Functional Approximations of Impulse Responses: New Insights into the Asymmetric Effects of Monetary Policy*

Regis Barnichon
FRB San Francisco, CREI, CEPR

Christian Matthes
FRB Richmond

June 2017

Abstract

Is expansionary monetary policy akin to pushing on a string? To study the asymmetric effects of monetary shocks, this paper develops a new, simple approach that consists in approximating impulse responses with basis functions in order to directly estimate the impulse responses from the data. Regardless of whether we identify monetary shocks from (i) a timing restriction, (ii) sign restrictions, or (iii) a narrative approach, the effects of monetary policy are highly asymmetric: A contractionary shock has a strong adverse effect on unemployment, but an expansionary shock has little effect.

JEL classifications: C14, C32, C51, E32, E52

*First draft: March 2014. We would like to thank Luca Benati, Francesco Bianchi, Christian Brownlees, Fabio Canova, Tim Cogley, Davide Debortoli, Wouter den Haan, Jordi Gali, Yuriy Gorodnichenko, Eleonora Granziera, Oscar Jorda, Thomas Lubik, Jim Nason, Kris Nimark, Mikkel Plagborg-Moller, Giorgio Primiceri, Ricardo Reis, Jean-Paul Renne, Barbara Rossi, Mark Watson, Yanos Zylberberg and seminar participants at the Barcelona GSE Summer Forum 2014, the 2014 NBER/Chicago Fed DSGE Workshop, William and Mary college, EUI Workshop on Time-Varying Coefficient Models, Oxford, Bank of England, NYU Alumni Conference, SED Annual Meeting (Warsaw), Universitaet Bern, Econometric Society World Congress (Montreal), the Federal Reserve Board, the 2015 SciencesPo conference on Empirical Monetary Economics and the San Francisco Fed for helpful comments. The views expressed here do not necessarily reflect those of the Federal Reserve Bank of Richmond or San Francisco or of the Federal Reserve System. Any errors are our own.

1 Introduction

There now exists a relatively broad consensus on the average effect of monetary policy on economic activity, and it is generally accepted that a monetary contraction (expansion) leads to a decline (increase) in economic activity.

However, one question at the core of monetary policy making remains unsettled: does monetary policy have asymmetric effects on economic activity? As captured by the string metaphor, does a contractionary monetary shock have a stronger effect –being akin to pulling on a string– than an expansionary shock –being akin to pushing on a string–?[1] Despite a large empirical literature on the effects of monetary policy, the evidence for asymmetric effects is relatively thin and inconclusive.[2]

Answering the question of asymmetry has been difficult in part for one important technical reason: the traditional approach to identify the dynamic effect of shocks relies on structural Vector-Autoregressions (VARs),[3] which are linear models. While VARs can accommodate certain types of non-linearities, some questions, such as the asymmetric effect of a monetary shock, cannot be answered within a VAR framework.

This paper proposes a new method –Functional Approximations of Impulse Responses (FAIR)– to estimate the (possibly asymmetric) dynamic effects of structural shocks. Instead of assuming the existence of a VAR representation, our approach consists in working directly with the structural moving-average representation of the economy. Then, to make the estimation of the moving-average representation feasible, we approximate the impulse response functions with Gaussian basis functions. While different families of basis functions are possible, Gaussian basis functions are particularly attractive for two reasons: (i) any mean-reverting impulse

[1]Theoretically, there are at least two reasons why monetary policy could have asymmetric effects. First, with downward wage/price rigidity, economic activity can display asymmetric responses to monetary changes (e.g., Morgan, 1993). Second, with borrowing constraints, increasing borrowing costs may curb investment and/or consumption, but cheap financing need not stimulate investment or consumption. This can happen if borrowing-constrained agents respond more to income drops than to income gains (Auclert, 2017), a possibility supported by recent evidence from Bunn et al. (2017) and Christelis et al. (2017).

[2]For instance, while Cover (1992) and Tenreyro and Thwaites (2016) find evidence of asymmetric effects, Ravn and Sola (1996, 2004) and Weise (1999) instead find nearly symmetric effects.

[3]See e.g., Christiano, Eichenbaum, and Evans (1999) and Uhlig (2005).

response function can be approximated by a linear combination of Gaussian basis functions, and (ii) a small number (one or two) of Gaussian functions can already capture a large variety of impulse response functions, and notably the typical impulse responses found in empirical or theoretical studies. For instance, the impulse response functions to monetary shocks are often found (or theoretically predicted) to be monotonic or hump-shaped (e.g., Christiano, Eichenbaum and Evans 1999, Walsh 2010). In such cases, a single Gaussian function can already provide an excellent approximation of the impulse response function.[4] In turn, the parsimony of the approach allows for the estimation of more general non-linear models.

FAIR can easily incorporate the identification schemes found in the structural VAR literature, and we explore the asymmetric effects of monetary shocks using three popular identification schemes: a recursive identification scheme, whereby monetary policy shocks can only affect macro variables with a one period lag (Christiano, Eichenbaum and Evans, 1999), (ii) a set identification scheme based on sign restrictions,[5] and (iii) a narrative identification scheme where a series of monetary shocks has been previously identified (possibly with measurement error) from narrative accounts (Romer and Romer, 2002).

Consistent with the string metaphor, our findings point towards the existence of strong asymmetries in the effects of monetary shocks, and Bayesian model comparison strongly favors a FAIR model with asymmetry over a linear VAR model. Regardless of whether we identify monetary shocks from a recursive ordering, from sign restrictions or from a narrative approach, we find that a contractionary shock has a strong adverse effect on unemployment, larger than implied by linear estimates, while an expansionary shock has little effect on unemployment. This finding is interesting in the context of the current debate among OECD countries on the appropriate timing of the lift-off of the policy rate. Our estimates suggest that an inappropriate (i.e., too strong or too early) increase in the policy rate could be a lot more costly (in terms of economic activity) than conventional (linear) estimates suggest, calling for a cautious approach

[4]Another advantage of using Gaussian basis functions is that prior elicitation can be much easier than with Bayesian estimation of standard VARs, because the coefficients to be estimated are directly interpretable as features of impulse responses.

[5]See e.g., Faust (1998), Canova and De Nicolo (2002), Uhlig (2005).

to monetary tightening.

Although our evidence for inflation is more uncertain, we find that inflation responds less strongly following a contractionary shock than following an expansionary shock. In other words, nominal quantities react more following an expansionary shock, while real quantities react more following a contractionary shock. Interestingly, this pattern is consistent with a Keynesian narrative where downward nominal rigidities lead to opposite asymmetric responses of unemployment and inflation.

Our use of basis functions to approximate impulse response functions relates to a large literature in statistics that relies on basis functions (of which Gaussian functions are one example) to approximate arbitrary functions (e.g., Hastie, Tibshirani and Friedman, 2009). In basis functions methods, the complexity of the model is typically controlled through a combination of restriction, selection and/or regularization methods. Our approach, which consists in using a limited number of basis functions, uses both selection and restriction to control the complexity of the model.[6]

In economics, our parametrization of impulse responses relates to an older literature on distributed lag models; in particular to the Almon (1965) lag specification, in which the successive weights, i.e., the impulse response function in our context, are given by a polynomial function, or to the Jorgenson (1966) specification, in which the successive weights are given by a rational function.[7] A recent promising approach related to ours is Plagborg-Moller (2016), who proposes a Bayesian method to directly estimate the structural moving-average representation of the data by using prior information about the shape and the smoothness of the impulse responses.

The economic literature has so far tackled the estimation of non-linear effects of shocks in

[6]It uses selection in the sense that our algorithm scans the dictionary of possible basis functions to find the Gaussian basis functions that best fit the data (in a maximum likelihood sense), and it uses restriction in the sense that we restrict ourselves to the class of impulse response functions that can be generated by a few Gaussian basis functions.

[7]In fact, the Jorgenson specification amounts to approximating an infinite moving average model with a finite order ARMA; a common approach in macro-econometrics. As mentioned earlier, we use Gaussian basis functions because of their parsimony and ease of interpretation.

two main ways: (i) Autoregressive Distributed Lags models (ADL) or Local Projections (LP, Jorda, 2005) combined with independently identified shocks, and (ii) non-linear structural VARs.[8]

In its simplest form, the first approach consists in estimating an ADL model, i.e., regressing a variable of interest on contemporaneous and lagged values of some independently identified shocks, while allowing for possible non-linear effects. In the context of monetary policy, Cover (1992), DeLong and Summers (1988) and Morgan (1993) identify monetary shocks from unanticipated money innovations (obtained from a money supply process regression, following Barro, 1977) and test whether the impulse response function depends on the sign of these innovations. While that approach was later abandoned because money supply regressions were suspected to poorly identify monetary shocks, this approach has been recently revived by Tenreyro and Thwaites (2016) using narratively identified shocks (Romer and Romer, 2002) and the Local Projections (LP) method pioneered by Jorda (2005).[9] The narrative approach was precisely developed in order to identify exogenous monetary innovations, and Jorda's method can easily accommodate non-linearities in the response function. However, the LP (or ADL) method suffers from two limitations: (i) shock identification, (ii) low efficiency. First, to explore the non-linear effects of shocks, LP requires a series of previously identified structural shocks, which limits its use to narrative identification schemes. In contrast, FAIR can easily accommodate the identifying schemes found in the structural VAR literature.[10] Second, LP is limited by efficiency considerations. Indeed, while the Local Projections approach is intentionally model-free –not imposing any underlying dynamic system–, this can come at an efficiency cost (Ramey, 2012), which makes inferences difficult, particularly in non-linear models. In contrast, by positing that the response function can be approximated by one (or a few)

[8]A third non-linear approach was recently proposed by Angrist et al. (2016) who develop a semi-parametric estimator to evaluate the (possibly asymmetric) effects of monetary policy interventions. They find asymmetric effects of monetary shocks consistent with our findings.

[9]The combination of Jorda's method with narratively identified shocks was first introduced in the context of fiscal policy by Auerbach and Gorodnichenko (2013).

[10]Moreover, imposing identifying restrictions in FAIR models is typically straightforward, because FAIR works directly with the structural moving average representation of the data.

Gaussian functions, FAIR imposes strong dynamic restrictions between the parameters of the impulse response function, which can improve efficiency.[11] Moreover, FAIR alleviates another source of inefficiency in LP, namely the presence of serial correlation in the LP residuals. By jointly modeling the behavior of key macroeconomic variables (in the spirit of a VAR), FAIR is effectively modeling the serial correlation present in LP residuals, which can further improve efficiency. A final advantage of FAIR over LP is that it can be used for model selection and model evaluation through marginal data density comparisons.

