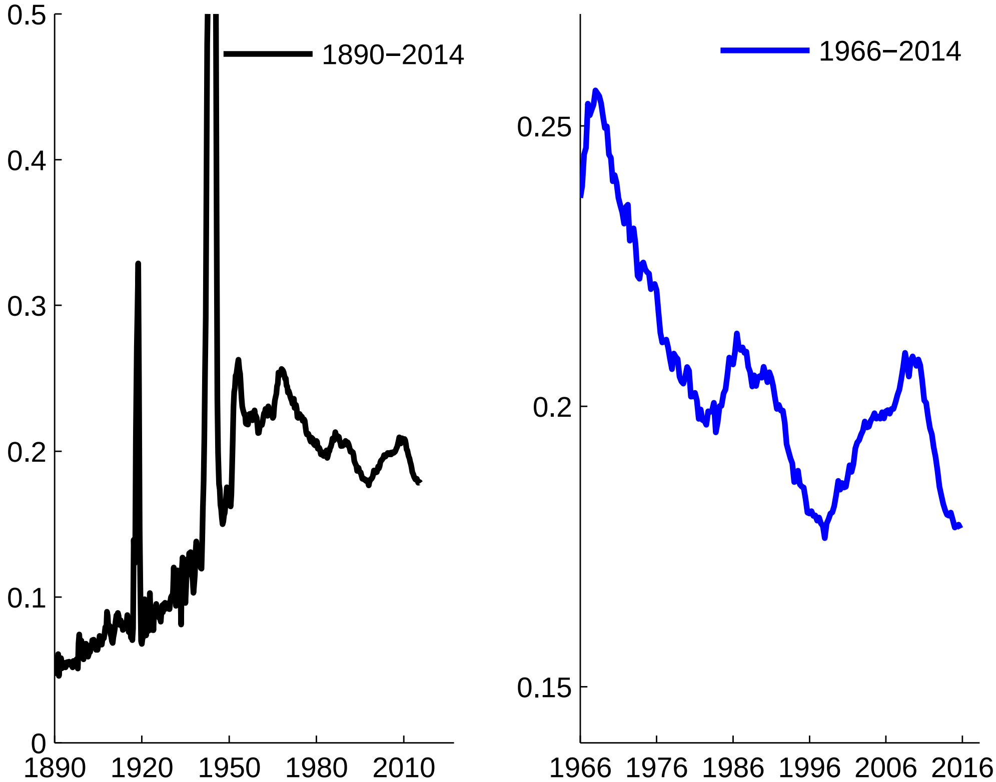

Figure A1: Government spending to potential output over 1890-2014 (left-panel) and 1966-2014 (right-panel)

