Functional Approximation of Impulse Responses

Regis Barnichon and Christian Matthes
March 2018

This online appendix presents a number of complementary results. Section 1 shows that any mean-reverting impulse response function can be approximated by a sum of Gaussian basis functions. Section 2 discusses the construction of the likelihood for FAIR models and Section 3 presents our estimation algorithm. Section 4 discusses the issue of identification in asymmetric vector moving-average models. Sections 5 to 7 present three different Monte-Carlo simulations to illustrate the workings of FAIR models as well as to evaluate their performances in finite sample. We evaluate the performances of FAIR in the linear case; first for a well-specified FAIR model and then for a mis-specified FAIR model, and we then evaluate the ability of FAIR models to detect asymmetric effects of shocks. Section 8 discusses the prior IRFs implied by our priors on the a-b-c parameters with a focus on the sign-restriction identification (section 3 of the main text); Section 9 clarifies the conceptual differences between a FAIR model and a (finite state) Markov-Switching (MS) model and discusses why a MS model cannot easily capture asymmetric impulse response functions.

1   Approximating IRFs with Gaussian basis functions

In this section, we show that any mean-reverting impulse response function can be approximated by a sum of Gaussian basis functions. The following Theorem is a simple extension of Alspach and Sorenson (1971, 1972).

Theorem 1. *Let f be a bounded continuous function on $\mathbb{R}$ that satisfies $\int_{-\infty}^{\infty} f(x)^2 dx < \infty$. There exists a function f_N defined by*

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

with a_n, b_n, $c_n \in \mathbb{R}$ for $n \in \mathbb{N}$, such that the sequence $\{f_N\}$ converges pointwise to f on every interval of $\mathbb{R}$.

Proof. Following Alspach and Sorenson (1971, 1972) in the context of approximating distributions, the problem of approximating a function f can be considered within the context of delta families of positive types.

Delta families are families of functions which converge to a delta function as a parameter characterizing the family converges to a limit value.

Let $\{\delta_\lambda\}$ be a family of functions on the interval $]-\infty, +\infty[$ which are integrable over every interval. $\{\delta_\lambda\}$ forms a delta family of positive type if the following conditions are satisfied:

1. For every constant $\gamma > 0$, δ_λ tends to zero uniformly for $\gamma \leq |x| \leq \infty$ as $\lambda \to \lambda_0$
2. There exist s in $\mathbb{R}$ so that $\int_{-s}^{s} \delta_\lambda(x) dx \longrightarrow 1$ as λ tends to some limit value λ_0
3. $\delta_\lambda(x) \geq 0$ for all x and λ

Defining

$$\delta_\lambda(x) \equiv G_\lambda(x) = \frac{1}{\sqrt{2\pi\lambda^2}} e^{-\frac{x^2}{\lambda^2}}, \tag{1}$$

it is easy to see that the Gaussian functions $\{G_\lambda\}$ form a delta family of positive type as $\lambda \to 0$ (i.e., $\lambda_0 = 0$). That is, the Gaussian function tends to the delta function as the variance tends to zero.[1]

We can then make use of the following theorem.

Theorem: The sequence $\{f_\lambda\}$ which is formed by the convolution of δ_λ and f

$$f_\lambda(x) = \int_{-\infty}^{+\infty} \delta_\lambda(x - u) f(u) du \tag{2}$$

converges uniformly to f as $\lambda \to \lambda_0$ for x on every interval $[x_0, x_1]$ of $\mathbb{R}$.

Proof. See Korevaar (1968). ∎

Using (1) in (2), the function f_λ given by

$$f_\lambda(x) = \int_{-\infty}^{+\infty} G_\lambda(x - u) f(u) du \tag{3}$$

converges uniformly to f as $\lambda \to 0$ for x in some arbitrary interval $[x_0, x_1]$ of $\mathbb{R}$.

Next, we want to approximate (3) with a Riemann sum. To do so, first rewrite f_λ as

$$f_\lambda(x) = \underbrace{\int_{-\infty}^{-s} G_\lambda(x - u) f(u) du}_{=A(\lambda, x)} + \int_{-s}^{+s} G_\lambda(x - u) f(u) du + \underbrace{\int_{s}^{+\infty} G_\lambda(x - u) f(u) du}_{=B(\lambda, x)} \tag{4}$$

---

[1]Note that this proof can be easily applied to other functions (such as the inverse quadratic function $x \to \frac{1}{1+\left(\frac{x}{\lambda}\right)^2}$) that form a delta family of a positive type, so that our approach is not restricted to Gaussian functions.

for $s > 1$.

Note that for any $s > 1$, we have

$$
\begin{aligned}
0 &\le \int_s^{+\infty} G_\lambda(u)du \\
&\le \frac{1}{\sqrt{2\pi\lambda^2}} \int_s^{+\infty} e^{-\frac{u}{\lambda^2}} du \text{ since } u^2 > u \text{ for any } u \text{ in } [s, +\infty[,\ s > 1 \\
&\le \left[\frac{-\lambda^2}{\sqrt{2\pi\lambda^2}} e^{-\frac{u}{\lambda^2}} \right]_s^{+\infty} = \frac{|\lambda|}{\sqrt{2\pi}} e^{-\frac{s}{\lambda^2}} \xrightarrow[\lambda \to 0]{} 0
\end{aligned}
$$

which shows that $\forall s > 1$, $\lim\limits_{\lambda \to 0} \int_s^{+\infty} G_\lambda(u)du = 0$. Symmetrically, we can show $\lim\limits_{\lambda \to 0} \int_{-\infty}^{-s} G_\lambda(u)du = 0$.

Going back to (4), we have

$$
0 \le |B(\lambda, x)| \le M \int_{-\infty}^{x-s} G_\lambda(t)dt
$$

where $M = \sup\limits_{x \in \mathbb{R}} |f(x)|$. Since $x \in [x_0, x_1]$, we can choose an $s > 1$ such that $x - s < -1$, so that we can apply the previous result and get

$$
\lim_{\lambda \to 0} |B(\lambda, x)| = 0. \tag{5}
$$

Proceeding symmetrically, we have $\lim\limits_{\lambda \to 0} |A(\lambda, x)| = 0$.

Finally, since the function $u \mapsto G_\lambda(x-u)f(u)$ is continuous over $[-s, s]$, we can approximate $\int_{-s}^{+s} G_\lambda(x-u)f(u)du$ with a Riemann sum. Denoting

$$
f_{\lambda,N}(x) = \sum_{n=1}^{N} G_\lambda(x - \xi_n) f(\xi_n) (\xi_n - \xi_{n-1})
$$

where $\xi_n = -s + n\frac{2s}{N}$, we get that

$$
\lim_{N \to \infty} f_{\lambda,N}(x) = \int_{-s}^{+s} G_\lambda(x - u)f(u)du. \tag{6}
$$

Denoting $a_n = f(\xi_n)(\xi_n - \xi_{n-1})$, $b_n = \xi_n$ and $c_n = \lambda$, using (6), (5) in (4) and combining with (3), we get that

$$
\lim_{\lambda \to 0} \left(\lim_{N \to \infty} f_{\lambda,N}(x) \right) = f(x)
$$

which completes the proof. ■

2 FAIR estimation

We now discuss how to estimate FAIR models using maximum likelihood or Bayesian methods. While the computational cost is not as trivial as OLS in the case of VARs, the estimation is simple and relatively easy thanks to modern computational capabilities.

The key is to construct the likelihood function $p(\boldsymbol{y^T}|\boldsymbol{\theta})$ of a sample of size T for a linear moving-average model with parameter vector $\boldsymbol{\theta}$ and where a variable with a superscript denotes the sample of that variable up to the date in the superscript. To start, we use the prediction error decomposition to break up the density $p(\boldsymbol{y^T}|\boldsymbol{\theta})$ as follows:[2]

$$p(\boldsymbol{y^T}|\boldsymbol{\theta}) = \prod_{t=1}^{T} p(\boldsymbol{y_t}|\boldsymbol{\theta}, \boldsymbol{y^{t-1}}). \tag{7}$$

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,[3] and we note that the density $p(\boldsymbol{y_t}|\boldsymbol{\theta}, \boldsymbol{y^{t-1}})$ 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

$$\boldsymbol{y_t} = \boldsymbol{\Psi}_0\boldsymbol{\varepsilon_t} + \sum_{h=1}^{H} \boldsymbol{\Psi}_h\boldsymbol{\varepsilon_{t-h}}. \tag{8}$$

Since the contemporaneous impact matrix 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 values of the components of $\boldsymbol{\varepsilon_t}$ from equation (8), that is we need that $\boldsymbol{\Psi}_0$ is invertible. We impose this restriction by assigning a minus infinity value to the likelihood whenever $\boldsymbol{\Psi}_0$ is not invertible. It is also at this stage that we impose the identifying restriction that we describe next. Finally, to initialize the recursion, we set the first H innovations $\{\boldsymbol{\varepsilon}_j\}_{j=-H}^{0}$ to zero.[4],[5]

3 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

---

[2]To derive the conditional densities in decomposition (7), our parameter vector $\boldsymbol{\theta}$ thus implicitly also includes the H initial values of the shocks: $\{\boldsymbol{\varepsilon}_{-H}...\boldsymbol{\varepsilon}_0\}$. We will keep those fixed throughout the estimation and discuss alternative initializations below.

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

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

[5]When H, the lag length of the moving average model is infinite, we truncate the model at some horizon H, large enough to ensure that the coefficients of the lag matrix $\boldsymbol{\Psi}_H$ are "close" to zero. Such a H exists since the variables are stationary.

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 $p(y^T|\theta)p(\theta)$. 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

- 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

Computational cost: To give a sense of the computational cost of our approach, on a 2016 MacBook Pro (3.3 GHz Intel Core i7 with 16 GB of RAM) 100 likelihood evaluations of a linear FAIR model with 1 basis function, 3 observables and 40 lags for 120 observations takes 1.98 seconds. A full estimation run (with initial optimization etc.) using the same data and model as well as hardware takes 36 minutes (with 20000 draws to tune the Metropolis-Hastings steps, 20000 draws from the posterior and 4 blocks in the Metropolis-Hastings algorithm).

5

4 Identifying restrictions in non-linear VMA models

We now discuss how to impose identifying restrictions in VMA models, notably non-linear VMA models with asymmetric impulse responses. We only discuss the non-linear model $\boldsymbol{y_t} = \sum_{h=0}^{\infty} \boldsymbol{\Psi}_h(\boldsymbol{\varepsilon_{t-h}},)\boldsymbol{\varepsilon_{t-h}}$, since it includes the simpler linear model $\boldsymbol{y_t} = \sum_{h=0}^{\infty} \boldsymbol{\Psi}_h \boldsymbol{\varepsilon_{t-h}}$.

