VAR Information and the Empirical Validation of DSGE Models

Mario Forni*
Università di Modena e Reggio Emilia, CEPR and RECent

Luca Gambetti†
Universitat Autonoma de Barcelona and Barcelona GSE

Luca Sala‡
Università Bocconi, IGIER and Baffi CAREFIN

Abstract

A shock of interest can be recovered, either exactly or with a good approximation, by means of standard VAR techniques even when the structural MA representation is noninvertible or nonfundamental, possibly because it has more shocks than variables. We propose a measure of how informative a VAR model is for a specific shock, or a subset of shocks, of interest. We show how to use such a measure for the validation of shocks' transmission mechanism of DSGE models through VARs. In an application, we validate a theory of news shocks. The theory does remarkably well for all variables, except for consumption and output, for which the model over-predicts the effects of news shocks.

JEL classification: C32, E32.

Keywords: Invertibility, news shocks, DSGE model, structural VAR.

---

*The financial support ... is gratefully acknowledged. Contact: Dipartimento di Economia Politica, via Berengario 51, 41100, Modena, Italy. Tel. +39 0592056851; e-mail: mario.forni@unimore.it

†The financial support ... is gratefully acknowledged. Contact: Office B3.1130 Departament d'Economia i Historia Economica, Edifici B, Universitat Autonoma de Barcelona, Bellaterra 08193, Barcelona, Spain. Tel. +34 935814569; e-mail: luca.gambetti@uab.cat

‡The financial support ... is gratefully acknowledged. Corresponding author. Contact: Department of Economics, Universitá Bocconi, Via Roentgen 1, 20136, Milan, Italy. Tel. +39 0258363062; e-mail: luca.sala@unibocconi.it

1 Introduction

Any theoretical model should in principle be validated by evaluating whether its implications are consistent with empirical facts.

DSGE models represent the most popular class of models in theoretical macroeconomics. They are widely used to study the propagation mechanisms of economic shocks and address policy-relevant questions. However, they often rely on arbitrary theoretical assumptions, which are hard to judge or evaluate form an empirical point of view. Maximum likelihood or Bayesian estimation methods provide the best constrained fit of the data; however it might still be the case that the theory fails in fitting the data satisfactorily (Sala, 2015).

On the other hand, structural VAR models have been for long time now the main tool in applied macroeconomic research. Structural VARs are, to a large extent, free of restrictions derived form economic theory. It is therefore a quite natural idea to use VAR models for the empirical validation of DSGE models, in the spirit of Sims (1989). A prominent example of the use of VAR models for validation purposes is represented by the technology-hours debate, where the empirical response of hours to technology shocks has been used to assess RBC and sticky prices models (Galì, 1999; Christiano, Eichenbaum and Vigfusson, 2004). The topic is extensively discussed in Canova (2002, 2007), Christiano, Eichenbaum and Vigfusson (2007), Chari, Keohe and McGrattan (2008), Giacomini (2013).

Validation through VARs is typically carried out by comparing the unrestricted VAR impulse response functions —let us say the "empirical" impulse response functions— with the constrained ones stemming from the DSGE model— let us call them the "theoretical" impulse response functions (even if they may result from estimation). A crucial problem with this procedure is: Does the VAR specification employed convey enough information to estimate the shocks of interest, *under the null hypothesis that the model is true*? If this is not the case, the VAR and the DSGE model are incompatible and the comparison is inconclusive, whatever the outcome might be. The key question is therefore: how can we know whether a VAR specification is deficient, according to the theoretical model?

In this paper we propose a measure of the informational deficiency of a given VAR specification with respect to any particular shock. This measure, which we call δ_i, tells us the best that we can do in approximating the shock of interest u_{it} by means of a VAR. It is defined as the fraction of unexplained variance of the orthogonal projection of u_{it} onto the VAR residuals. It can be computed for any DSGE model, endowed with a set of values for the parameters. It takes on values between zero and one; $\delta_i = 0$ means perfect information for; $\delta_i = 1$ means no information. If $\delta_i = 0$ we say that the VAR is *informationally sufficient* for u_{it}. A VAR which is sufficient, or approximately sufficient, can be used for model validation.

To better clarify the practical use of our measure, let us explicit, step by step, the validation procedure we have in mind. First, consider the log-linear equilibrium representation of a calibrated/estimated DSGE model. Second, compute the measure for a particular VAR specification (whose variables form a subset of the variables modeled in the DSGE). For the sake of simplicity, let us focus on just a single shock u_{it}. Then use the following criterion, which can be named "δ-criterion". If δ_i is larger than a pre-specified threshold level, say 0.1, reject the specification and choose another vector of observables. If it is smaller, estimate the VAR (with real data) and identify the shocks of interest using restrictions consistent with the theoretical model. Finally, verify whether the theoretical impulse response functions lie within the confidence bands obtained with the VAR. If they do, the model is validated. If they do not, there is something wrong with either the parameter calibration/estimation or the model itself.