A second strand in the literature has relied on regime-switching VAR models –notably threshold VARs (e.g., Hubrich and Terasvirta, 2013) and Markov-switching VARs (Hamilton, 1989)– to capture certain types of non-linearities.[12,13] However, while regime-switching VARs can capture state dependence (whereby the value of some state variable affects the impulse response functions), they cannot capture asymmetric effects of shocks (whereby the impulse response to a structural shock depends on the sign of that shock). Indeed, with regime-switching VAR models, it is assumed that the economy can be in a finite number of regimes, and that each regime corresponds to a different set of VAR coefficients. However, if the true data generating process features asymmetric impulse responses, a new set of VAR coefficients would be necessary each period, because the behavior of the economy at any point in time depends on all structural shocks up to that point. As a result, such asymmetric data generating processes cannot be easily captured by threshold VARs or Markov-switching models that only handle a finite (and typically small) number of state variables. In contrast, by working directly with the structural moving-average representation, FAIR can easily capture asymmetric impulse response functions.

Section 2 describes how we approximate impulse responses using Gaussian basis functions,

[11] Naturally, this statement also implies that our results are valid under the assumption that response functions can be well approximated by a few Gaussian functions. In this respect, our approach is best seen as complementing the model-free approach of Jorda (2005).

[12] For examples in the monetary policy literature, see Beaudry and Koop (1993), Thoma (1994), Potter (1995), Kandil (1995), Koop, Pesaran and Potter (1996), Koop and Potter, (1998), Ravn and Sola (1996, 2004), Weise (1999), Lo and Piger (2005).

[13] Another prominent class of non-linear VARs includes models with time-varying coefficients and/or time-varying volatilities (e.g., Primiceri, 2005).

Section 3 discusses the key steps of the estimation methodology; Section 4 generalizes our approach to non-linear models; Section 5 uses FAIR models to the study of the asymmetric effects of monetary shocks; Section 6 concludes.

2 Empirical model

Our goal in this paper is to study how the effect of monetary policy changes depends on the sign of the policy intervention.

To capture this possibility, we need a model that allows the impulse response functions to depend on the sign of the shock.[14] Our empirical model is thus a (non-linear) structural moving-average model, in which the behavior of a vector of macroeconomic variables is dictated by its response to past and present structural shocks. Specifically, denoting $\boldsymbol{y}_t$ a vector of stationary macroeconomic variables, the economy is described by

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

where $\boldsymbol{\varepsilon}_t$ is the vector of i.i.d. structural innovations with $E\boldsymbol{\varepsilon}_t = \mathbf{0}$ and $E\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' = \mathbf{I}$, K is the number of lags, which can be finite or infinite. $\boldsymbol{\Psi}_k$ is the matrix of lag coefficients –i.e., the matrix of impulse responses at horizon k–.

Model (1) is a *non-linear* vector moving average representation of the economy, because the matrix of lag coefficients $\boldsymbol{\Psi}_k$, i.e., the impulse responses of the economy, can depend on the values of the structural innovations $\boldsymbol{\varepsilon}$: With $\boldsymbol{\Psi}_k$ a function of $\boldsymbol{\varepsilon}_{t-k}$, the impulse responses to a given structural shock depend on the value of that shock at the time of shock. For instance, a positive shock may trigger different impulse responses than a negative shock.

Importantly, our starting point is *not* a structural Vector AutoRegression (VAR). While the use of a VAR is a common way to estimate a moving-average model, it relies on the existence of a VAR representation. However, in a non-linear world where $\boldsymbol{\Psi}_k$ depends on the sign of the

[14] As we argue in two paragraphs, a VAR is ill-suited to capture such non-linearities.

shocks $\boldsymbol{\varepsilon}$ as in (1), the existence of a VAR is severely compromised, because inverting (1) is generally not possible. Thus, in this paper, we work with an empirical method that side-steps the VAR and instead directly estimates the vector moving average model (1).

2.1 Functional Approximation of Impulse Responses (FAIR)

Estimating moving-average model is notoriously difficult, because the number of free parameters $\boldsymbol{\Psi}_k$ in (1) is very large or infinite. To address this issue, we use a new approach –Functional Approximation of Impulse Responses or FAIR–, which consists in representing the impulse response functions as expansions in basis functions.

Since the intuition and benefits of our approach can be understood in a linear context, this section introduces FAIR in a linear model, i.e., where $\boldsymbol{\Psi}_k(\boldsymbol{\varepsilon}_{t-k}, z_{t-k}) = \boldsymbol{\Psi}_k$. We postpone non-linear models to a later section.

Denote $\psi(k)$ an element of matrix $\boldsymbol{\Psi}_k$, so that $\psi(k)$ is the value of the impulse response function ψ at horizon k. A functional approximation of ψ consists in decomposing ψ into a sum of basis functions, i.e., in modeling $\psi(k)$ as a *basis function expansion*:

$$\psi(k) = \sum_{n=1}^{N} a_n g_n(k), \quad \forall k > 0 \tag{2}$$

with $g_n : \mathbb{R} \to \mathbb{R}$ the nth basis function, $n = 1, .., N$. Different families of basis functions are possible, and in this paper we use Gaussian basis functions and posit

$$\psi(k) = \sum_{n=1}^{N} a_n e^{-(\frac{k-b_n}{c_n})^2}, \quad \forall k > 0 \tag{3}$$

with a_n, b_n, and c_n parameters to be estimated. Since model (3) uses N Gaussian basis functions, we refer to this model as a functional approximation of order N.[15]

While different families of basis functions are possible, Gaussian basis functions are partic-

[15] The functional approximation of ψ may or may not include the contemporaneous impact coefficient, that is one may choose to use the approximation (3) for $k > 0$ or for $k \geq 0$. In this paper, we treat $\psi(0)$ as a free parameter for additional flexibility.

ularly attractive in our context for two reasons.

First, as Theorem 1 below establishes, any mean-reverting impulse response function can be approximated by a sum of Gaussian basis functions.

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. See Appendix. ■

Second, only a very small number of Gaussian basis functions are needed to approximate a large class of impulse response functions, in fact most impulse responses encountered in macro applications. Intuitively, impulse response functions are often found (or theoretically predicted) to be monotonic or hump-shaped (e.g., Christiano, Eichenbaum, and Evans, 1999).[16] In such cases, one or two Gaussian basis functions can already provide a very good approximate description of the impulse response.

Figure 1 shows how one Gaussian basis function can capture a hump-shape impulse response, while two Gaussian basis functions can capture an overshooting/oscillating impulse response.

In the context of monetary policy, Figure 2 plots the impulse response functions of unemployment, the price level and the fed funds rate to a shock to the fed funds rate estimated from a standard VAR specification with a recursive ordering,[17] along with the corresponding

[16] In New-Keynesian models, the impulse response functions are generally monotonic or hump-shaped (see e.g., Walsh, 2010).

[17] We describe the exact specification in section 5.

estimates of a FAIR model with only *one* Gaussian function, i.e., using the approximation

$$\psi(k) \simeq a e^{-\frac{(k-b)^2}{c^2}}, \quad \forall k > 0. \tag{4}$$

We can see that a FAIR with only one Gaussian basis function already does a good job at capturing the impulse responses implied by the VAR. With two Gaussian basis functions, the impulse responses are very close to those of the VAR (Figure 2). For illustration, Figure 3 plots the Gaussian basis functions used for each impulse response in that case.

The relatively small number of free parameters (only three per impulse response function in the one-Gaussian case), has two important advantages. First, it allows us to directly estimate the impulse response functions from the moving-average representation (1).[18] Second, it will allow us to later add more degrees of freedom and allow for asymmetric effects of shocks to monetary policy.

As a last comment, we note two additional advantages of FAIR models with only *one* Gaussian basis function: (i) ease of interpretation, and (ii) ease of prior elicitation.

In a one-basis function FAIR model like (4), the a, b and c coefficients can be easily interpreted, because the impulse response function is summarized by three parameters –the peak effect, the time to peak effect, and the persistence of the impulse response–, which are generally considered the most relevant characteristics of an impulse response function.[19] As illustrated in Figure 1, parameter a is the height of the impulse-response, which corresponds to the maximum effect of a unit shock, parameter b is the timing of this maximum effect, and parameter c captures the persistence of the effect of the shock, as the amount of time τ required for the effect of a shock to be 50% of its maximum value is given by $\tau = c\sqrt{\ln 2}$.

Then, the ease of interpretation of the a, b and c parameters in turn makes prior elicitation

[18]For instance, with 3 variables, we only have $3 * 3^2 = 27$ parameters (ignoring intercepts) to estimate to capture the whole set of impulse response functions $\{\boldsymbol{\Psi}_k\}_{k=1}^{\infty}$. In comparison, a corresponding quarterly VAR with 4 lags has 36 parameters, while a monthly VAR with 12 lags has 108 parameters.

[19]For instance, when comparing the effects of monetary shocks across different specifications, Coibion (2012) focuses on the peak effect of the monetary shock, which in a FAIR model with one basis function is simply parameter a.

easier than in standard VARs, in which the VAR coefficients have a less direct economic interpretation.

3 Bayesian estimation

To estimate our model, we use a Bayesian approach, which is particularly well suited for models that approximate the true DGP (Fernandez-Villaverde and Rubio-Ramirez, 2004). In particular, Bayes factors will allow us to evaluate FAIR models against VAR models, even though the two classes of models are non-nested.[20] Bayesian model comparison will also offer us a natural way to select the order of the FAIR, i.e., the number of Gaussian basis functions used in the approximation.

In this section, we describe the implementation and estimation of FAIR models. We first describe how we construct the likelihood function by exploiting the prediction-error decomposition, discuss structural identification, then present the estimation routine based on a multiple-block Metropolis-Hasting algorithm, discuss prior elicitation, the determination of the order of the FAIR and identification issues related to fundamentalness. We conclude by discussing how to deal with non-stationary data. For clarity, we focus here on the estimation of *linear* moving-average models and leave non-linear models for the next section.

3.1 Constructing the likelihood function

We need to construct the likelihood function $p(\boldsymbol{y^T}|\boldsymbol{\theta})$ of a sample of size T for a linear moving-average model (1) 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

[20] Bayes factors are functions of the marginal data densities for the two models that are being compared. Since marginal data densities can be rewritten as products of one-step ahead forecast densities, Bayes factors also offer insights about the relative forecasting abilities of the two models that are being compared.

follows:[21]

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

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,[22] 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

$$\boldsymbol{y_t} = \boldsymbol{\Psi}_0\boldsymbol{\varepsilon_t} + \sum_{k=1}^{K} \boldsymbol{\Psi}_k\boldsymbol{\varepsilon_{t-k}}. \tag{6}$$

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 (6), 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 K innovations $\{\boldsymbol{\varepsilon_j}\}_{j=-K}^{0}$ to zero.[23,24]

3.2 Structural identifying assumptions

While the structural vector moving-average model (1) is under-identified without additional restrictions, FAIR models can easily incorporate the main identification schemes found in the structural VAR literature. In fact, since FAIR models work directly with the structural moving-average representation, imposing certain restrictions such as sign-restictions is arguably easier in FAIR than in VARs.

Given our focus on the effects of monetary policy, we will discuss three approaches that are

[21]To derive the conditional densities in decomposition (5), our parameter vector $\boldsymbol{\theta}$ thus implicitly also includes the K initial values of the shocks: $\{\boldsymbol{\varepsilon}_{-K}...\boldsymbol{\varepsilon}_0\}$. We will keep those fixed throughout the estimation and discuss alternative initializations below.