As described in the main text, we impose the identifying restriction when we construct the likelihood, so that constructing the likelihood and imposing identifying restrictions are intimately linked, and we thus describe them jointly. 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. As described in section 2 of this appendix, in order to construct the likelihood recursively, the system of equations

$$\boldsymbol{\Psi}_0(\boldsymbol{\varepsilon_t})\boldsymbol{\varepsilon_t} = \mathbf{u}_t \tag{9}$$

need to have a unique solution vector $\boldsymbol{\varepsilon_t}$ given $\mathbf{u}_t = \boldsymbol{y_t} - \sum_{h=0}^{H} \boldsymbol{\Psi}_h(\boldsymbol{\varepsilon_{t-h}})\boldsymbol{\varepsilon_{t-1-h}}$. That is, we must ensure that there is a one-to-one mapping from $\boldsymbol{\varepsilon_t}$ to $\boldsymbol{\Psi}_0(\boldsymbol{\varepsilon_t})\boldsymbol{\varepsilon_t}$. 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.[6]

Consider an asymmetric model where $\boldsymbol{\Psi}_h$ 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 $\boldsymbol{\varepsilon_t}$.

We now show how to address the issue when we allow the identified shocks to have asymmetric effects on the impulse response functions. We successively consider each identification scheme used in the paper: (i) recursive ordering, (ii) narrative identification, and (iii) sign restrictions.

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 $\boldsymbol{y_t^{>\ell}}$.

---

[6]Note that if the impact matrix $\boldsymbol{\Psi}_0$ is a constant and does not depend on $\boldsymbol{\varepsilon_t}$ (so that $\boldsymbol{\Psi}_h$ depends on $\boldsymbol{\varepsilon_t}$ only for $h > 0$), 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 $\mathbf{u}_t$.

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

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

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

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

with $\ell \in \{1,..,L\}$, $\boldsymbol{\Psi_0}^{<\ell,<\ell}$ and $\boldsymbol{\Psi_0}^{>\ell,>\ell}$ 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 $\boldsymbol{\varepsilon_t}$. For instance, in the monetary model used in section 5, the policy rate is ordered last, and the recursive identification scheme states that shocks to the policy rate do not affect unemployment and inflation contemporaneously, i.e., that the last column of $\boldsymbol{\Psi_0}$ is filled with zeros except for the diagonal element.

We consider a model with asymmetry in response to the structural shock $\{\varepsilon_{\ell,t}\}_{\ell \in \{1,..,L\}}$ and we establish the following proposition:

Proposition 1. *Consider the non-linear moving average model defined in section 4 of the main text with*

$$
\begin{aligned}
\boldsymbol{\Psi}_h(\boldsymbol{\varepsilon_{t-h}}) &= \boldsymbol{\Psi}_h(\varepsilon_{\ell,t-h}) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad (10) \\
&= \left[\boldsymbol{\Psi}_h^+ 1_{\varepsilon_{\ell,t-h}>0} + \boldsymbol{\Psi}_h^- 1_{\varepsilon_{\ell,t-h}<0} \right], \ \ \forall h \in \{0,..,H\}, \ \forall t \in \{1,..,T\} \quad (11)
\end{aligned}
$$

with $\ell \in \{1,..,L\}$, $\varepsilon_{\ell,t}$, 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 H initial values of the shocks $\{\boldsymbol{\varepsilon}_{-H}...\boldsymbol{\varepsilon}_0\}$, 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*

$$\boldsymbol{\Gamma} = \left(\begin{array}{cc} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{array} \right)$$

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* $\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}$ *nonsingular. Then, the inverse of* $\boldsymbol{\Gamma}$ *satisfies*

$$\boldsymbol{\Gamma}^{-1} = \left(\begin{array}{cc} \mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{BF}^{-1}\mathbf{CA}^{-1} & -\mathbf{A}^{-1}\mathbf{BF}^{-1} \\ -\mathbf{F}^{-1}\mathbf{CA}^{-1} & \mathbf{F}^{-1} \end{array} \right)$$

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

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

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

$$\mathbf{u}_t = \boldsymbol{\Psi_0}(\varepsilon_{\ell,t})\boldsymbol{\varepsilon_t} \tag{12}$$

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

Notice that (12) implies the sub-system with ℓ equations

$$\mathbf{u}_t^{\leq \ell} = \left(\begin{array}{cc} \boldsymbol{\Psi}_0^{<\ell,<\ell} & \mathbf{0}^{<\ell,1} \\ \boldsymbol{\Psi}_0^{\ell,<\ell} & \boldsymbol{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t}) \end{array} \right) \varepsilon_t^{\leq \ell} \tag{13}$$

and notice that the matrix in (13) depends on $\varepsilon_{\ell,t}$ only through the scalar $\boldsymbol{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t})$. Denoting $\mathbf{A} \equiv \boldsymbol{\Psi}_0^{<\ell,<\ell}$ a $(\ell-1)\times(\ell-1)$ invertible matrix (from Assumption 1), $\mathbf{C} \equiv \boldsymbol{\Psi}_0^{\ell,<\ell}$ a $1 \times (\ell-1)$ matrix, $\mathbf{B} \equiv \mathbf{0}$ of dimension $(\ell-1) \times 1$, and $D(\varepsilon_{\ell,t}) \equiv \boldsymbol{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t})$ the (ℓ,ℓ) coefficient of $\boldsymbol{\Psi}_0$ (a scalar), we can use Lemma 1 to invert the system (13) and obtain

$$\varepsilon_t^{\leq \ell} = \frac{1}{D(\varepsilon_{\ell,t})} \left(\begin{array}{cc} D(\varepsilon_{\ell,t})\mathbf{A}^{-1} & \mathbf{0}^{<\ell,1} \\ -\mathbf{CA}^{-1} & 1 \end{array} \right) \mathbf{u}_t^{\leq \ell}. \tag{14}$$

The last row of (14) provides the equation $\varepsilon_{\ell,t} = \frac{1}{D(\varepsilon_{\ell,t})}(\begin{array}{cc} -\mathbf{CA}^{-1} & 1 \end{array})\mathbf{u}_t$, 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

8

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 (14). Then, with $\mathbf{\Psi}_0^{<\ell,<\ell}$ and $\mathbf{\Psi}_0^{>\ell,>\ell}$ invertible, (12) has a unique solution vector $\boldsymbol{\varepsilon_t}$. ∎

Proposition 1 ensures that the system (9) 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 $\mathbf{\Psi}_0$ a function of $\boldsymbol{\varepsilon_t}$, we have

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

where J_t is the Jacobian of the (one-to-one) mapping from $\boldsymbol{\varepsilon_t}$ to $\mathbf{\Psi}_0(\boldsymbol{\varepsilon_t})\boldsymbol{\varepsilon_t}$ and where $p(\boldsymbol{\varepsilon_t})$ is the density of $\boldsymbol{\varepsilon_t}$.[7]

Finally, note that while we considered the case of a partially identified model, we can proceed similarly for a fully identified model with $\mathbf{\Psi}_0$ lower triangular and show that the shock vector $\boldsymbol{\varepsilon_t}$ is uniquely determined by (9) even when all shocks have asymmetric effects.

4.2 Narrative identification scheme

For a narrative identification scheme, we can use the previous results on recursive identification, since the use of narratively identified shocks can be cast as a partial identification scheme.

Indeed, if one orders the narratively identified shocks series first in $\boldsymbol{y_t}$, we can assume that $\mathbf{\Psi}_0$ has its first row filled with 0 except for the diagonal coefficient, which implies that the narratively identified shock does not react contemporaneously to other shocks (as should be the case if the narrative shocks were correctly identified). With Assumption 1 satisfied with $\ell = 1$, Proposition 1 then imply that (9) has a unique solution vector $\boldsymbol{\varepsilon_t}$ even when the narratively identified shocks have asymmetric effects.

4.3 Set identification from sign restrictions

We now consider the case of a set identification scheme based on sign restrictions. Denote ε_t^r the structural shock of interest identified from sign restrictions.

---

[7]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.

We establish the conditions under which system (9) has a unique solution vector in a model with asymmetry:

Proposition 2. *Consider the asymmetric moving average model defined in section 4 of the main text with*

$$\boldsymbol{\Psi}_h(\boldsymbol{\varepsilon_{t-h}}) = \boldsymbol{\Psi}_h(\varepsilon_{t-h}^r) \tag{15}$$

$$= \left[\boldsymbol{\Psi}_h^+ 1_{\varepsilon_{t-h}^r > 0} + \boldsymbol{\Psi}_h^- 1_{\varepsilon_{t-h}^r < 0} \right], \quad \forall h \in \{0, .., H\}, \ \forall t \in \{1, .., T\} \tag{16}$$

with ε_t^r the structural shock identified from sign restrictions. Then, given $\{\boldsymbol{y_t}\}_{t=1}^T$, given the model parameters and given H initial values of the shocks $\{\boldsymbol{\varepsilon_{-H}}...\boldsymbol{\varepsilon_0}\}$, the series of shocks $\{\boldsymbol{\varepsilon_t}\}_{t=1}^T$ is uniquely determined provided that $sgn(\det \boldsymbol{\Psi}_0^+) = sgn(\det \boldsymbol{\Psi}_0^-)$.

Proof. Without loss of generality, let us order the variables such that ε_t^r, the shock with asymmetric effects, is ordered last. We can then write $\boldsymbol{\Psi}_0(\varepsilon_t^r)$ (of dimension $L \times L$) as

$$\boldsymbol{\Psi}_0(\varepsilon_t) = \begin{pmatrix} \mathbf{A} & \mathbf{B}(\varepsilon_t^r) \\ \mathbf{C} & D(\varepsilon_t^r) \end{pmatrix}$$

with $\mathbf{A}$ a $(L-1) \times (L-1)$ invertible matrix, $\mathbf{C}$ a $1 \times (L-1)$ matrix, $\mathbf{B}(\varepsilon_t^r)$ a matrix of dimension $(L-1) \times 1$ that depends on ε_t^r, and $D(\varepsilon_t^r) \equiv \boldsymbol{\Psi}_{0,LL}(\varepsilon_t^r)$ a scalar. Notice that only the last column of $\boldsymbol{\Psi}_0$ depends on ε_t^r.