The concept of "sufficient information" used here is due to Forni and Gambetti, (2014). It is a generalization of the concept of "nonfundamentalness", or "noninvertibility" (Lippi and Reichlin, 1993), which has been widely debated in the recent literature.[1] Precisely, the structural representation of the variables included in the VAR is fundamental if and only if the VAR is informationally sufficient for all of the structural shocks. Correspondingly, the deficiency measure proposed in the present paper can be regarded as a generalization of existing fundamentalness conditions, including the well-known "Poor Man's Condition" of Fernandez-Villaverde et al. (2007) (PMC hereafter). The generalization works in two dimensions: (a) our measure is shock-specific; (b) it provides information about the "degree" of nonfundamentalness.

The validation procedure described above can in principle be performed by using the PMC in place of the δ-criterion. So what is the usefulness of our measure? Why not using instead the PMC (or other existing fundamentalness conditions)? The answer is that the PMC is extremely and unnecessarily restrictive. If the number of variables is smaller than the number of shocks, we cannot have fundamentalness and the PMC is not even defined. Since modern DSGE models have several shocks, small VARs, which are a useful tool, especially for short samples, are automatically ruled out. Second, considering specifications including as many variables as shocks, the existence of a specification satisfying the PMC cannot be taken for granted. Under fiscal foresight, or in presence of news shocks, nonfundamentalness is endemic (Leeper, Walker and Yang, 2013; Sims, 2012). For instance, in the news-shock DSGE model

[1] Early papers are Hansen and Sargent (1991) and Lippi and Reichlin (1993, 1994b). A partial list of recent papers includes Giannone and Reichlin (2006), Fernandez-Villaverde, Rubio-Ramirez, Sargent and Watson (2007), Ravenna (2007), Yang (2008), Forni, Giannone, Lippi and Reichlin (2009), Mertens and Ravn (2010), Sims (2012), Leeper, Walker and Yang (2013), Forni, Gambetti, Lippi and Sala (2013a, 2013b), Forni, Gambetti and Sala (2014), Forni and Gambetti (2014), Beaudry and Portier (2015).

studied here, which has seven shocks, there are no seven-variable VAR specifications satisfying the PMC.

On the other hand, the researcher is often interested in assessing the transmission mechanisms of a single shock. We show below that there are economically interesting examples in which a VAR is informationally sufficient for a single shock but not for all shocks, i.e. we do not have fundamentalness, but we have what may be called "partial fundamentalness". In this cases the PMC is unnecessarily restrictive, in the sense that the VAR is perfectly informative for the shock of interest even if the PMC does not hold. In addition, VAR deficiency, though different from zero, may be very small. In these cases the VAR performs reasonably well, despite nonfundamentalness. Indeed, what is really important for a reliable VAR analysis is not whether we have fundamentalness or not, but how much information the VAR specification conveys for the shocks of interest.

The fact that a VAR might perform well despite nonfundamentlness has first been observed by Sims (2012) and further documented in Beaudry and Portier (2015). In this paper we show that the performance of a VAR in recovering u_{it} and the related impulse response functions is closely related to our deficiency measure. Our work is also related to Soffritti (2015), which (independently of us) proposes a fundamentalness measure. The main difference with respect to ours is that it is global, rather than shock-specific, and, like PMC, is only defined for square systems, having as many variables as shocks.

Nonundamentalness is not the only one problem for VAR validation of DSGE models. Two additional problems are worth mentioning. First, most DSGE models are non-linear and linear representations result from approximation. In the present paper we do not address this problem; rather, we get rid of it by assuming an exact linear representation for the economic variables. Second, VAR estimation might entail a large truncation bias, as stressed within the present context in Chari, Kehoe and McGrattan (2008). Although lag truncation is not the main focus here, we discuss a natural extension of our deficiency measure to the case of the K-order VAR —δ_i^K. Finite-order VAR deficiency can be useful in practice, in that it provides a lower bound for the total bias due to nonfundamentalness and lag truncation.