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

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

[24]When K, the lag length of the moving average (1), is infinite, we truncate the model at some horizon K, large enough to ensure that the lag matrix coefficients $\boldsymbol{\Psi}_K$ are "close" to zero. Such a K exists since the variables are stationary.

popular in the monetary literature: a recursive identification scheme (Christiano, Eichenbaum and Evans, 1999), (ii) a narrative identification scheme where a series of monetary shocks has been previously identified (possibly with measurement error) from narrative accounts (Romer and Romer, 2002), and (iii) a set identification scheme based on sign restrictions.[25] We also mention how FAIR could open the door for more general identification schemes based on shape-restrictions.

Short-run restrictions and recursive ordering

Short-run restrictions in a fully identified model consists in imposing $\frac{L(L-1)}{2}$ restrictions on $\boldsymbol{\Psi}_0$ (of dimension $L \times L$), and a common approach is to impose that $\boldsymbol{\Psi}_0$ is lower triangular, so that the different shocks are identified from a timing restriction. This identifying scheme is popular in the case of monetary policy, where monetary shocks are assumed to only affect macro variables with a one period lag (Christiano, Eichenbaum and Evans, 1999).

In a partially identified model, one can impose a timing restriction for one shock only. In the case of a monetary model, this will amount to ordering the monetary policy variable last and imposing that $\boldsymbol{\Psi}_0$ has its last column filled with 0 except for the diagonal coefficient. The submatrix $\tilde{\boldsymbol{\Psi}}_0$ made of the first $(L-1)$ rows and $(L-1)$ columns of $\boldsymbol{\Psi}_0$ is then left unrestricted, apart from invertibility to ensure that equation (6) defines a unique shock vector $\boldsymbol{\varepsilon}_t$.

Narrative identification

In a narrative identification scheme, a series of shocks has been previously identified from narrative accounts. For that case, we can proceed as with the recursive identification, because the use of narratively identified shocks can be cast as a partial identification scheme.

We order the narratively identified shocks series first in $\boldsymbol{y}_t$, and we assume that $\boldsymbol{\Psi}_0$ 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.

In other words, we are assuming that the narrative shocks are contemporaneously correlated

[25] Other identification schemes, for instance long-run restrictions, are also straightforward to impose.

with the true monetary shocks and uncorrelated with other structural shocks. Note that this assumption is weaker than the common assumption (e.g., Romer and Romer, 2004) that the narratively identified shocks are perfectly correlated with the true monetary shocks. Our procedure allows the narrative shocks to contain measurement error, as long as the measurement error is independent of structural shocks.[26]

Sign restrictions

Set identification through sign restrictions consists in imposing sign-restrictions on the sign of the $\boldsymbol{\Psi}_k$ matrices, i.e., on the impulse response coefficients at different horizons. Again, because a FAIR model works directly with the moving average representation and the $\boldsymbol{\Psi}_k$ matrices, imposing sign-restrictions in FAIR is straightforward. This is in contrast with VARs where imposing sign-restrictions in a non-dogmatic fashion is non trivial, because the impulse-responses are non-linear transformations of the VAR coefficients (Baumeister and Hamilton, 2015).

One can impose sign-restrictions on only the impact coefficients (captured by $\boldsymbol{\Psi}_0$, which could be left as a free parameter in this case) and/or sign restrictions on the impulse response over a specific horizon (captured by the $\{a_n, b_n, c_n\}$ FAIR coefficients that model $\boldsymbol{\Psi}_k$).[27] Note that while imposing sign-restrictions is not trivial in VAR setting (see e.g., Baumeister and Hamilton, 2015)

Identification restrictions through priors

Because FAIR works directly with the structural moving-average representation, the parameters to be estimated can be interpreted as "features" of the impulse responses, and one could go beyond sign-restrictions and envision set identification through shape restrictions (see

[26]In fact, our approach is closely related to recent work on external instruments in macro-econometrics, where narrative shocks are used as external instruments to identify the structural VAR (Stock 2008, Mertens and Ravn, 2013) and thus *only* need to be correlated with the true shocks of interest and uncorrelated with other shocks. See Stock and Watson (2017) for a unifying treatment of the problem.

[27]To implement parameter restrictions on $\boldsymbol{\Psi}_0$ and/or $\{a_n, b_n, c_n\}$, we assign a minus infinity value to the likelihood whenever the restrictions are not met.

e.g., Lippi and Reichlin, 1994 for an early application of this idea). Using the insights from Baumeister and Hamilton (2015), one could implement shape restrictions through informative priors on the coefficients of $\mathbf{\Psi}_0$ and on the $\{a_n, b_n, c_n\}_{n=1}^N$ coefficients. For instance, one could posit priors on the location of the peak effect, posit priors on the persistence of the effect of the shock, among other possibilities.[28]

3.3 Estimation routine

To estimate our model, we use a Metropolis-within-Gibbs algorithm (Robert & Casella 2004, Haario et al., 2001) 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).

To initialize the Metropolis-Hastings algorithm in an area of the parameter space that has substantial posterior probability, we follow a two-step procedure: first, we estimate a standard VAR using OLS on our data set, calculate the moving-average representation, and we use the impulse response functions implied by the VAR as our starting point. More specifically, we calculate the parameters of our FAIR model to best fit the VAR-based impulse response functions.[29] Second, we use these parameters as a starting point for a simplex maximization routine that then gives us a starting value for the Metropolis-Hastings algorithm.

3.4 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 FAIR to

[28] See Plagborg-Moller (2016) for a related idea.

[29] Specifically, we set the parameters of our model (the a, b and c coefficients) to minimize the discrepancy (sum of squared residuals) between the two sets of impulse responses.

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 corresponding standard-deviations are set as follows: $\sigma_{ij,a} = 10$, $\sigma_{ij,b} = K$ and $\sigma_{ij,c} = K$ (recall that K is the length of the moving-average).[30] While there is clearly some arbitrariness in choosing the tightness of our priors, it is important to note that they are sufficiently loose to let us explore a large class of alternative specifications.[31] More generally, the use of informative priors is not critical for our approach, and we could have used improper uniform priors, but the use of proper priors allows us to compute posterior odds ratios, which are important to select the order of the moving-average and to compare different FAIR models.

3.5 Choosing N, the number of Gaussian basis functions

To choose N, the order of the FAIR model, we use posterior odds ratios (assigning equal probability to any two models) to compare models with increasing number of basis functions. We select the model with the highest posterior odds ratio.[32]

3.6 Fundamentalness

In a linear moving average model, different representations (i.e., different sets of coefficients and innovation variances) can exhibit the same first two moments, so that with Gaussian-distributed innovations, the likelihood can display multiple peaks, and the moving average model is inherently underidentified. Since a FAIR model works off directly with the moving-

[30]Going back to our intuitive interpretation of the three parameters of a Gaussian basis function in Section 2, note that these priors are very loose. This is easy to see for a and b. For c, recall that $c\sqrt{\ln 2}$ is the the half-life of the effect of a shock. If $c = K$, this already corresponds to very persistent impulse response functions, since $K\sqrt{\ln 2} = 38$ quarters.

[31]For our monetary policy application, we verified that the prior did not influence our conclusions by using uninformative priors: We estimated the asymmetric FAIR model with improper flat priors, and we obtained very similar results.

[32]This approach can be seen as analogous to the choice of the parameter lag in VAR models. While the Wold theorem shows that any covariance-stationary series can be written as a VAR(∞), one must select a finite lag order p that reasonable approximate the VAR(∞) (e.g., Canova, 2007). The usual approach is to use information criteria such as AIC and BIC, which is similar to our present approach. Just as in the case of lag length choice in a VAR (where this is rarely, if ever, done), we could alternatively treat N as a discrete parameter. We choose to use one value for N at a time to highlight how different choices for N affect estimated impulse responses.

average representation, it cannot distinguish between invertible (also called "fundamental") and non-invertible representations. By using the VAR-based impulse responses as starting values, we implicitly focus on the invertible part of the parameter space.[33,34]

3.7 Dealing with non-stationary data

As can be seen from Theorem 1, FAIR models can only capture impulse response functions that are bounded and integrable, which restricts our approach to stationary series. If the data are non-stationary, we can (i) allow for a deterministic trend in equation (1) and/or (ii) difference the data, and then proceed exactly as described above.

If a deterministic trend is suspected, we allow for a polynomial trend in each series, and we jointly estimate the parameters of the impulse responses (the $\boldsymbol{\Psi}_k$ coefficients) and the polynomial parameters.

If a stochastic trend is suspected, we can transform the data into stationary series by differencing the data. Importantly, the presence of co-integration does not imply that a FAIR model in first-difference is misspecified.[35] After estimation, one can even test for co-integration by testing whether the matrix sum of moving-average coefficients ($\sum_{k=1}^{K}\sum_{l=0}^{k}\boldsymbol{\Psi}_l$) is of reduced rank (Engle and Yoo, 1987).

4 FAIR for non-linear moving-average models

Since FAIR models work directly with the moving-average representation of the data, FAIR can easily allow for non-linear effects of shocks, in particular asymmetry.

[33]Since a VAR is obtained by inverting the fundamental moving-average representation, it automatically selects the fundamental representation (e.g., Lippi and Reichlin, 1994).

[34]An alternative estimation procedure to handle both invertible and non-invertible representations would be to use the Kalman filter with priors on the K initial values of the shocks $\{\boldsymbol{\varepsilon}_{-K}...\boldsymbol{\varepsilon}_0\}$, as recently proposed by Plagborg-Moller (2016). However, unlike our proposed approach, this procedure would be difficult to implement in non-linear models. Note also that the non-uniqueness of the moving average representation was proven for linear models (under Gaussian shocks). When we consider non-linearities, the non-uniqueness of the moving-average representation is not guaranteed anymore, and identification may be easier. In practice (and in Monte-Carlo simulations), the likelihood did not display multiple peaks when we allowed for asymmetry.

[35]The reason is that a FAIR model directly works with the moving-average representation and does not require inversion of the moving-average, unlike VAR models.

4.1 Introducing asymmetry

Specifically, with asymmetric effects of (in our case) monetary shocks, the matrix of impulse responses $\boldsymbol{\Psi}_k$ depend on the sign of the structural shock, i.e., we let $\boldsymbol{\Psi}_k$ take two possible values: $\boldsymbol{\Psi}_k^+$ or $\boldsymbol{\Psi}_k^-$, so that a model with asymmetric effects of shocks would be

$$\boldsymbol{y_t} = \sum_{k=0}^{K} \left[\boldsymbol{\Psi}_k^+ (\boldsymbol{\varepsilon_{t-k}} \odot \mathbf{1}_{\boldsymbol{\varepsilon_{t-k}} > \mathbf{0}}) + \boldsymbol{\Psi}_k^- (\boldsymbol{\varepsilon_{t-k}} \odot \mathbf{1}_{\boldsymbol{\varepsilon_{t-k}} < \mathbf{0}}) \right] \tag{7}$$

with $\boldsymbol{\Psi}_k^+$ and $\boldsymbol{\Psi}_k^-$ the lag matrices of coefficients for, respectively, positive and negative shocks and $\odot$ denoting element-wise multiplication.