We will make use of the following lemma:

Lemma 2. *Consider the same matrix Γ as in Lemma 1. We have*

$$\det \boldsymbol{\Gamma} = \det(\mathbf{A}) \det(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}).$$

Proof. Rewrite Γ as

$$\boldsymbol{\Gamma} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{pmatrix}$$

and the lemma follows. ∎

Using Lemma 1 and noting that $D(\varepsilon_t^r)$ is a scalar, we have that the inverse of $\boldsymbol{\Psi}_0$ satisfies

$$\boldsymbol{\Psi}_0^{-1} = \frac{1}{D(\varepsilon_t^r) - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}(\varepsilon_t^r)} \begin{pmatrix} \left(D(\varepsilon_t^r) - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}(\varepsilon_t^r)\right)\mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\varepsilon_t^r) \\ -\mathbf{C}\mathbf{A}^{-1} & 1 \end{pmatrix}.$$

The last row of the system $\boldsymbol{\varepsilon_t} = \boldsymbol{\Psi}_0^{-1}\mathbf{u}_t$ provides the equation

$$\varepsilon_t^r = \frac{1}{D(\varepsilon_t^r) - \mathbf{CA}^{-1}\mathbf{B}(\varepsilon_t^r)}(\ -\mathbf{CA}^{-1}\ \ 1\ )\mathbf{u}_t, \qquad (17)$$

which defines ε_t^r. Since the right hand side of (17) only depends on ε_t^r through $D(\varepsilon_t^r) - \mathbf{CA}^{-1}\mathbf{B}(\varepsilon_t^r)$, the sign of the right hand side of (17) depends on ε_t^r only through the sign of $D(\varepsilon_t^r) - \mathbf{CA}^{-1}\mathbf{B}(\varepsilon_t^r)$.[8] Using Lemma 2, we have $\det \boldsymbol{\Psi}_0(\varepsilon_t^r) = \left(D(\varepsilon_t^r) - \mathbf{CA}^{-1}\mathbf{B}(\varepsilon_t^r)\right)\det \mathbf{A}$, so that the sign of the right hand side of (17) depends on ε_t^r only through the sign of $\det \boldsymbol{\Psi}_0(\varepsilon_t^r)$. Thus, with $sgn(\det \boldsymbol{\Psi}_0^+) = sgn(\det \boldsymbol{\Psi}_0^-)$, the sign of ε_t^r is uniquely pinned down, so that with $\mathbf{A}$ invertible, the system $\mathbf{u}_t = \boldsymbol{\Psi_0}(\varepsilon_t^r)\boldsymbol{\varepsilon_t}$ has a unique solution vector. ∎

Proposition 2 states that the system $\mathbf{u}_t = \boldsymbol{\Psi_0}(\varepsilon_t^r)\boldsymbol{\varepsilon_t}$ determines a unique solution vector $\boldsymbol{\varepsilon_t}$ as long as $sgn\left(\det \boldsymbol{\Psi}_0^+\right) = sgn\left(\det \boldsymbol{\Psi}_0^-\right)$, i.e., as long as $sgn\left(D^+ - \mathbf{CA}^{-1}\mathbf{B}^+)\right) = sgn\left(D^- - \mathbf{CA}^{-1}\mathbf{B}^-)\right)$, which means that there is a unique solution vector $\boldsymbol{\varepsilon_t}$ as long as the asymmetry on the impact coefficients is not too strong.

In practice, we impose this restriction by assigning a minus infinity value to the likelihood whenever $sgn(\det \boldsymbol{\Psi}_0^+) \neq sgn(\det \boldsymbol{\Psi}_0^-)$. Then, to construct the likelihood, we proceed as described in the recursive identification section by using the fact that there is a one-to-one mapping from $\boldsymbol{\varepsilon_t}$ to $\boldsymbol{\Psi}_0(\boldsymbol{\varepsilon_t})\boldsymbol{\varepsilon_t}$.

5   Simulation study #1: a well-specified linear FAIR model

In this section we describe our first simulation study, where we consider the case of a linear FAIR(1) data generating process (DGP), so that the FAIR estimation is carried out with a correctly specified model. We first describe the DGP, and then assess (i) the finite performances of FAIR against a VAR, (ii) the convergence property of our MCMC, (iii) whether FAIR models are well identified, and (iv) to what extent the initialization of the initial shocks to zero matters for inference.

5.1   The Data Generating Process

The true DGP is a trivariate $\text{FAIR}_{\mathcal{G}}(1)$ model with 40 lags. The impulse responses (which completely characterize the DGP) are given in Figure A1.

With this DGP, we try to capture a number of different patterns of impulse responses regularly observed in linear analysis, while at the same time trying to keep the DGP simple via the use of one basis function only. We simulate 50 data sets of 120 observations each - we

---

[8]In fact, we have $D(\varepsilon_t^r) - \mathbf{CA}^{-1}\mathbf{B}(\varepsilon_t^r) = \boldsymbol{\Psi}_{0,LL}(\varepsilon_t^r) - \sum_{\ell=1}^{L-1}\left(\mathbf{CA}^{-1}\right)_\ell \boldsymbol{\Psi}_{0,\ell L}(\varepsilon_t^r).$

have in mind the DGP mimicking a monthly model with 10 years of data. The data-generating process imposes a recursive identification scheme (the matrix of initial impacts of the shock is lower triangular), which we also impose in our estimation. Beyond this restriction we do not impose *any* priors to focus solely on the in the likelihood function and highlight that FAIR parameters can be well estimated without informative priors.

For estimation, we use 20000 draws for the FAIR(1) model (20000 additional draws for burn-in/fine-tuning the proposal densities), and we only use 1 block in the Metropolis-Hastings algorithm.

5.2   FAIR performances and comparison with a VAR

To assess the performances of FAIR, we compare our results with those of a Gaussian VAR(12). We use very loose (conjugate Normal-Wishart) priors for the VAR, so that just as in the case of the FAIR model, all relevant information comes from the likelihood function. Note that as long as the VAR can capture the lagged dynamics of the data (we use large number of lags in the VAR exactly to give it a fair shot to match impulse responses, echoing results in De Graeve & Westermark (2013)), it will be able to capture the true impulse responses via a simple Cholesky identification scheme.

For each of our 50 Monte Carlo samples, we compute the mean squared error for all impulse responses (3 shocks and 3 observables for a total of 9 impulse response paths) for the first 25 time periods. The average mean squared error (across responses and Monte Carlo replications) is 150 percent higher for the VAR. This result holds across *all* Monte Carlo samples as well, with the 25th percentile of the increase in mean squared errors across replications being already 65 percent.

While we use a VAR with relatively many lags to give the VAR a good chance to match impulse responses, a natural response to our results might be that they are driven by the VAR lag length. To rule this out, we also estimated VAR(4) models on our 50 datasets. The same pattern of results emerges - the mean squared error across responses and Monte Carlo samples is 106 percent higher in the VAR case than with a FAIR(1) and the 25th percentile of the increase in mean squared errors across replications is 35 percent. Looking at the impulse responses over the first 40 time periods yields an even stronger picture: The average mean squared error of a VAR(12) relative to a FAIR model is 284 percent. The 25th percentile of the increase in mean squared errors across replications is 146 percent in that case.

To illustrate the range of impulse response estimates for the VAR(12) and for FAIR(1), figures A2 and A3 plot the true responses along with the estimated responses across our 50 Monte Carlo samples. Note that the y-scale is substantially smaller for FAIR estimates than for the VAR, so to make comparison easier, figure A4 plots the FAIR impulse responses *on*

the same scale as for the VAR. We can see that VAR estimates display substantially more variability than FAIR estimates. First, VAR estimates are more noisy while FAIR estimates are smooth by construction. Second, and more importantly, VAR estimates are very variable across replications, and the VAR-estimated IRF can show large deviations from the true IRF (e.g., the response of variable 3 to shock 1 in the lower-left corner). Intuitively, when estimating IRFs from a finite-order VAR, researchers faces a difficult bias-variance trade-off, balancing between the need to estimate a high-enough order VAR and the need to keep the number of free parameters small. In situations where the order of the VAR must be large to guarantee a good approximation of the DGP (as in this Monte Carlo simulation), a FAIR$_\mathcal{G}$ model can provide a useful alternative to estimate IRFs. Third, because VAR-based IRF are basically linear combinations of damped sine-cosine functions, the VAR-based IRFs can display counter-factual oscillations.[9] With its tight parametrization, a FAIR$_\mathcal{G}$ with only a few basis functions avoids this problem.

Importantly, our goal was not to claim that FAIR models are superior to VARs. Instead, this simulation is meant to convey that FAIR models can provide a useful alternative approach, especially in short samples.

5.3 Convergence

To get a sense of how well our MCMC algorithm mixes, we (randomly) focus on 15 of the 50 data sets and run a second Markov chain on them. For simplicity, we only focus on the estimated impulse responses *on impact* that are not restricted to be zero. Figure A5 shows the trace plots for 2000 consecutive (post burn-in) draws from both chains and one dataset. The straight lines show the averages for the 2000 draws. We can see that relative to the estimated posterior means displayed in those 2000 draws, the difference in means is small. This conclusion carries over to all 15 datasets. If we compute the difference between the estimated posterior means relative to the posterior mean for the original chain, the median difference (across all impulse responses and datasets) is less than 5 percent of the estimated posterior mean of the original chain.

5.4 Identification

To assess whether FAIR models are well identified, we plot slices of the likelihood surface at the true values (holding all other parameters constant at their true values. We do this for two random samples, one of size 120 (as in our MC simulations, shown in Figure A6) and one with

---

[9]Going back to our second point, the order of the VAR may need to be very high to avoid these oscillations. However, this may be difficult in practice, because the number of parameters to estimate grows exponentially with the number of lags.

1200 observations (shown in Figure A7). What we can see is that (at least locally, which is all we can learn from this exercise) the FAIR model is well identified even in small samples.

5.5 Shock Initialization

Finally, to assess whether setting the initial shocks to 0 (when we construct the likelihood) matters for inference, we have conducted an experiment where we simulated data from our tri-variate $\text{FAIR}_{\mathcal{G}}(1)$ DGP, generating 20 samples of 120 observations each. For each sample, we estimated two versions of the model: one where the true initial shocks (set to 0) were used, and one where the initial shocks were set to a series of standard normal random variables. For illustration, we focus on the peak effects to the third shock in our trivariate FAIR (the results for the other shocks are similar). Figure A8 shows the estimated median peak effects across samples for the two specifications. The true value for the first two peak effects is 1, while the true value for the third response is 2. For all samples and specifications, the estimated mean values fluctuate around the true value, with no discernible difference in the average performance across the two specifications

6 Simulation study #2: A mis-specified linear FAIR model

We now consider a simulation in which the true DGP is a VAR(4), so that the $\text{FAIR}_{\mathcal{G}}$ is mis-specified. The idea of the exercise is to show that a parsimonious FAIR model that only *approximates* the true DGP can be helpful.