In the application, we test a theory of news shocks. The model is a New-Keynesian DSGE, similar to that used by Blanchard, Lorenzoni and L'Huillier (2013). It features several frictions, such as internal habit formation in consumption, adjustment costs in investment, variable capital utilization, Calvo price and wage stickiness. It includes seven exogenous sources of fluctuations, a news permanent shock and a surprise temporary shock in technology, an investment-specific shock, a monetary policy shock, a shock to price markups, a shock to wage markups and a shock to government expenditures. We find that the PMC is not satisfied for a VAR including TFP, GDP, consumption, investment, hours, interest rate and inflation, confirming that news shocks, as already stressed in the literature, generate a problem of non-invertibility. However, our measure for the news shock in that VAR is zero, implying that the VAR is informationally sufficient for the news shock. To identify, consistently with the model, we impose the restriction that the news shock is the only one driving total factor productivity in the long run. Then we estimate the news shocks and the corresponding impulse response functions (using real data). Results show that the theory does remarkably well. Indeed the theoretical impulse response functions lie within the VAR bands for almost all variables, the only exceptions being consumption and output, for which the model overestimates the effects of news shocks.

The remainder of the paper is organized as follows. Section 2 introduces the theoretical framework and the measure, shows some examples and discusses the possible applications. Section 3 is devoted to the empirical application. Technical results are shown in Section 4. Section 5 concludes. The Appendix reports the details of the DSGE model used in Section 3.

2 Theory: Main ideas and two examples

2.1 Informational deficiency and the deficiency measure

As observed above, a relevant problem for the validation of DSGE models through VARs is that most DSGE models are non-linear. The theoretical impulse response functions result from a linear approximation which in principle may be inaccurate. Since our focus is on informational deficiency, we do not address this issue here. Rather, we rule out the problem by assuming directly that the macroeconomic variables in the model have an exact Moving Average representation (possibly derived from a state-space representation), where the structural shocks propagates through linear impulse response functions. As a consequence, our results hold true for any theoretical model which can be cast in the MA form (not necessarily general equilibrium models).

Let us focus on the section of the macroeconomic model corresponding to the variables used in the VAR, i.e. the entries of the n-dimensional vector x_t. We assume that the vector x_t, possibly after transformation, has the Moving Average (MA) representation

$$x_t = \sum_{k=0}^{\infty} A_k u_{t-k} = A(L)u_t, \tag{1}$$

where $u_t = (u_{1,t} \ \cdots \ u_{q,t})'$ is a q-dimensional white noise vector of mutually orthogonal macroeconomic shocks, and $A(L) = \sum_{k=0}^{\infty} A_k L^k$ is an $n \times q$ matrix of square-summable impulse response functions.

Representation (1) is "structural" in the sense that the vector u_t includes all of the exogenous shocks driving x_t. However, we do not assume that all of the shocks in u_t have a structural economic interpretation: some of them may be statistical residuals, devoid of economic interest, which arise from measurement errors. This enables us to evaluate VAR deficiency —and therefore its performance— with respect to the shocks of interest when actual variables are affected by errors of a pre-specified size.

We do not assume that the number of variables n is equal to the number of shocks q. In other words, representation (1) is not necessarily square. In particular, it can be "short", with more shocks than variables, $q > n$. Short systems are relevant for applied work for two reasons. First, several empirical analyses are based on small-scale VARs, with just two or three variables. If the economy is driven by a larger number of shocks, the above MA system will be short. Second, most variables are in practice affected by measurement errors and/or small shocks of limited economic interest, so that, even if we have as many variables as major structural shocks (or even more variables than shocks) the system may be short because measurement errors are included in the vector u_t.[2]

Given model (1), we want to evaluate whether a VAR in x_t conveys the information needed to recover the shocks of interest and the corresponding impulse response functions. In practice, the impulse response function obtained with a VAR are affected by estimation errors arising from the finiteness of the sample size. Such finiteness requires specification of low-order VARs, which might be affected by truncation bias. Since our main focus here is on the nonfundamentalness bias, we replace finite-sample, finite-order VARs with orthogonal projections on infinite-dimensional information spaces.

Within this conceptual framework, the first step of the standard structural VAR procedure is to compute the orthogonal decomposition

$$x_t = P(x_t|H^x_{t-1}) + \epsilon_t, \tag{2}$$

where H^x_t is the closed linear space $\overline{\text{span}}(x_{1,t-k}, \dots, x_{n,t-k}, k = 0, \dots, \infty)$ and ϵ_t is the Wold innovation.

The second and final step of the structural VAR procedure is identification, whose goal is to approximate the shock of interest, say u_{it}, as a linear combination of the entries of ϵ_t. The standard practice is to find this combination by imposing restrictions derived from economic theory. The quality of the result depends on the particular restrictions chosen. We shall focus

[2] "Tall" systems, i.e. systems with more variables than shocks, are also interesting from a theoretical point of view, but are unlikely to occur in practice, because of measurement errors. We shall not consider them further in the present work.

here on the best possible result, which is given by the projection of u_{it} onto the entries of ϵ_t:

$$u_{it} = M\epsilon_t + e_{it}. \tag{3}$$

The variance of the above residuals measures the approximation error. Deficiency is defined as the fraction of unexplained variance in the above projection:

$$\delta_i = \sigma^2_{e_i}/\sigma^2_{u_i}. \tag{4}$$

For simplicity of notation we do not explicit the dependence of δ_i on x_t. The deficiency measure can be computed from the theoretical model (1), that is from $A(L)$, according to the formula provided in Section 4.