Denoting $\psi_{ij}^+(k)$, the i-row j-column coefficient of $\boldsymbol{\Psi}_k^+$ (that is, the impulse response of variable j to a positive shock i), a FAIR model with asymmetry would then be

$$\psi_{ij}^+(k) = \sum_{n=1}^{N} a_{ij,n}^+ e^{-\left(\frac{k - b_{ij,n}^+}{c_{ij,n}^+}\right)^2}, \quad \forall k \in (0, K] \tag{8}$$

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

4.2 Bayesian estimation

The Bayesian estimation of FAIR models with asymmetric impulse responses proceeds similarly to linear case, but the construction of the likelihood involves one additional complication that we briefly mention here and describe in detail in the Appendix.

The additional complication comes from the fact that one must make sure that the system $\boldsymbol{\Psi_0}(\boldsymbol{\varepsilon_t}, \boldsymbol{z_t})\boldsymbol{\varepsilon_t} = \mathbf{u}_t$ has a unique solution vector $\boldsymbol{\varepsilon_t}$ given a set of model parameters and given some vector $\mathbf{u}_t$. With the contemporaneous impact matrix $\boldsymbol{\Psi_0}$ a function of $\boldsymbol{\varepsilon_t}$, a unique solution is a priori not guaranteed. However, we show in the Appendix that there is a unique solution when we allow the identified shocks to have with asymmetric effects in (i) the (full or partial) recursive identification scheme, (ii) the narrative identification scheme, and (iii) the sign-restriction identification scheme under the restriction that $sgn(\det \boldsymbol{\Psi}_{\mathbf{0}}^{+}) = sgn(\det \boldsymbol{\Psi}_{\mathbf{0}}^{-})$.

Compared to the linear case, the non-linear models require some initial values and prior distribution for the parameters controlling the non-linearities. As initial guesses, we set the parameters capturing asymmetry to zero (i.e., no non-linearity).[36] This approach is consistent with the starting point of this paper: structural shocks have linear effects on the economy, and we are testing this hypothesis against the alternative that shocks have some non-linear effects. We then center the priors for these parameters at zero with flat (but proper) priors.

4.3 Assessing FAIR performances from Monte-Carlo simulations

We conducted a number of Monte-Carlo simulations to study the performance FAIR in finite sample, first for linear models, then for asymmetric models. While we leave the details of these simulations for the appendix, we obtained two sets of results. First, in a linear data generating process, we found that a FAIR model can generate more accurate impulse response estimates (in a mean-squared error sense) than a well-specified VAR model. Second, for a data generating process with asymmetry, we found that FAIR can accurately detect the presence of asymmetries and deliver good estimates of the magnitudes of these non-linearities.

5 The asymmetric effects of monetary shocks

To study the non-linear effects of monetary shocks, we consider a model of the US economy in the spirit of Primiceri (2005), where $\boldsymbol{y}_t$ includes the unemployment rate, the PCE inflation rate and the federal funds rate. The data cover 1959Q1 to 2007Q4, and we exclude the latest recession where the fed funds rate was constrained at zero and no longer captured variations in the stance of monetary policy. When constructing the likelihood, we consider a moving-average model with $K = 45$, chosen to be large enough such that the lag matrix coefficients $\boldsymbol{\Psi}_k$ are

[36] An alternative would be to obtain initial estimates about possible non-linear effects. One option could be to combine Jorda's (2005) local projection method (which can accommodate non-linearities) with the structural shocks recovered from the VAR in order to get first estimates of the non-linear impulse responses.

close enough to zero for $k > K$.[37] For FAIR models, we leave the non-zero coefficients of the contemporaneous impact matrix $\mathbf{\Psi}_0$ as free parameters.

We now report results obtained for three different identification schemes: (i) a timing restriction, (ii) a narrative approach, and (iii) sign restrictions.

5.1 Recursive identifying restriction

We first identify monetary shocks through a timing restriction, and we assume that monetary policy affects the economy with a lag, so that the matrix $\mathbf{\Psi}_0$ has its last column filled with 0 except for the diagonal coefficient.

As a preliminary test and to illustrate the working of FAIR, we start by checking that a linear one-basis-function FAIR model performs well against a standard VAR with 4 lags. Table 1 reports the (log) marginal data densities for the FAIR model and for the VAR model, so that a model comparison can be readily obtained by computing the Bayes factor (obtained by taking the exponential of the difference in (log) marginal data densities) after positing equal priors for the two competing models. Encouragingly for our approach, Bayesian model comparison favors the more parsimonious one-basis-function FAIR model with a Bayes factor of about 400.

We now discuss the results from asymmetric FAIR models in which the impulse responses to monetary shocks depend on the sign of the shock. As shown in columns (3) to (5) of Table 1, Bayesian model comparison favors a two-basis-function FAIR model, and from now on we will report and discuss the results obtained using a two-basis-function FAIR model.

First, we note that Bayesian model comparison strongly favors a model with asymmetry in the impulse responses to monetary shocks: the (log) marginal data density of an asymmetric two-basis-function FAIR is respectively 20 log-points larger than the linear (symmetric) FAIR model and 25 log-points larger than the VAR model, which imply Bayes factors of respectively about 10^8 and 10^{11}.

To visualize the asymmetry present in the impulse responses, Figure 4 plots the impulse

[37] As a robustness check, we considered a higher moving-average lag-length with $K = 55$. Results were identical.

responses (in percentage points) of unemployment, the price level and the federal funds rate to a one standard-deviation monetary shock. The thick lines denote the impulse response functions implied by the posterior mode, and the error bands are the 5th and 95th posterior percentiles.[38] When comparing impulse responses to positive and negative shocks, it is important to 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).

Monetary shocks have very asymmetric effects: following a contractionary monetary shock, which represents a 70 basis points increase in the fed funds rate, unemployment increases by about 0.15 percentage points (ppt), whereas a (linear) VAR implies only a 0.10 ppt increase. In contrast, following an expansionary monetary shock (a 70 basis points decrease in the fed funds rate), the response of unemployment is small (a decline of 0.04 percentage points) and non-significantly different from zero. To assess the statistical significance of our result, Figure 7 plots the posterior distribution of the *peak* responses of unemployment and inflation for expansionary monetary shocks (x-axis) and contractionary shocks (y-axis). The dashed red line denotes identical peak responses, i.e., symmetric peak effects, so that the figure can be seen as a test of the difference in the peak responses: a posterior density lying above or below the red line indicates statistical evidence for asymmetric impulse responses. The top panels plot the posterior distributions using the recursive identification scheme, and shows strong evidence in favor of asymmetry with a posterior probability of 0.98 that the peak response of unemployment is larger following a contractionary shock than an expansionary shock. Complementary to Figure 7, Figure 8 plots the posterior distribution of the *cumulative* responses of unemployment following respectively a contractionary and an expansionary shock. Although the uncertainty range for cumulated impulse responses is necessarily larger than for

[38] To be specific, this figure and subsequent figures show paths of the moving average coefficients $\psi(k)$.

peak responses, the statistical evidence for asymmetric impulse responses is still strong with a 0.94 posterior probability that the cumulative response of unemployment is larger following a contractionary shock than following an expansionary shock.

The response of the price level also displays an interesting asymmetric pattern: the price level appears more sticky following a contractionary shock than following an expansionary shock. Looking at Figures 7 and Figure 8, the posterior probability that the peak responses or the cumulative responses differ with the sign of the shock are respectively 0.91 and 0.85. While this statistical evidence is more marginal, such an asymmetric response of inflation is what one would expect if downward price (or wage) rigidity was responsible for the asymmetric response of unemployment.[39]

We also find asymmetry in the response of the fed funds rate to a monetary shock, but it is relatively mild. A monetary shock generates a slightly more persistent increase in the fed funds rate than its expansionary counterpart. This can be seen in the bottom right panel of Figure 4 where the response of the fed funds rate is slightly more short-lived following an expansionary shock.[40]

5.2 Narrative identification

We now evaluate the presence of asymmetry using monetary shocks identified through the narrative approach by Romer and Romer (2004) and extended until 2007 by Coibion et al. (2012). As pointed out by Coibion (2012), the advantage of the narrative procedure is that one should be able to more precisely identify the effects of monetary shocks than with a relatively small model like the one considered above, since the Romer and Romer measure controls for

[39]The existence of downward wage rigidity is supported empirically by the scarcity of nominal wage cuts relative to nominal wage increases (e.g., Card and Hyslop, 1997).

[40]One way to gauge how much of the asymmetric response of unemployment can be explained by the asymmetric response of the fed funds rate is to proceed as in the government spending multiplier literature (e.g., Ramey and Zubairy, 2016) and to compute the total change in unemployment relative to the total change in the fed funds rate, that is to compute the multiplier $m = \sum_{k=0}^{K} \psi_k^U / \sum_{k=0}^{K} \psi_k^{ffr}$ for respectively positive and negative shocks. After "controlling" for the total change in the fed funds rate, the asymmetry is still present with $m^+ = .24 > m^- = .12$ with m^+ the multiplier associated with a contractionary shock (an increase in the fed funds rate) and m^- the multiplier associated with an expansionary shock.

much of the endogenous fluctuations in the interest rate as well as the Fed's information set.

We estimate an asymmetric two-basis-function FAIR model with 4 variables included in the following order: the Romer and Romer shocks, unemployment, inflation and the fed funds rate, and we posit that the contemporaneous matrix $\mathbf{\Psi}_0$ 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. Recall that this assumption is weaker than the common assumption (e.g., Romer and Romer, 2004) that the narratively identified shocks are perfectly correlated with the true monetary shocks, and that our procedure allows the narrative shocks to contain measurement error.

Figure 5 plots the asymmetric impulse responses to an innovation to the Romer and Romer shocks. Confirming our previous results, unemployment displays a very asymmetric response: there is no significant movement in unemployment following an expansionary shock, but there is a large increase following a contractionary shock. The posterior probabilities of asymmetric effects for unemployment are >0.99 for both the peak response and the cumulative response (Figures 7 and 8, middle panels).

While the evidence for asymmetric responses of inflation is again more marginal, the pattern is the same as with the recursive identification scheme: inflation reacts more strongly following an expansionary shock than following a contractionary shock. The posterior probability of asymmetric effects for inflation is 0.93 for the peak response and 0.87 for the cumulative response (Figures 7 and 8, middle panels).