6.1 The Data Generating Process

To construct a plausible VAR-based DGP, we proceed as follows. We first estimate a structural VAR on US data (using a recursive identification scheme), invert it to obtain a set of impulse responses $\left\{ \hat{\boldsymbol{\Psi}}_h \right\}_{h=0}^{\infty}$, and we modify these baseline impulse responses to introduce asymmetry. From these impulse responses, we generate simulated data from

$$\boldsymbol{y_t} = \sum_{h=0}^{\infty} \hat{\boldsymbol{\Psi}}_h \boldsymbol{\varepsilon_{t-h}} \tag{18}$$

with $\boldsymbol{\varepsilon_t}$ Normally distributed, $E\boldsymbol{\varepsilon_t} = 0$ and $E\boldsymbol{\varepsilon_t}\boldsymbol{\varepsilon_t'} = \mathbf{I}$.

We use 50 Monte-Carlo replications with a sample size $T = 200$, which roughly corresponds to the sample size available for the US.

The DGP is obtained from estimating the quarterly VAR(4) considered in the main text with the unemployment rate, the PCE inflation rate and the federal funds rate over 1959-2007.

The impulse response functions to a monetary shock can be seen in Figure 1 of the main text.

6.2   FAIR performances and comparison with a VAR

For each simulated dataset, we estimate (i) a FAIR model with two Gaussian basis functions (with priors defined exactly as in the main text), and (ii) a (well-specified) VAR(4), and we evaluate the Mean-Square Error (MSE) of the estimated impulse response function over the horizons $h = 1...25$.[10] Importantly, we stack the odds in favor of the VAR and against the FAIR model, because the estimated VAR is a correctly specified model.

The first row of Table A1 presents the average MSEs over the simulations. For unemployment and inflation, the FAIR model is respectively 25 percent and 50 percent more accurate on average than the VAR. For the fed funds rate, the MSE is small in both cases, but again with a slight advantage for FAIR. The advance of FAIR over VAR is not as large in this simulation as in the previous one, but recall that FAIR is now mis-specified.[11] Table A1 also presents the average length and coverage rate of the confidence bands capturing the 95 percent posterior probability and compares it with the confidence bands implied by a Bayesian VAR with loose, but proper, Normal-Wishart priors. We report the average length and coverage rate at the time of the peak effect of the shock of the variable of interest. We can see that the average lengths are smaller for FAIR than for the VAR, while the coverage rate of FAIR remains good.

7   Simulation study #3:   An asymmetric DGP with a mis-specified FAIR model

We now present a simulation exercise to evaluate whether FAIR is able to detect non-linearities.

7.1   The Data Generating Process

For the DGP, we proceed as in the previous section except that we modify the baseline impulse responses to introduce asymmetry. We start from a VAR with (log) GDP, inflation and the fed funds rate, where we detrend GDP with a quadratic trend. Although we could have used the same VAR as previously, we preferred this one, because the price puzzle is more substantial in this specification (Figure A9), so that the Monte-Carlo exercise will be a more stringent test on a FAIR model with one Gaussian function that cannot capture the oscillating pattern

---

[10]Specifically, we report $MSE = \sum_{h=1}^{25}(\hat{\psi}(h) - \psi(h))^2$ where $\hat{\psi}$ is the estimated impulse response function and ψ is the true function.

[11]As with the earlier simulation study with a well-specified linear FAIR, the reason for the superior performances of FAIR is the fact that the VAR often shows counterfactual oscillation patterns. In contrast, FAIR is disciplined by its stricter parametrization.

in inflation. Again, the goal of the exercise is to assess whether a FAIR model that only approximates the main feature of the impulse responses can still recover non-linearities.

We consider a DGP where the impulse response functions to monetary shocks depend on the sign of the shock. To introduce asymmetry, we modify the impulse responses $\left\{ \hat{\boldsymbol{\Psi}}_h \right\}_{h=0}^{\infty}$ to make them depend on the sign of the monetary shock, and Figure A9 plots the asymmetric impulse response functions. For realism, the level of asymmetry that we simulate is chosen to roughly match the magnitude of the asymmetry we later find in US data. Note that we do not impose asymmetry for the response of the fed funds rate. This is done to test whether our procedure incorrectly reports the existence of asymmetry when there is none.

From these impulse responses, we the generate simulated data from

$$y_t = \sum_{h=0}^{\infty} \hat{\boldsymbol{\Psi}}_h(\varepsilon_{t-h}) \varepsilon_{t-h} \tag{19}$$

with ε_t Normally distributed, $E\varepsilon_t = 0$ and $E\varepsilon_t \varepsilon_t' = \mathbf{I}$.

As in simulation study #2, we use 50 Monte-Carlo replications with a sample size $T = 200$.

7.2 FAIR performances

We estimate a one-basis-function FAIR model with asymmetry on each set of simulated data, and Table A2 presents summary statistics for $a^+ - a^-$, which captures the amount of peak asymmetry for each one of the three variables in the model.

A number of results emerge. First, as shown by the frequency of rejection of zero coefficient for $a^+ - a^-$, the algorithm can detect asymmetry when it exists (case of output and inflation, first row of Table A2), even when the impulse response is not generated by one Gaussian, and even when, as with inflation, there is a strong oscillating pattern that cannot be captured by a one Gaussian approximation.[12] This is encouraging, because it supports our motivating idea that by approximating the most important feature of an impulse response, one can detect important non-linearities. Moreover, the algorithm does not detect asymmetry when there is none (case of the fed funds rate). Second, looking at the mean and standard-deviation of the estimates across Monte-Carlo replications (second row of Table A2), we can see that the algorithm under-estimates the amount of asymmetry (both for output and inflation). This indicates that in our empirical application on US data, our algorithm may under-estimate the magnitude of asymmetry present in the data. Third, the dispersion (third row) in the estimates across the Monte-Carlo replications is reasonably small, while the coverage rate of

---

[12]Specifically, the 90 percent posterior probability of $a^+ - a^-$ excludes zero for output and inflation respectively 94 and 90 percent of the time.

the posterior distribution – the frequency with which the true value lies within 90 percent of the posterior distribution–, is also good (fourth row).

8 Prior IRFs and sign-restrictions identification

In this section, we give two examples of how the *a-b-c* priors translate into priors on the IRFs. Specifically, Figure A10 plots the prior IRF of unemployment as used in all identification schemes and the prior IRF of inflation as used in the sign-restriction schemes (the prior IRFs are in response to a 100bp monetary shock).

For the IRF of unemployment, recall that we used the following *a-b-c* priors: we centered the priors on the values for a, b and c obtained by matching the impulse responses obtained from the VAR, and we set the standard-deviations of the priors as $\sigma_a = 10$ ppt, $\sigma_b = 10$ quarters and $\sigma_c = 20$ quarters with the constraint $c > 0$. The upper panel of Figure A10 shows that the prior IRF for unemployment is very loose: Notice that the scale of the y-axis is *two* orders of magnitude larger than the initial guess (plotted in figure 1 of the main text). Moreover, the shape of the IRF is also little restricted by our priors.

Turning to inflation, Figure A10 (bottom panel) shows the sign-restriction in action. Recall that we imposed (section 3 of the main text) that the loading on the second basis function is negative ($a_{\pi,2} < 0$), while the first basis function (meant to capture a possible price puzzle) can load positively or negatively but is restricted to peak within a year ($b_{\pi,1} \leq 4$) with a "half-life" of at most a year ($c_{\pi,1}\sqrt{\ln 2} \leq 4$). Figure A10 shows that our identification restriction on inflation is that the price puzzle cannot last for too long, roughly not more than 2 years, as the response of inflation must be negative after that. Note again that apart from this sign restriction, the prior on the magnitude or the shape of the IRF is very loose.

9 Contrasting a FAIR model and a finite state Markov-switching model

In this section, we clarify the conceptual differences between a FAIR model and a (finite state) Markov-Switching (MS) model, notably why a MS model cannot easily capture asymmetric impulse response functions.

Consider as Data-Generating Process (DGP) a univariate non-linear $\text{FAIR}_{\mathcal{G}_1}$ model[13]

$$y_t = \sum_{h=0}^{H} \psi(h, \varepsilon_{t-h})\varepsilon_{t-h} \tag{20}$$

where ε_t is an i.i.d. innovation with $E\varepsilon_t = 0$ and $E\varepsilon_t^2 = 1$, and H is the number of lags, which can be finite or infinite, and

$$\psi(h, \varepsilon_{t-h}) = a(\varepsilon_{t-h})e^{-\left(\frac{h-b(\varepsilon_{t-h})}{c(\varepsilon_{t-h})}\right)^2}$$

In (20), the lag coefficient $\psi(h, \varepsilon_{t-h},)$ is the impulse response of y_t at horizon h to innovation ε_t. In this non-linear model, the impulse response function ψ depends on the values of the innovations ε_t (for instance, positive vs negative), because the a-b-c coefficients are (continuous) functions of ε_t.

Turning to the Markov-switching model, an MS model for y_t would write

$$y_t = \sum_{h=0}^{H} M_{s_t}(h)\varepsilon_{t-h} \tag{21}$$

where $M_{s_t}(h)$ depends on the value of the state variable s_t.

Note first that a finite state MS model will not be able to perfectly reproduce the moving-average coefficients, because the a-b-c coefficients (and thus ψ) can take on an infinite number of values, depending on the realization of the continuous variable ε.

Next, consider a more specific version of (20) with asymmetric impulse responses, i.e., where the a-b-c coeficients depend *only* on the sign of the shock. The FAIR model in that case is

$$y_t = \sum_{h=0}^{H} \psi^+(h)\varepsilon_{t-h}\mathbf{I}(\varepsilon_{t-h} > 0) + \sum_{h=0}^{H} \psi^-(h)\varepsilon_{t-h}\mathbf{I}(\varepsilon_{t-h} \leq 0) \tag{22}$$

with $\psi^+(h) = a^+ e^{-(\frac{h-b^+}{c^+})^2}$ and similarly for ψ^-.