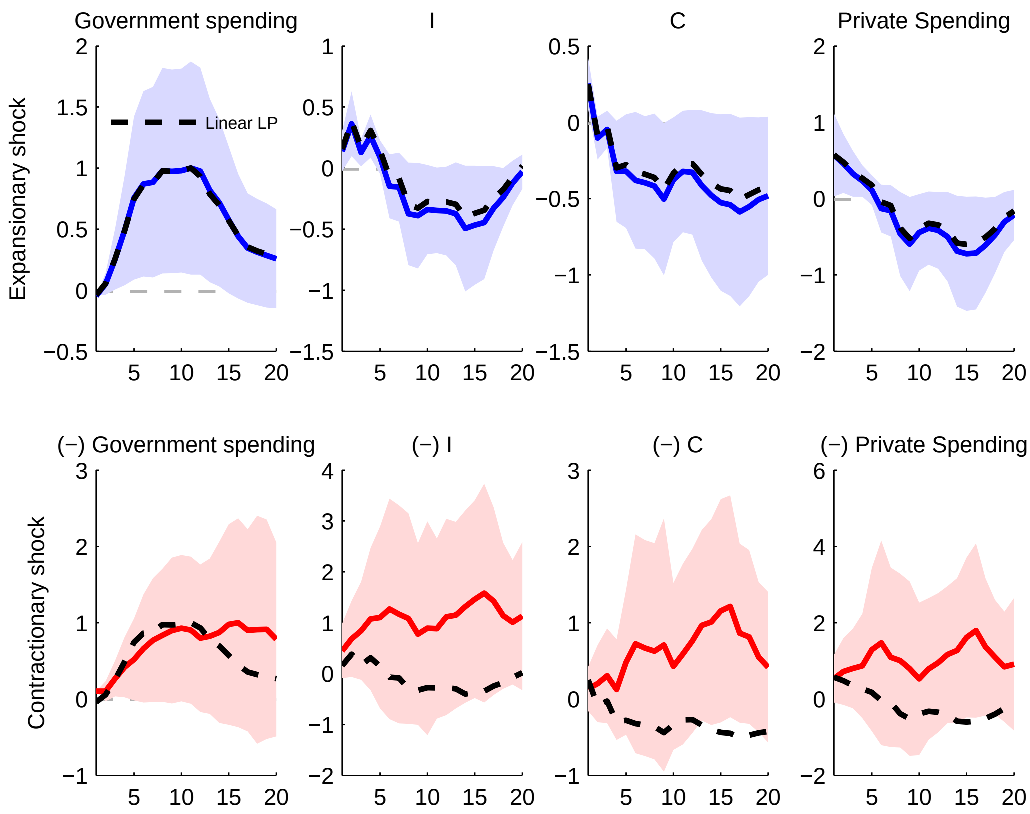

Figure A2: **Narrative identification scheme, Local Projections without a polynomial trend, 1947-2014**. Impulse response functions (in percent) of Government spending, Investment (I), and Consumption (C) following an expansionary Ramey news shock (left panel) and a contractionary Ramey news shock (right panel). Estimates from Local Projections. The dashed lines report the point estimates from a linear (i.e., symmetric) model. For ease of comparison between the left and right panels, the responses to a contractionary shock are multiplied by -1 in the right panels. The shaded areas are the 90 percent confidence bands calculated using Newey-West standard errors.

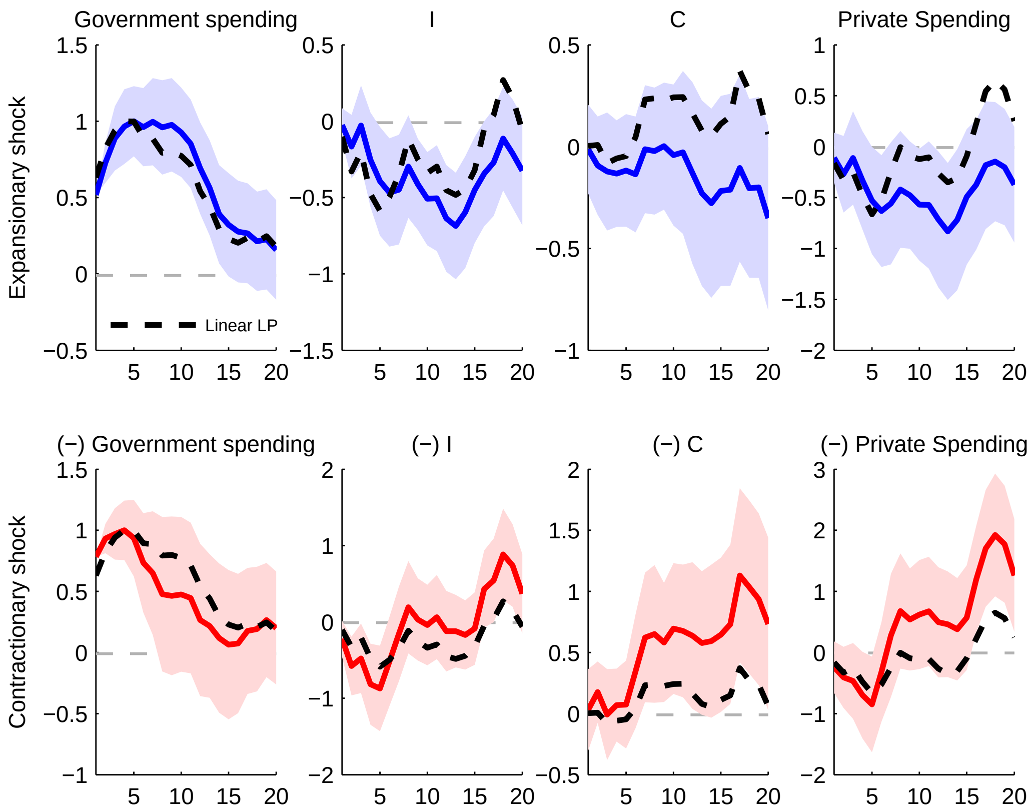

Figure A3: **Recursive identification scheme, Local Projections without a polynomial trend, 1947-2014**. Impulse response functions (in percent) of Government spending, Investment (I) and Consumption (C) following an expansionary government spending shock (left panel) and a contractionary shock (right panel). The dashed lines report the point estimates from a linear (i.e., symmetric) model. For ease of comparison between the left and right panels, the responses to a contractionary shock are multiplied by -1 in the right panels. The shaded areas are the 90 percent confidence bands calculated using Newey-West standard errors.