5.3 Identification through sign restrictions

Finally, we evaluate the presence of asymmetry using monetary shocks identified through sign restrictions. We posit that monetary shocks are the only shocks that raise the fed funds rate and lower inflation. We use a one-basis-function FAIR model, so that the sign restrictions for inflation and the fed funds rate are imposed over the whole horizon.[41]

[41] Other identification schemes are possible, and a two-basis-function FAIR model would allow us to impose the sign restriction over a specific horizon. We also experimented with imposing the additional restriction that

As initial guess in our optimization routine, we use the structural impulse responses implied by a Cholesky ordering, and we use flat priors with $a \in [-10, 10]$ (as well as for the intercepts and the coefficients of $\mathbf{\Psi}_0$), $b \in [0, K]$ and $c \in [0, K]$.[42]

Figure 6 plots the asymmetric impulse responses to a monetary shock. Again, the evidence for asymmetry is strong: while a contractionary shock raises unemployment significantly, an expansionary shock generates a much smaller (and non-significant) change in unemployment. The posterior probabilities of asymmetric effects for unemployment are >0.99 for both the peak response and the cumulative response (Figures 7 and 8, bottom panels).

Consistent with our previous results, the response of the price level is asymmetric with a strong response following an expansionary shock but a weak response following a contractionary shock. This time however, statistical significance is very strong with posterior probabilities of asymmetric effects in the peak and cumulative responses of inflation of 0.98 and 0.96 (Figures 7 and 8, bottom panels).

To summarize, following a contractionary shock, quantities react, while following an expansionary shock, prices react; a pattern consistent with the presence of downward price (or wage) rigidity.

6 Conclusion

This paper proposes a new method to estimate the (possibly non-linear) dynamic effects of structural shocks by using Gaussian basis functions to approximate impulse response functions. We apply our approach to the study of monetary policy and find that the effect of a monetary intervention depends strongly on the sign of the intervention. A contractionary shock has a strong adverse effect on unemployment, larger than implied by linear estimates, but an expansionary shock has, on average, no significant effect on unemployment. While the evidence

the unemployment increases following a contractionary monetary shock. The estimated impulse responses were similar.

[42]The latter prior variance imposes that the effect of a shock can have a half-life as large as $K\sqrt{\ln 2} = 38$ quarters (recall $K = 45$ in our monetary application), which represents an extremely persistent impulse response.

for inflation is more uncertain, inflation displays a weaker response following a contractionary shock than following an expansionary shock, consistent with a Keynesian narrative whereby the asymmetry emerges (at least in part) out of downward price/wage rigidities.

Although this paper studies the asymmetric effects of monetary policy, Functional Approximations of Impulse Responses may be useful in many other contexts. In particular, FAIR can be used to estimate the non-linear effects of other important shocks where the existence of asymmetry and state-dependence remains an important and unresolved question; notably fiscal policy shocks (Auerbach and Gorodnichenko, 2012, Ramey and Zubairy, 2016) or credit supply shocks (Gilchrist and Zakrajsek, 2012). Moreover, the parametrization offered by FAIR models and the associated efficiency gains may be useful even for linear models, where the sample size is small and/or the data are particularly noisy.

References

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

[2] Angrist, J., Jorda O and G. Kuersteiner. "Semiparametric estimates of monetary policy effects: string theory revisited," Journal of Business and Economic Statistics, forthcoming, 2016

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

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

[5] Auclert, A, "Monetary Policy and the Redistribution Channel," Manuscript, May 2017

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

[7] Auerbach A and Y Gorodnichenko. "Fiscal Multipliers in Recession and Expansion." In Fiscal Policy After the Financial Crisis, edited by Alberto Alesina and Francesco Giavazzi, pp. 63–98. University of Chicago Press, 2013

[8] Barro, Robert J., "Unanticipated Money Growth and Unemployment in the United States," American Economic Review, LXVII, 101-15, 1977

[9] Baumeister, C. and J. Hamilton, "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," Econometrica, 83(5), 1963-1999, 2015

[10] Beaudry, Paul, and Gary Koop. "Do Recessions Permanently Change Output?" Journal of Monetary Economics 31 (1993), 149 - 63

[11] Bunn P., J. Le Roux, K. Reinold and P. Surico "The Consumption Response to Positive and Negative Income Shocks," CEPR Working Paper DP No. 11829, 2017

[12] Canova, F. "Methods for Applied Macroeconomic Research," Princeton University Press, 2007

[13] Canova, F. and G. De Nicolo, "Monetary Disturbances Matter for Business Fluctuations in the G-7", Journal of Monetary Economics, 49, 1131-1159, 2002

[14] Card, D. and D. Hyslop, "Does Inflation Greases the Wheels of the Labor Market?" in Reducing Inflation: Motivation and Strategy, C. Romer and D. Romer, eds., University of Chicago Press, 1997

[15] Casella, G. and R. L. Berger, Statistical Inference, Duxbury, 2002

[16] Christelis D., D Georgarakos, T. Jappelli, L. Pistaferri and M. van Rooij, "Asymmetric Consumption Effects of Transitory Income Shocks," CSEF Working Papers 467, Centre for Studies in Economics and Finance (CSEF), 2017

[17] Cover, J. "Asymmetric Effects of Positive and Negative Money-Supply Shocks," The Quarterly Journal of Economics, Vol. 107, No. 4, pp. 1261-1282, 1992

[18] Coibion, O. "Are the Effects of Monetary Policy Shocks Big or Small?," American Economic Journal: Macroeconomics, vol. 4(2), pages 1-32, April, 2012

[19] Coibion, O., Y. Gorodnichenko, L. Kueng, and J. Silvia, "Innocent Bystanders? Monetary Policy and Inequality in the US," NBER Working Papers 18170, National Bureau of Economic Research, Inc, 2012

[20] Christiano, L., M. Eichenbaum, and C. Evans. "Monetary policy shocks: What have we learned and to what end?," Handbook of Macroeconomics, volume 1, chapter 2, pages 65-148, 1999

[21] DeLong B and L. Summers, "How Does Macroeconomic Policy Affect Output?," Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution, vol. 19(2), pages 433-494, 1988

[22] Engle, R. and B. Yoo, "Forecasting and testing in co-integrated systems," Journal of Econometrics, vol. 35(1), pages 143-159, May 1987

[23] Faust, J., "The Robustness of Identified VAR Conclusions about Money", Carnegie-Rochester Series on Public Policy, 49, 207-244, 1998

[24] Fernandez-Villaverde, J. and J. Rubio-Ramirez, "Comparing dynamic equilibrium models to data: a Bayesian approach," Journal of Econometrics, vol. 123(1), pages 153-187, November, 2004

[25] Gilchrist, S and E Zakrajsek, "Credit Spreads and Business Cycle Fluctuations," American Economic Review, American Economic Association, vol. 102(4), pages 1692-1720, June 2012

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

[27] Hamilton, J. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," Econometrica 57, 357-384, 1989

[28] Hastie, T., R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*, Springer 2009

[29] Hubrich, K. and T. Terasvirta. "Thresholds and smooth transitions in vector autoregressive models," Advances in Econometrics "VAR Models in Macroeconomics, Financial Econometrics, and Forecasting", Vol. 31, 2013

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

[31] Jorgenson D., "Rational Distributed Lag Functions," Econometrica, 34 134-49, 1966

[32] Koop G, M. Pesaran and S. Potter "Impulse response analysis in nonliner multivariate models," Journal of Econometrics, 74 119-147, 1996

[33] Koop G and S. Potter "Dynamic asymmetries in US unemployment," Journal of Business & Economic Statistics Volume 17, Issue 3, 1999

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

[35] Lippi, M. & Reichlin, L. "VAR analysis, nonfundamental representations, Blaschke matrices," Journal of Econometrics 63(1), 307–325, 1994

[36] Lippi, M., Reichlin, L., "Diffusion of technical change and the decomposition of output into trend and cycle," Review of Economic Studies 61 (1) (206), 19–30, 1994b

[37] Lo, M. C., and J. Piger "Is the Response of Output to Monetary Policy Asymmetric? Evidence from a Regime-Switching Coefficients Model," Journal of Money, Credit and Banking, 37(5), 865–86, 2005

[38] Mertens K. and M. Ravn. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States," American Economic Review, American Economic Association, vol. 103(4), pages 1212-1247, June 2013

[39] Morgan, D. "Asymmetric Effects of Monetary Policy." Federal Reserve Bank of Kansas City Economic Review 78, 21-33, 1993

[40] Plagborg-Moller, P. "Bayesian Inference on Structural Impulse Response Functions," Working Paper, 2016

[41] Potter, S. "A non-linear approach to US GNP," Journal of Applied Econometrics," Vol 10 109-125, 1995

[42] Primiceri, G. "Time Varying Structural Vector Autoregressions and Monetary Policy," Review of Economic Studies, vol. 72(3), pages 821-852, 2005

[43] Ravn M and M Sola. "A Reconsideration of the Empirical Evidence on the Asymmetric Effects of Money-supply shocks: Positive vs. Negative or Big vs. Small," Archive Discussion Papers 9606, Birkbeck, 1996

[44] Ravn M. and M. Sola. "Asymmetric effects of monetary policy in the United States," Review, Federal Reserve Bank of St. Louis, issue Sep, pages 41-60, 2004

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

[46] Ramey V. and S. Zubairy. "Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data ," Journal of Political Economy, Forthcoming, 2016

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

[48] Romer, C., and D. Romer. "A New Measure of Monetary Shocks: Derivation and Implications," American Economic Review 94 (4): 1055–84, 2004

[49] Stock, J., What's New in Econometrics: Time Series, Lecture 7. Short course lectures, NBER Summer Institute, at http://www.nber.org/minicourse_2008.html, 2008

[50] Stock, J. and M. Watson, "Identification and Estimation of Dynamic Causal Effects in Macroeconomics," prepared for the Sargan Lecture to the Royal Economic Society, April 2017

[51] Tenreyro, S. and G. Thwaites."Pushing on a String: US Monetary Policy Is Less Powerful in Recessions." American Economic Journal: Macroeconomics, 8(4): 43-74, 2016

[52] Thoma, M. "Subsample Instability and Asymmetries in Money-Income Causality." Journal of Econometrics 64, 279-306, 1994

[53] Uhlig, H. "What are the effects of monetary policy on output? Results from an agnostic identification procedure," Journal of Monetary Economics, vol. 52(2), pages 381-419, March 2005

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

[55] Weise, C. "The Asymmetric Effects of Monetary Policy: A Nonlinear Vector Autoregression Approach," Journal of Money, Credit and Banking, vol. 31(1), pages 85-108, February 1999

Appendix A1: Proof of Theorem 1

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

$$\delta_\lambda(x) \equiv G_\lambda(x) = \frac{1}{\sqrt{2\pi\lambda^2}} e^{-\frac{x^2}{\lambda^2}}, \tag{9}$$

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.[43]

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

$$f_\lambda(x) = \int_{-\infty}^{+\infty} \delta_\lambda(x-u) f(u) du \tag{10}$$

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). ■

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

Using (9) in (10), the function f_λ given by

$$f_\lambda(x) = \int_{-\infty}^{+\infty} G_\lambda(x-u) f(u) du \tag{11}$$

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 (11) with a Riemann sum. To do so, first rewrite f_λ as

$$f_\lambda(x) = \underbrace{\int_{-\infty}^{-s} G_\lambda(x-u) f(u) du}_{=A(\lambda,x)} + \int_{-s}^{+s} G_\lambda(x-u) f(u) du + \underbrace{\int_{s}^{+\infty} G_\lambda(x-u) f(u) du}_{=B(\lambda,x)} \tag{12}$$

for $s > 1$.