Note that DGP (22) can now be described with a finite number of states, since ψ can only take a finite number of values. However, a very large number of Markov states (2^H) would be needed to perfectly approximate the MA coefficients of (22), much larger than typical empirical applications (for say quarterly data with $H = 20$, that would be 2^{20} states).

Moreover, even if the MS model could allow for 2^H states, the *typical* Markov-Switching

---

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

model would be misspecified. Indeed, in empirical applications, the discrete Markov state driving it is typically either assumed to be exogenous (and independent of the other shocks in the model, in particular independent of ε_t) or dependent on lagged endogenous variables (e.g., Sims, Waggoner and Zha, 2008). However, $\psi(0, \varepsilon_t)$ –the contemporaneous moving-average coefficient in model (22)– depends on the *current* period shock. This will lead to a misspecification of the typical MS model.

10 The asymmetric effects of shocks: additional results

To give a complete set of results for the asymmetric effects of shocks, we plot the impulse responses to positive and negative monetary shocks under the narrative and sign identification schemes (Figure A11 to A12).[14]

For both identification schemes, a contractionary monetary shock significantly increases unemployment whereas an expansionary monetary shock has little on effect on unemployment (and non-significantly different from zero). A similar mirror image pattern holds for inflation.

---

[14]When comparing impulse responses to positive and negative shocks, keep in mind that the impulse responses to expansionary monetary shocks (a decrease in the fed funds rate) were multiplied by -1 in order to ease comparison across impulse responses. With this convention, when there is no asymmetry, the impulse responses are identical in the upper panels (responses to a contractionary monetary shock) and in the bottom panels (responses to an expansionary monetary shock).

References

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

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

[3] De Graeve, F. and A. Westermark, "Un-truncating VARs," Working Paper Series 271, Sveriges Riksbank, 2013

[4] Korevaar J. *Mathematical Methods*, Vol. 1, pp 330-333. Academic Press, New York, 1968

[5] Robert, C. and G. Casella "Monte Carlo Satistical Methods" Springer, 2004

[6] Sims, C., D. Waggoner, and T. Zha "Methods for inference in large multiple-equation Markov-switching models," Journal of Econometrics, Elsevier, vol. 146(2), pages 255-274, October 2008
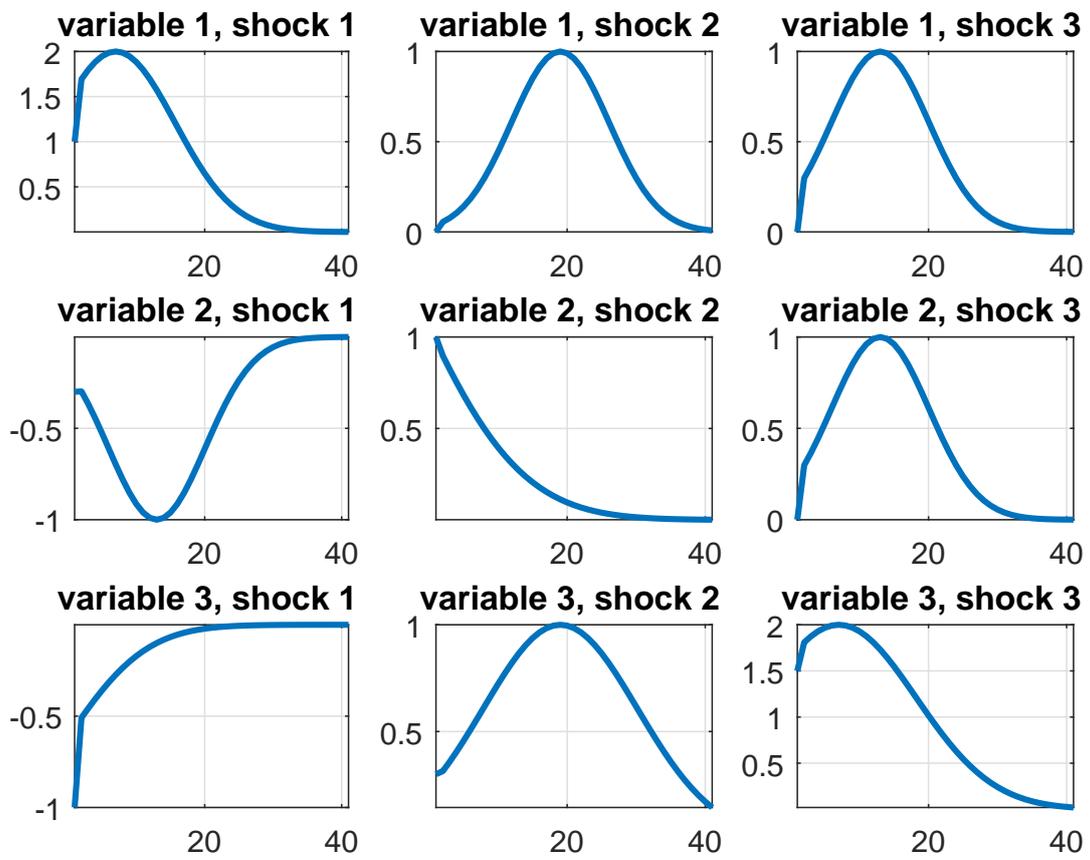
Figure A1: Data-generating process
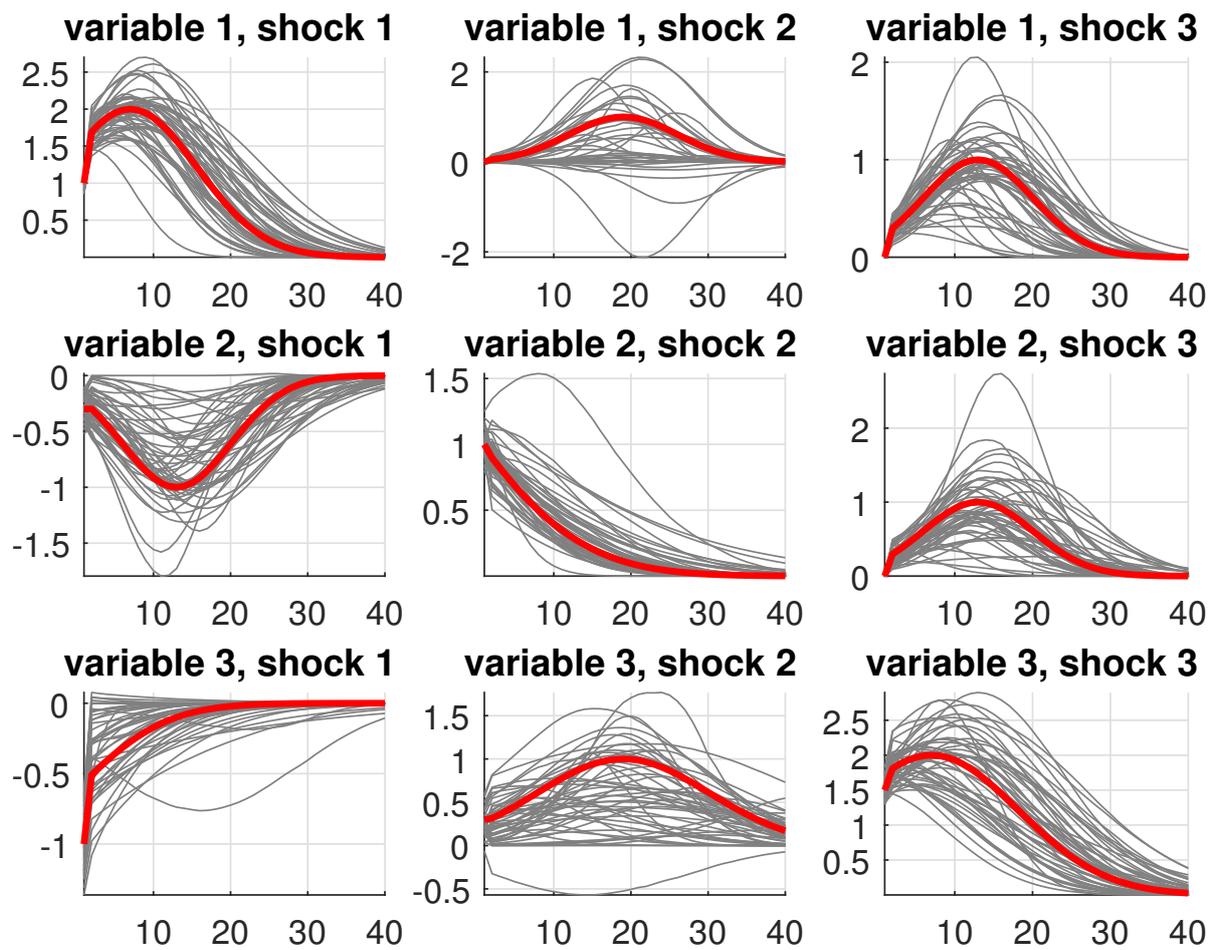
Figure A2: FAIR impulse responses
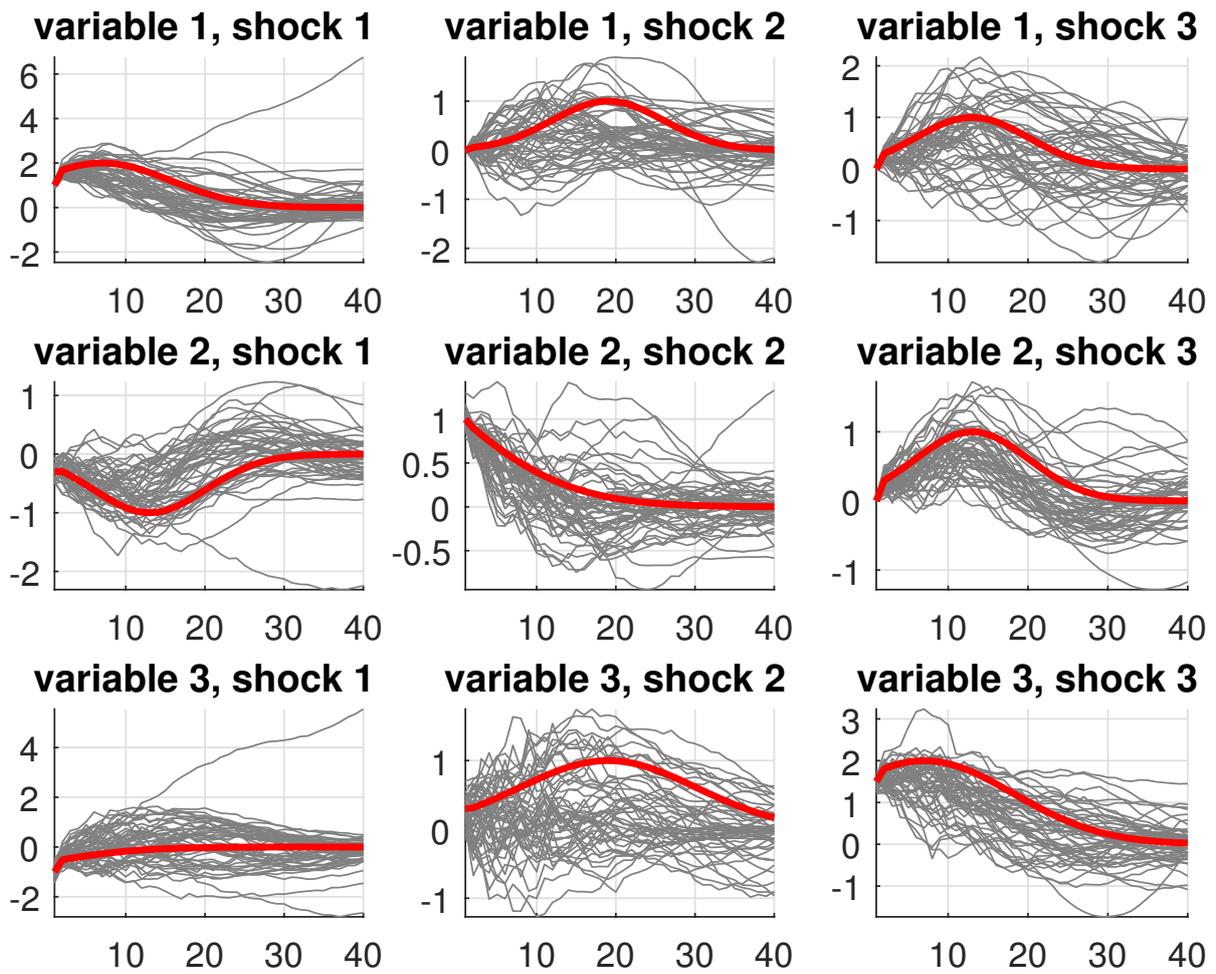
Figure A3: VAR impulse responses
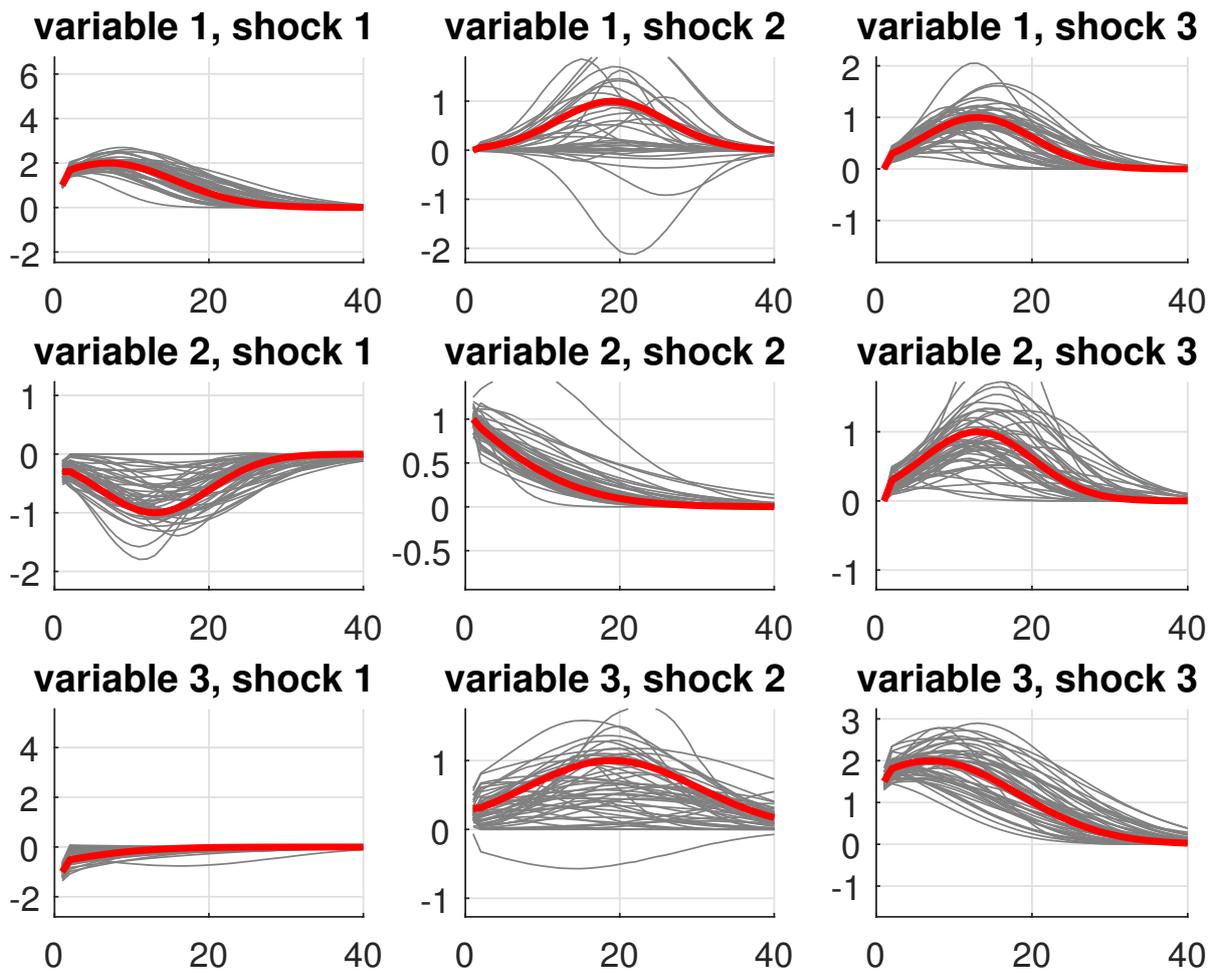
Figure A4: FAIR impulse responses
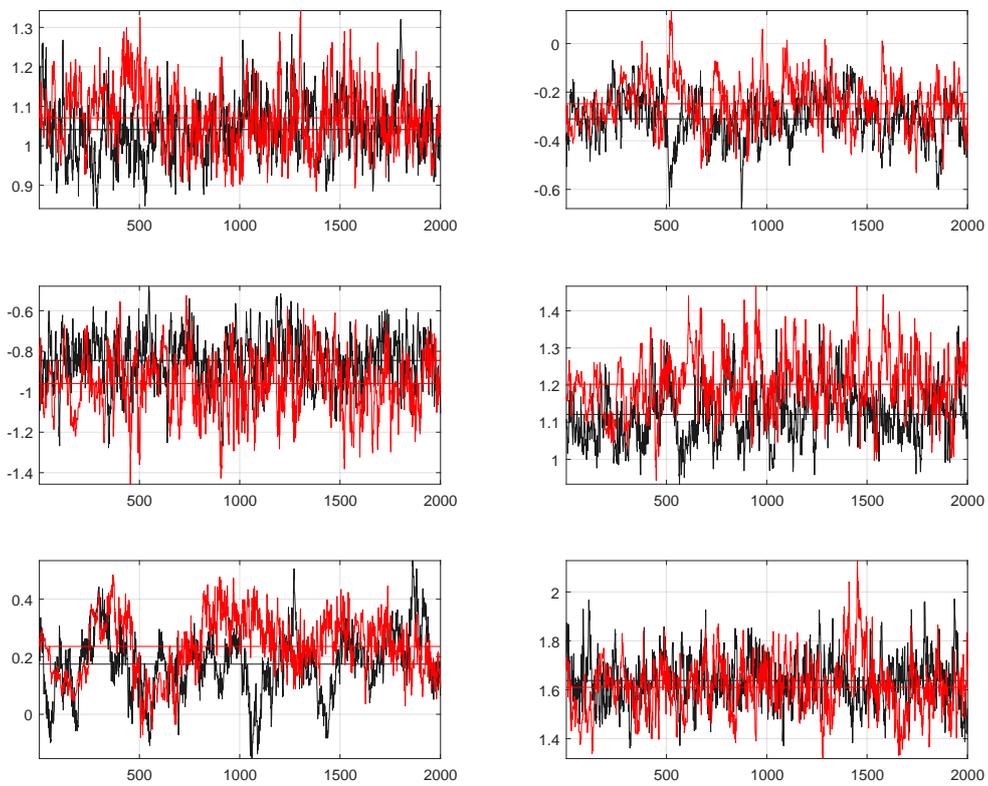
Figure A5: Trace plots for impact coefficients