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

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

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

Going back to (12), we have

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

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

$$\lim_{\lambda \to 0} |B(\lambda, x)| = 0. \tag{13}$$

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

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

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

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

$$\lim_{N\to\infty} f_{\lambda,N}(x) = \int_{-s}^{+s} G_\lambda(x-u)f(u)du. \tag{14}$$

Denoting $a_n = f(\xi_n)\left(\xi_n - \xi_{n-1}\right)$, $b_n = \xi_n$ and $c_n = \lambda$, using (14), (13) in (12) and combining with (11), we get that

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

which completes the proof.

Appendix A2: Identifying restrictions in non-linear Moving-Average models

We now detail how to impose the different identifying restrictions used in the paper. We only discuss the non-linear model $\boldsymbol{y_t} = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k(\boldsymbol{\varepsilon_{t-k}}, \boldsymbol{z_{t-k}})\boldsymbol{\varepsilon_{t-k}}$, since it includes the simpler linear model $\boldsymbol{y_t} = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k\boldsymbol{\varepsilon_{t-k}}$.

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 the main text, in order to construct the likelihood recursively, the system of equations

$$\boldsymbol{\Psi}_0(\boldsymbol{\varepsilon_t}, \boldsymbol{z_t})\boldsymbol{\varepsilon_t} = \mathbf{u}_t \tag{15}$$

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.[44]

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

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

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

$\mathbf{\Gamma}^{>i,>j}$ 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 $\boldsymbol{\varepsilon_t}$. 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 $\mathbf{\Psi_0}$ 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 with*

$$\boldsymbol{\Psi}_k(\boldsymbol{\varepsilon_{t-k}}) = \boldsymbol{\Psi}_k(\varepsilon_{\ell,t-k}) \tag{16}$$

$$= \left[\boldsymbol{\Psi}_k^+ 1_{\varepsilon_{\ell,t-k}>0} + \boldsymbol{\Psi}_k^- 1_{\varepsilon_{\ell,t-k}<0}\right], \ \forall k \in \{0,..,K\}, \ \forall t \in \{1,..,T\} \tag{17}$$

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

Proof. We first establish the following lemma:

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

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

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

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

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

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

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

$$\mathbf{u}_t = \boldsymbol{\Psi_0}(\varepsilon_{\ell,t})\boldsymbol{\varepsilon_t} \tag{18}$$

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

Notice that (18) implies the sub-system with ℓ equations

$$\mathbf{u}_t^{\leq\ell} = \begin{pmatrix} \mathbf{\Psi}_0^{<\ell,<\ell} & \mathbf{0}^{<\ell,1} \\ \mathbf{\Psi}_0^{\ell,<\ell} & \mathbf{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t}) \end{pmatrix} \boldsymbol{\varepsilon}_{\boldsymbol{t}}^{\leq\ell} \tag{19}$$

and notice that the matrix in (19) depends on $\varepsilon_{\ell,t}$ only through the scalar $\mathbf{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t})$. Denoting $\mathbf{A} \equiv \mathbf{\Psi}_{\mathbf{0}}^{<\ell,<\ell}$ a $(\ell-1)\times(\ell-1)$ invertible matrix (from Assumption 1), $\mathbf{C} \equiv \mathbf{\Psi}_{\mathbf{0}}^{\ell,<\ell}$ a $1\times(\ell-1)$ matrix, $\mathbf{B} \equiv \mathbf{0}$ of dimension $(\ell-1)\times 1$, and $D(\varepsilon_{\ell,t}) \equiv \mathbf{\Psi}_{0,\ell\ell}(\varepsilon_{\ell,t})$ the (ℓ,ℓ) coefficient of $\mathbf{\Psi}_0$ (a scalar), we can use Lemma 1 to invert the system (19) and obtain

$$\boldsymbol{\varepsilon}_{\boldsymbol{t}}^{\leq\ell} = \frac{1}{D(\varepsilon_{\ell,t})} \begin{pmatrix} D(\varepsilon_{\ell,t})\mathbf{A}^{-1} & \mathbf{0}^{<\ell,1} \\ -\mathbf{C}\mathbf{A}^{-1} & 1 \end{pmatrix} \mathbf{u}_t^{\leq\ell}. \tag{20}$$

The last row of (20) 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 (20). Then, with $\mathbf{\Psi}_{\mathbf{0}}^{<\ell,<\ell}$ and $\mathbf{\Psi}_{\mathbf{0}}^{>\ell,>\ell}$ invertible, (18) has a unique solution vector $\boldsymbol{\varepsilon}_{\boldsymbol{t}}$. ■

Proposition 1 ensures that the system (15) 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 $\boldsymbol{\varepsilon_t}$.[45]

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 (15) even when all shocks have asymmetric effects.

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 $\boldsymbol{\Psi}_0$ 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 (15) has a unique solution vector $\boldsymbol{\varepsilon}_t$ even when the narratively identified shocks has asymmetric and state dependent effects.

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.

We establish the conditions under which system (15) has a unique solution vector in a model with asymmetry:

Proposition 2 *Consider the asymmetric moving average model defined in (7) with*

$$\begin{aligned} \boldsymbol{\Psi}_k(\boldsymbol{\varepsilon_{t-k}}) &= \boldsymbol{\Psi}_k(\varepsilon_{t-k}^r) & (21) \\ &= \left[\boldsymbol{\Psi}_k^{+} 1_{\varepsilon_{t-k}^r>0} + \boldsymbol{\Psi}_k^{-} 1_{\varepsilon_{t-k}^r<0}\right], \ \ \forall k \in \{0,..,K\}, \ \forall t \in \{1,..,T\} & (22) \end{aligned}$$

[45] 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.

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 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 $sgn(\det \boldsymbol{\Psi}_0^+) = sgn(\det \boldsymbol{\Psi}_0^-)$.

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

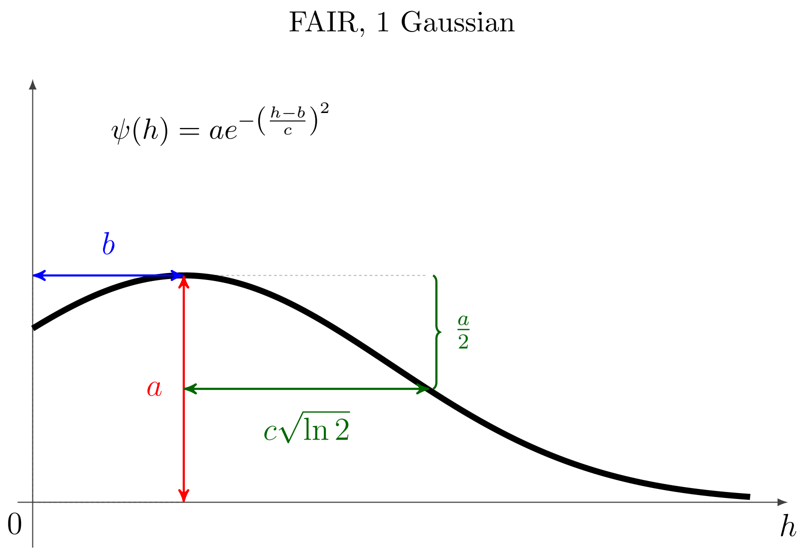
$$\varepsilon_t^r = \frac{1}{D(\varepsilon_t^r) - \mathbf{CA}^{-1}\mathbf{B}(\varepsilon_t^r)} \begin{pmatrix} -\mathbf{CA}^{-1} & 1 \end{pmatrix} \mathbf{u}_t, \tag{23}$$

which defines ε_t^r. Since the right hand side of (23) 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 (23) depends on ε_t^r only through the sign of $D(\varepsilon_t^r) - \mathbf{CA}^{-1}\mathbf{B}(\varepsilon_t^r)$.[46] 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 (23) 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 $\mathbf{u}_t = \boldsymbol{\Psi}_{\mathbf{0}}(\varepsilon_t^r)\boldsymbol{\varepsilon_t}$ has a unique solution vector. ■

Proposition 2 states that the system $\mathbf{u}_t = \boldsymbol{\Psi}_{\mathbf{0}}(\varepsilon_t^r)\boldsymbol{\varepsilon_t}$ 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 $\boldsymbol{\Psi}_0(\boldsymbol{\varepsilon_t})\boldsymbol{\varepsilon_t}$.

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



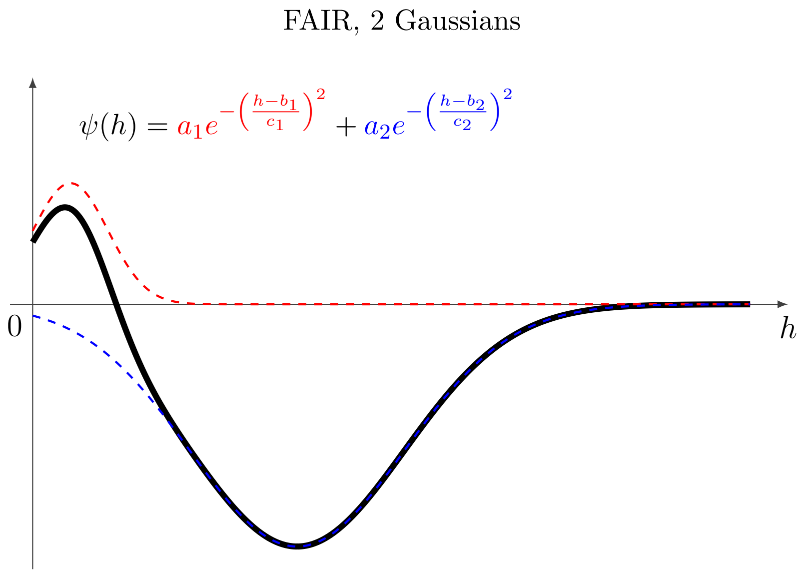


Figure 1: **Functional Approximation of Impulse Responses (FAIR)**: Examples with one Gaussian basis function (top panel) or two Gaussian basis functions (bottom panel).