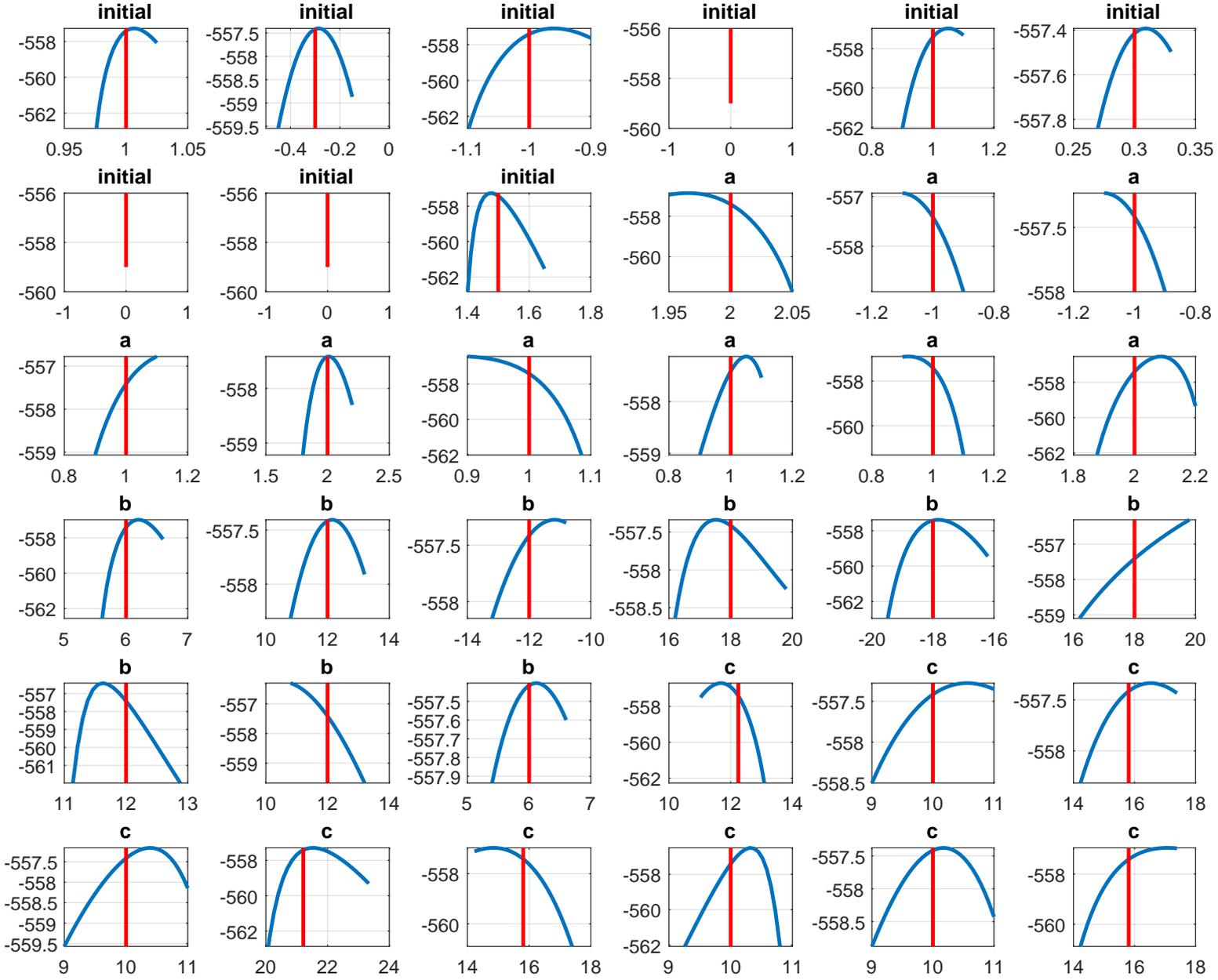
Figure A6: likelihood surfaces, 120 observations. The titles in the subplots denote which group of parameters the specific parameter plotted belongs to. If only a red line appears then that parameters is restricted to a specific value in estimation.
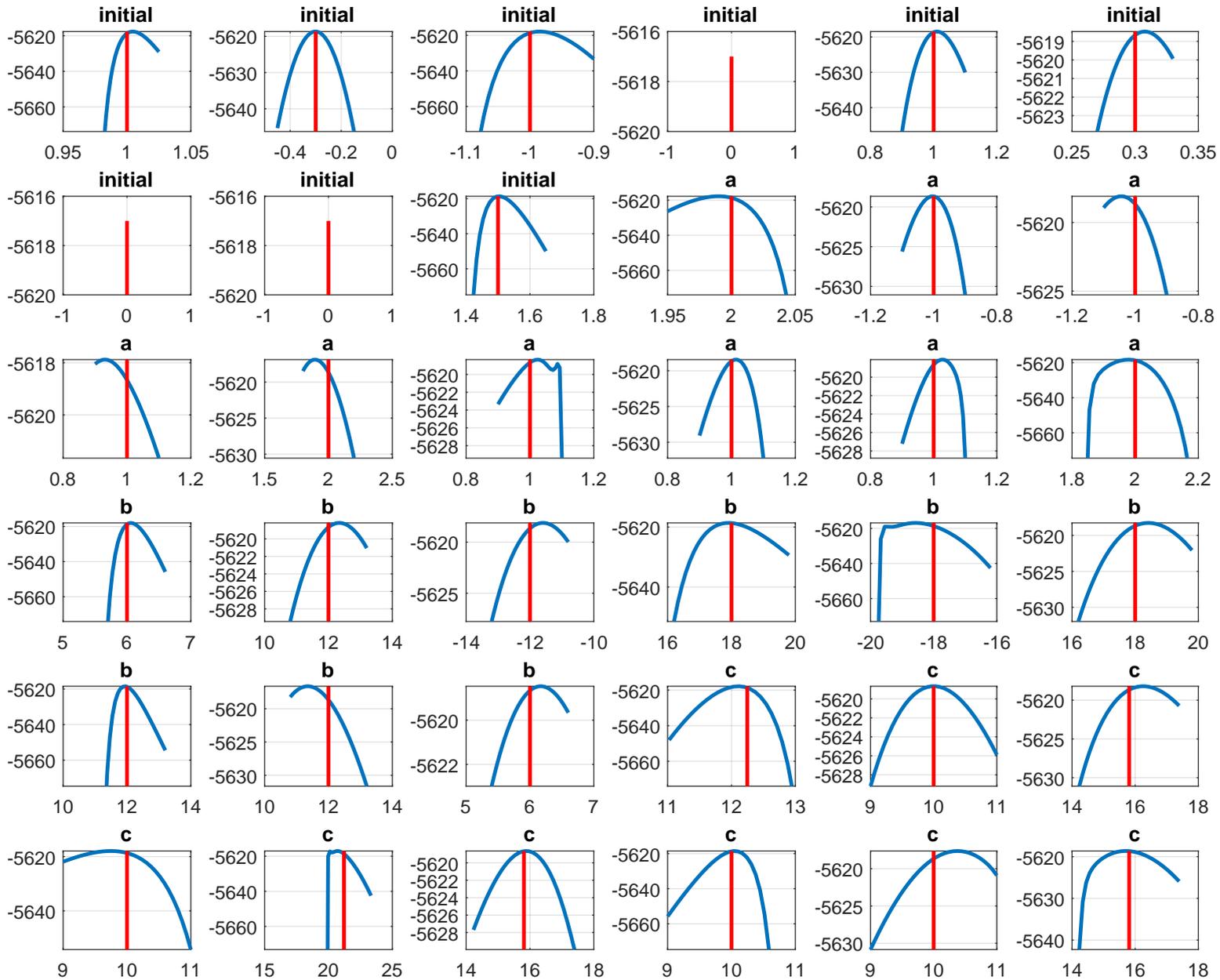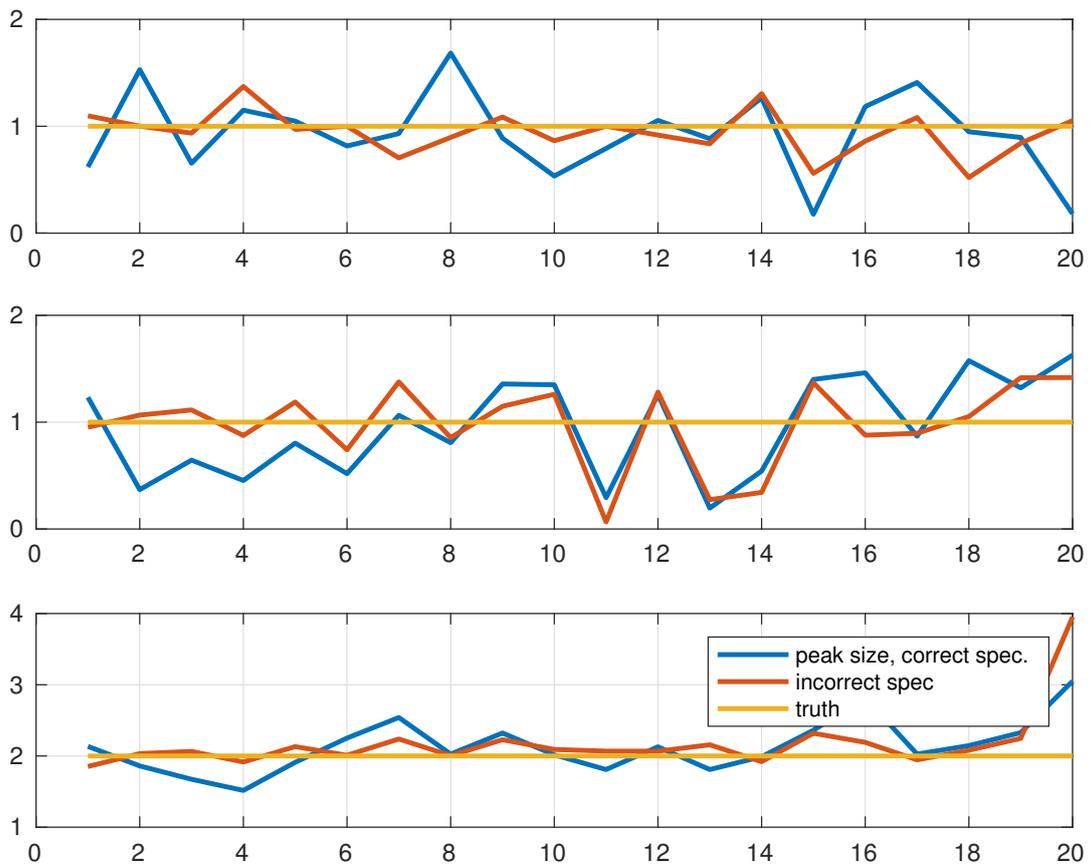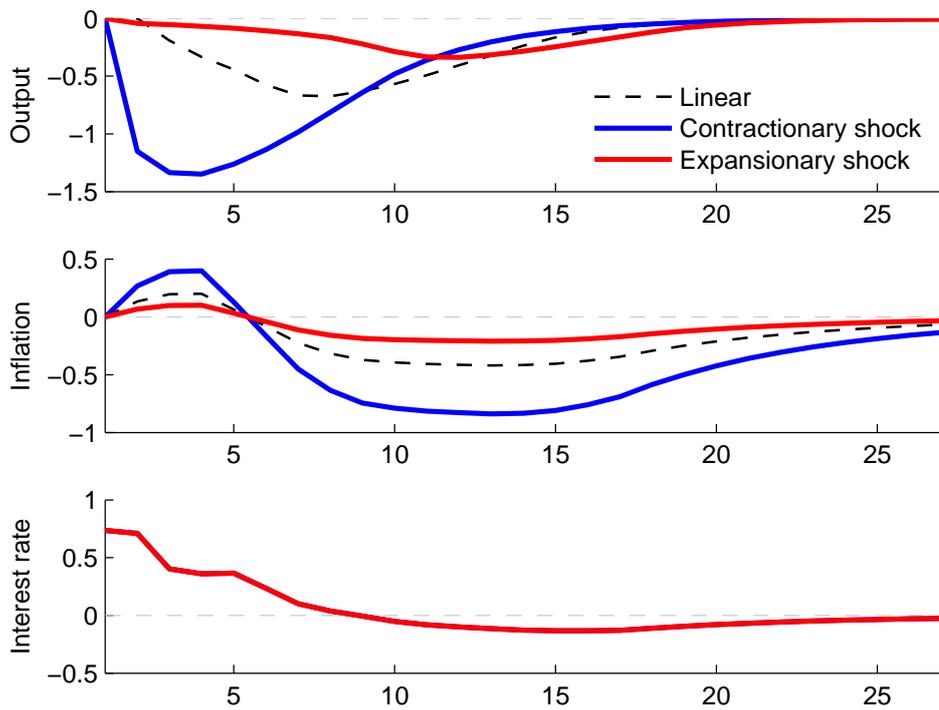
Figure A7: likelihood surfaces, 1200 observations. The titles in the subplots denote which group of parameters the specific parameter plotted belongs to. If only a red line appears then that parameters is restricted to a specific value in estimation.