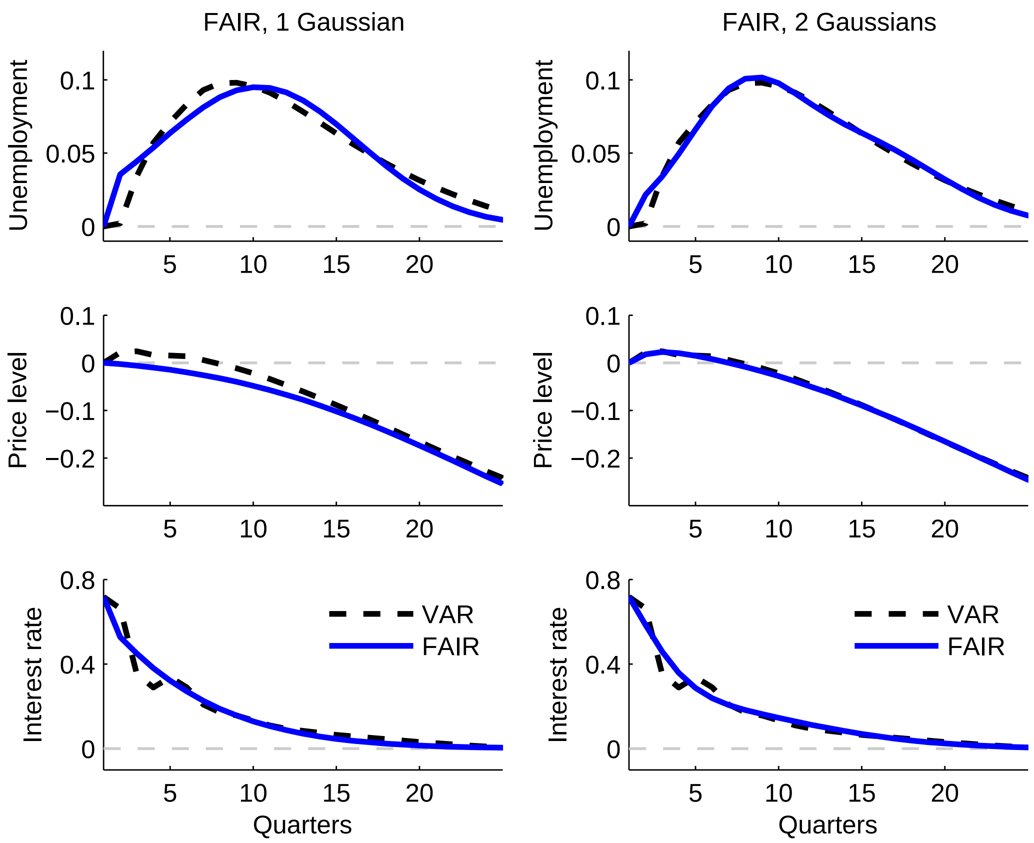


Figure 2: Impulse response functions of the unemployment rate (in ppt), the (log) price level (in percent) and the federal funds rate (in ppt) to a one standard-deviation monetary shock. Impulse responses estimated with a VAR (dashed-line) or approximated using one Gaussian basis function (FAIR, left-panel, thick line) or two Gaussian basis functions (FAIR, right panel thick line). Estimation using data covering 1959-2007.

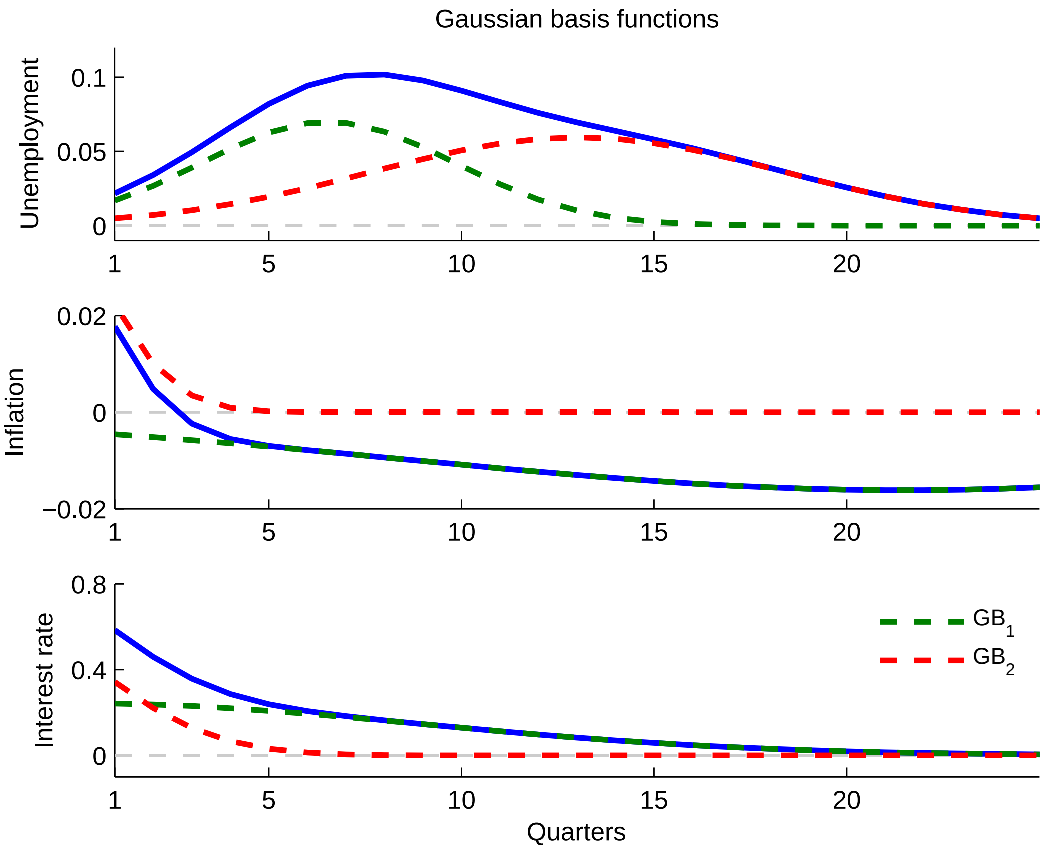


Figure 3: Gaussian basis functions (dashed lines) used by a FAIR to approximate the responses of unemployment, inflation and the fed funds rate to a monetary shock. The basis functions are appropriately weighted so that their sum gives the functional approximation of the impulse response functions (solid lines) reported in the right-panels of Figure 1.

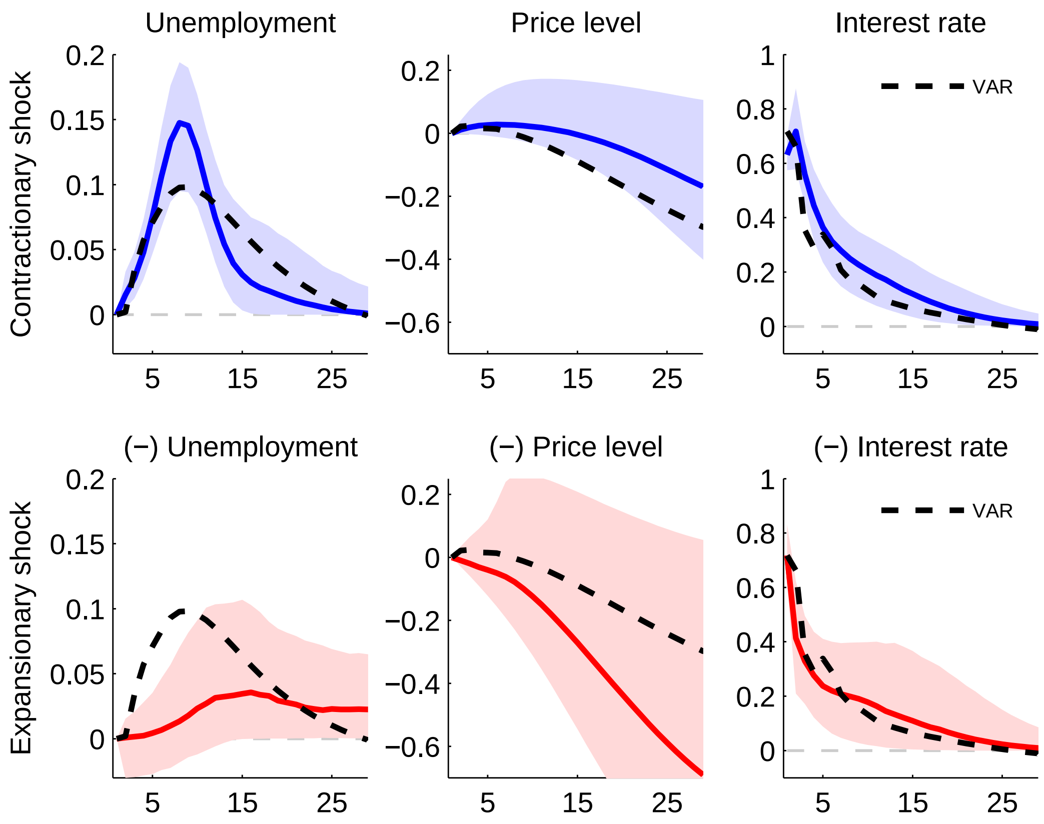


Figure 4: **Recursive identification scheme, FAIR, 1959-2007**: Impulse response functions of the unemployment rate (in ppt), the (log) price level (in percent) and the federal funds rate (in ppt) to a one standard-deviation monetary shock identified from a recursive ordering. Estimation from a VAR (dashed-line) or 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.

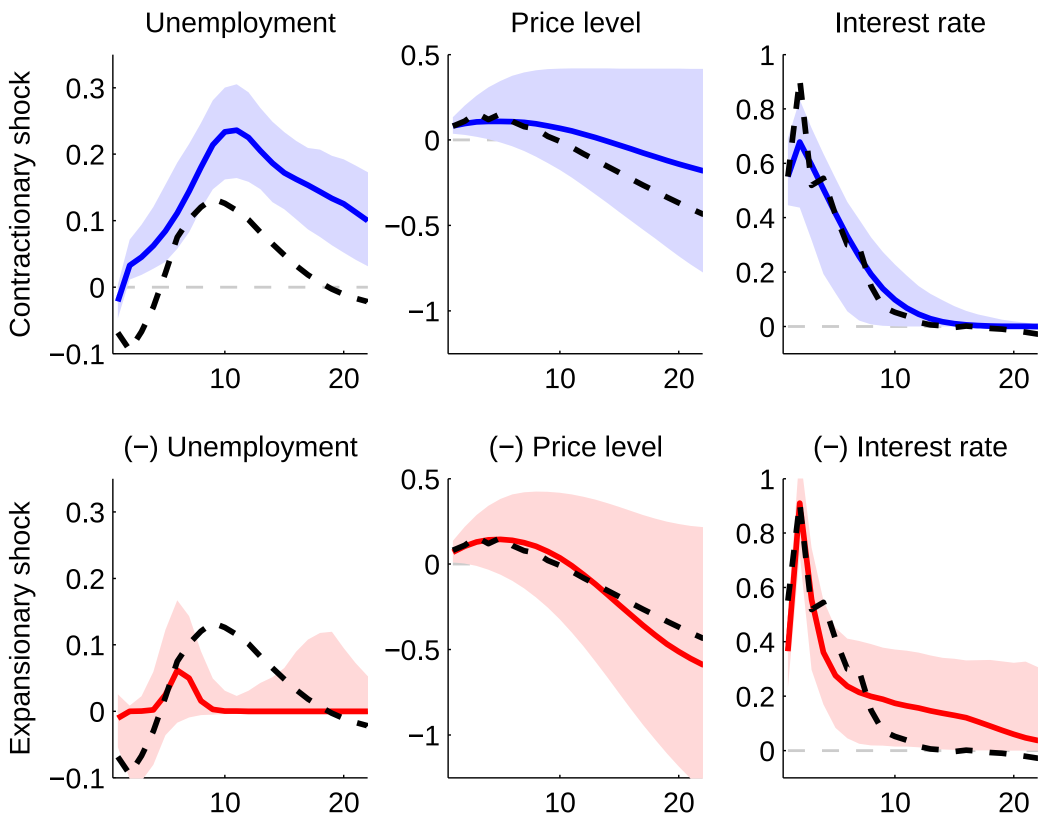


Figure 5: **Narrative identification scheme, FAIR, 1966-2007**: Impulse response functions of the unemployment rate (in ppt), the (log) price level (in percent) and the federal funds rate (in ppt) to a one standard-deviation Romer and Romer monetary shock. Estimation from a VAR (dashed-line) or from 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.

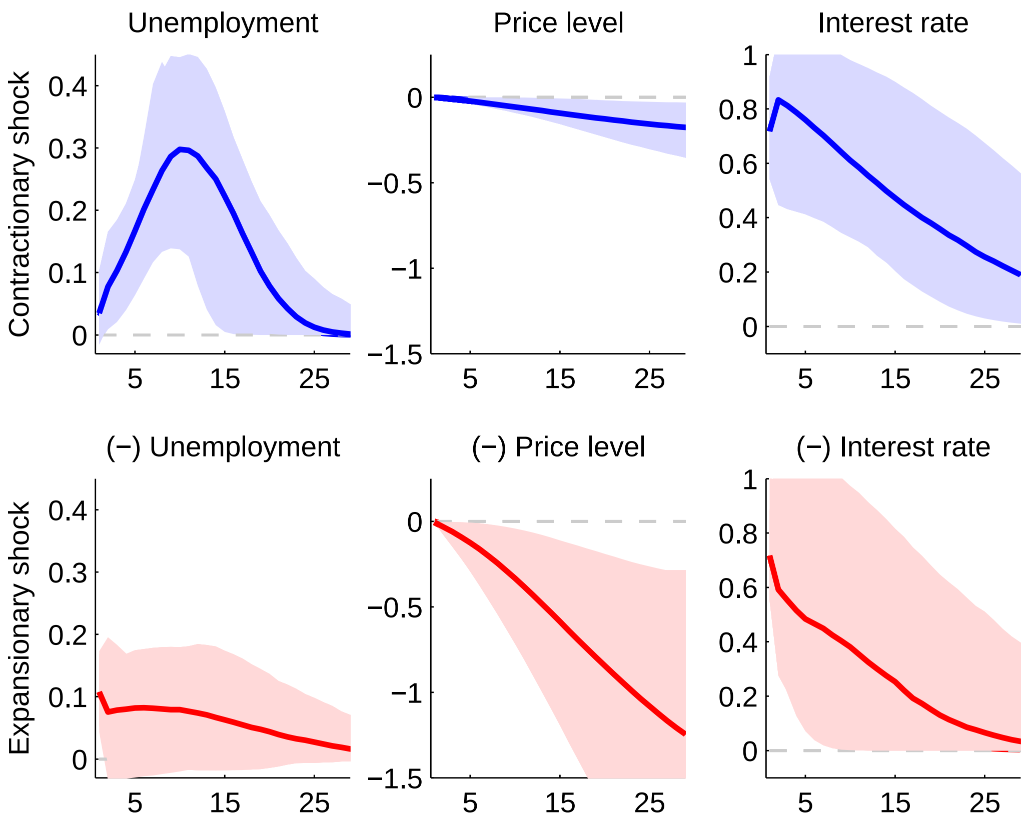


Figure 6: **Sign-restrictions identification scheme, FAIR, 1959-2007**: Impulse response functions of the unemployment rate (in ppt), the (log) price level (in percent) and the federal funds rate (in ppt) to a one standard-deviation monetary shock identified with sign restrictions. Estimation from a 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.

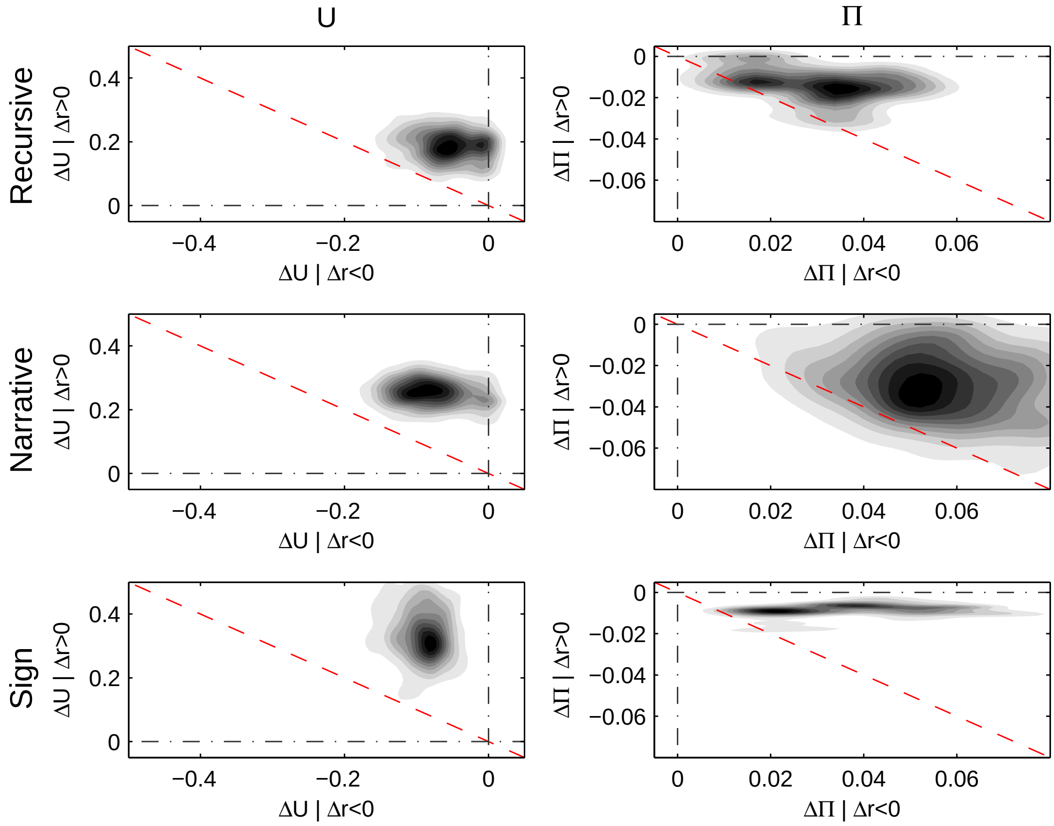


Figure 7: Posterior distribution of the peak responses of Unemployment (ΔU, left panel) and Inflation (ΔΠ, right panel) to a monetary shock (Δr). The dashed red line denotes symmetric peak responses. Results from a recursive identification scheme over 1959-2007 ("Recursive", top panel), a narrative identification scheme over 1966-2007 ("Narrative", middle panel), and a set identification scheme with sign restrictions over 1959-2007 ("Sign", bottom panel).

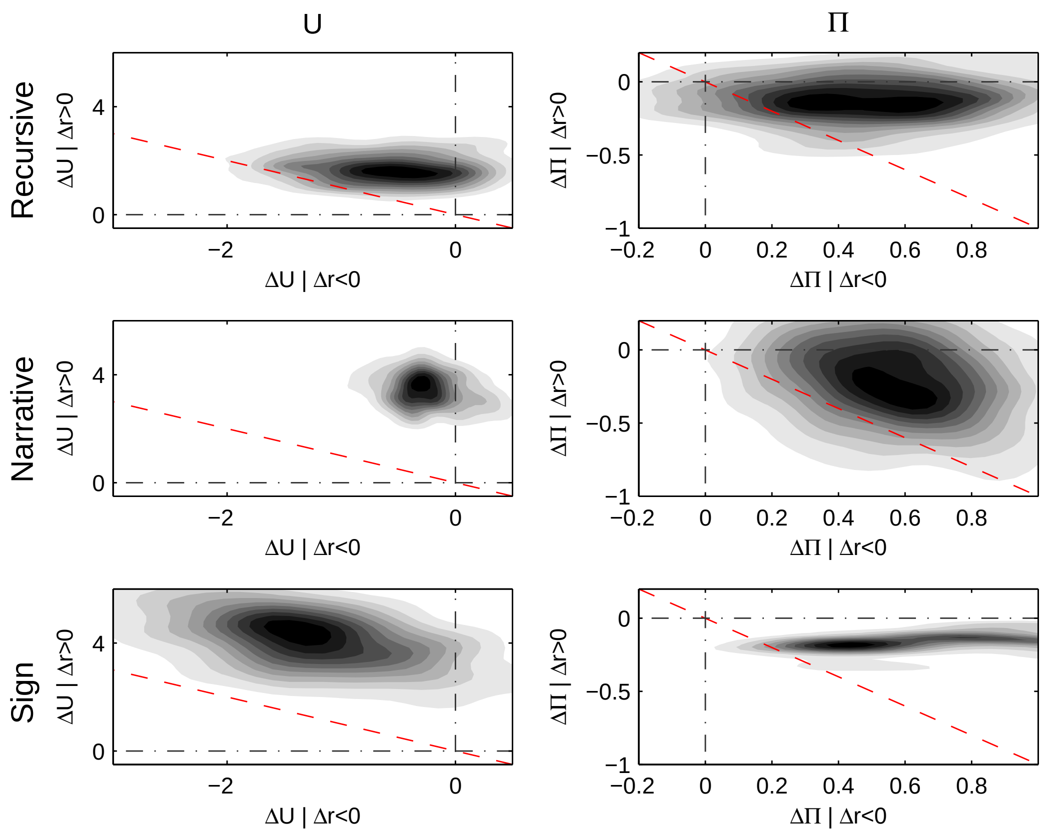


Figure 8: Posterior distribution of the cumulative responses of Unemployment (ΔU, left panel) and Inflation (ΔΠ, right panel) to a monetary shock (Δr). The dashed red line denotes symmetric cumulative responses. Results from a recursive identification scheme over 1959-2007 ("Recursive", top panel), a narrative identification scheme over 1966-2007 ("Narrative", middle panel), and a set identification scheme with sign restrictions over 1959-2007 ("Sign", bottom panel).

Table 1: Marginal data densities, recursive identification

	VAR	FAIR(1)	FAIR(1) Asymmetry	FAIR(2) Asymmetry	FAIR(3) Asymmetry
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
(log) marginal data density	112	118	127	138	107

Note: Trivariate models with unemployment, PCE inflation and the fed funds rate estimated over 1959-2007. The VAR estimates and confidence bands are obtained from a Bayesian VAR with Normal-Whishart priors. FAIR(N) demotes a FAIR model with N Gaussian basis functions.