Figure A8: Peak effects

Figure A9: Monte Carlo simulation with asymmetric impulse responses to monetary shocks. The thick blue lines report the simulated impulse responses to a contractionary shock, and the thick red lines report the simulated impulse responses to an expansionary shock (with the responses to an expansionary shock multiplied by -1 for clarity of exposition). The dashed lines are the impulse responses estimated from a VAR over 1959-2007.
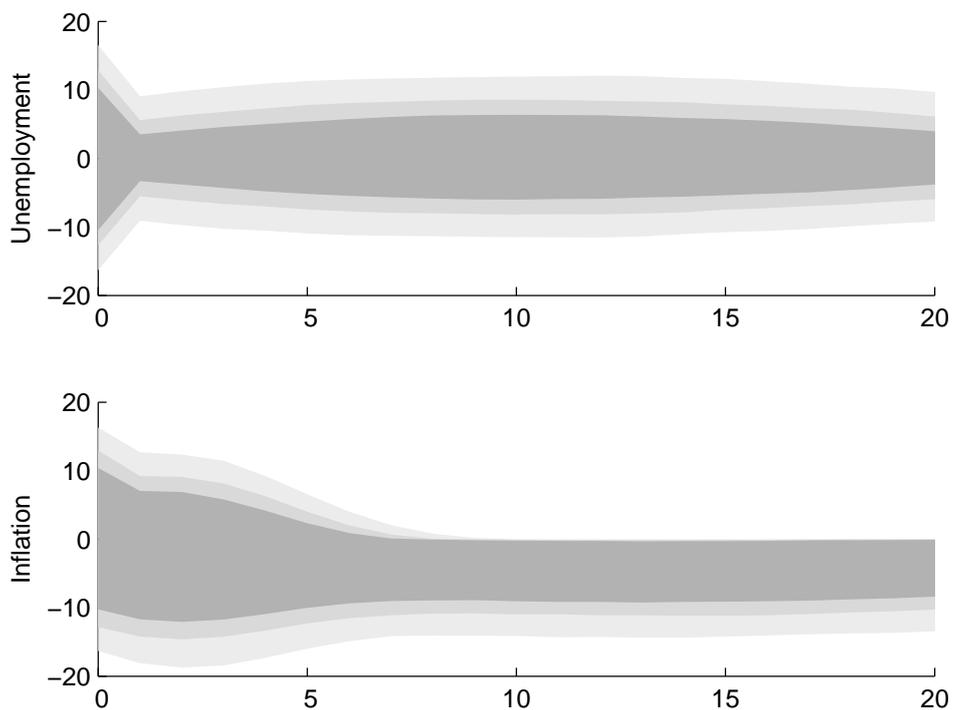
Figure A10: IRF prior distribution for unemployment (all three identification schemes) and for inflation, as used in the sign-restrictions identification scheme. Shaded area denote respectively the 70th, 80th and 90th percentiles of the IRF prior distribution implied by the prior distribution on FAIR parameters.
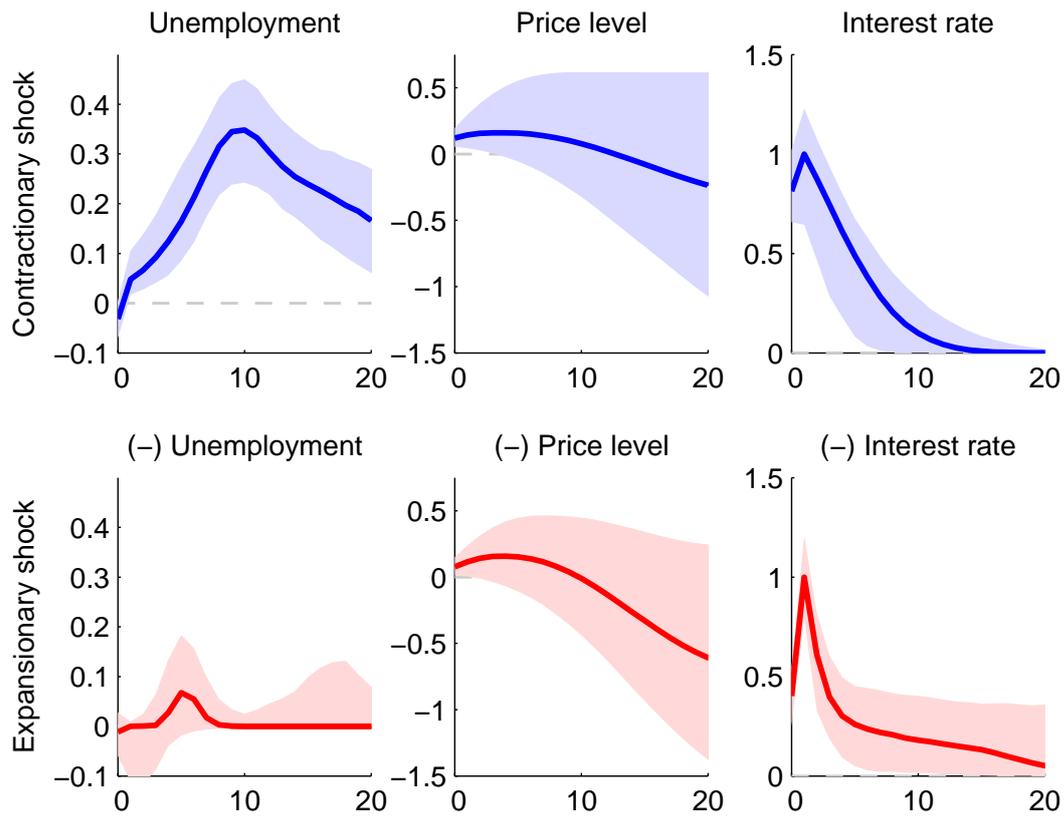
Figure A11: **Asymmetric IRFs, narrative identification**: FAIR estimates of the IRFs of unemployment (in ppt), the (log) price level (in percent) and the federal funds rate (in ppt) to 100bp monetary shock identified by Romer and Romer (2004). Shaded bands denote the 5th and 95th posterior percentiles. For ease of comparison, responses to the expansionary shock are multiplied by -1. Sample 1969-2007
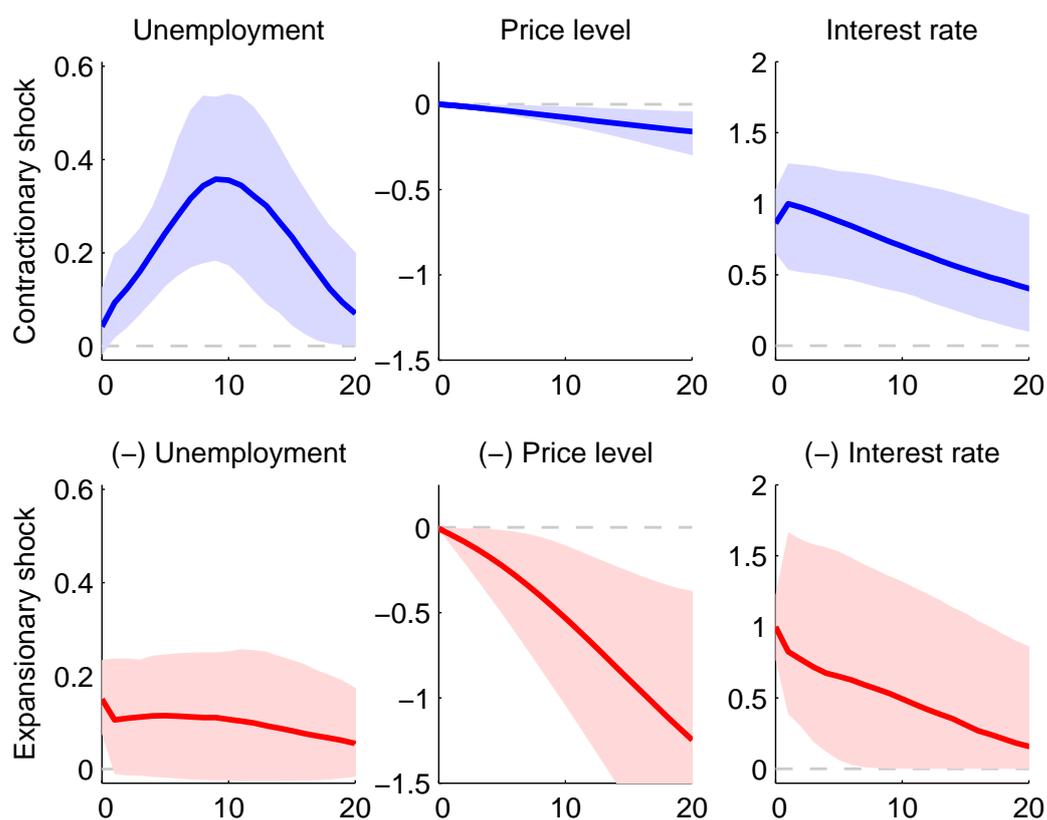
Figure A12: **Asymmetric IRFs, sign-restrictions identification**: FAIR estimates of the IRFs of unemployment (in ppt), the (log) price level (in percent) and the federal funds rate (in ppt) to a 100bp monetary shock identified with sign restrictions. Estimation from a FAIR with asymmetry (plain line). Shaded bands denote the 5th and 95th posterior percentiles. For ease of comparison, responses to the expansionary shock are multiplied by -1. With this convention, when there is no asymmetry, the impulse responses are identical in the upper panels (responses to a contractionary monetary shock) and in the bottom panels (responses to an expansionary monetary shock). Sample 1959-2007.