We say that x_t is informationally sufficient for u_{it} if and only if u_{it} is an exact linear combination of the entries of ϵ_t, i.e. $\delta_i = 0$. In Section 4 we show that projecting u_{it} onto the entries of ϵ_t is equivalent to projecting it onto the VAR information set H^x_t (so that the two-step structural VAR procedure is able to produce the optimal approximation, given the available information). It follows that we have sufficiency for u_{it} if and only if $u_{it} \in H^x_t$.

By the very definition of informational sufficiency, if the identification scheme is correct, in the sense that it produces the identification vector M, then a sufficient VAR delivers u_{it} without error, whereas a deficient VAR produces an approximation, whose error is measured by δ_i.

However, the ultimate goal of the validation procedure are the impulse response functions, rather than the shock u_{it} itself. In Section 4 we show that a VAR, which is sufficient for u_{it} and correctly identified (but possibly deficient for the other structural shocks), delivers the correct impulse response functions. Hence δ_i provides a meaningful indication about the performance of the theoretical VAR in approximating the impulse response functions.

Of course, in practical situations we have finite-sample, finite-order VARs, which are affected by estimation and lag-truncation errors. Such VARs, however, will provide estimates whose bias converges to δ_i, as the sample size and the truncation lag increase at appropriate rates.

Although the truncation bias is not our focus here, the deficiency measure can be naturally extended to the case of finite-order VARs. Deficiency of a VAR(K) with respect to u_{it}, denoted by δ_i^K, is given by the fraction of unexplained variance of the projection of u_{it} onto the truncated VAR information space spanned by present and past values of the x's, until the maximum lag K. The sequence δ_i^K is nonincreasing in K. The difference $\delta_i^K - \delta_i$ tells us the additional bias due to lag truncation, as far as estimation of u_{it} is concerned. Unfortunately, this is not true for the impulse response functions (see Section 4 for details). Despite this, δ_i^K can prove useful in the context of a validation exercise, in that it is a lower bound for the total bias due to the cumulated effects of nonfundamentalness and lag truncation.

2.2 Beyond the Poor Man's condition

How does our defiency measure relate to fundamentalness and existing fundamentalness conditions?

We have fundamentalness when all of the shocks in u_t belong to the econometrician information set, i.e. $u_{it} \in H_t^x$, for all i. Hence we have fundamentalness if, and only if, the VAR is informationally sufficient for all shocks, that is $\delta_i = 0$ for all i. Sufficient information is then a notion of "partial fundamentalness", a straightforward shock-specific generalization of the fundamentalness concept.

Note that short systems are never fundamental (see Section 4, Proposition 1). This is quite intuitive: if we have just n variables we cannot estimate consistently more than n orthogonal shocks. Square system, as is well known, can be either fundamental or not, depending on the roots of the determinant of $A(L)$: we have fundamentalness if there are no roots smaller than 1 in modulus.

Fernandez-Villaverde et al. (2007) proposes a fundamentalness condition, the PMC, based on the state-space representation of the economy. Consider the following linear equilibrium representation of a DSGE model

$$
\begin{aligned}
s_t &= As_{t-1} + Bu_t && (5)\\
x_t &= Cs_{t-1} + Du_t && (6)
\end{aligned}
$$

where s_t is an m-dimensional vector of stationary "state" variables, x_t is the n-dimensional vector of variables observed by the econometrician, and u_t is the q-dimensional vector of shocks with $q \leq m \leq n$. A, B, C and D are conformable matrices of parameters, B has a left inverse B^{-1} such that $B^{-1}B = I_q$. Representation (5)-(6) can always be cast in form (1). If the matrix D is square and invertible, the matrix $A(L)$ apparing in representation (1) can be written as

$$A(L) = DB^{-1}\left[I - (A - BD^{-1}C)L)\right](I - AL)^{-1}B. \quad (7)$$

The PMC is that all the eigenvalues of the matrix $A - BD^{-1}C$ are strictly less than one in modulus. It is easily seen that, if the PMC holds, the MA representation of x_t is invertible and u_t can be represented as a linear combination of the present and past values of x_t. In other words, the PMC implies fundamentalness, i.e. sufficient information for all shocks.[3] Hence, if the system is square, and the PMC holds, then $\delta_i = 0$ for all i.

Summing up, the deficiency measure can be regarded as a generalization of the PMC (as well as other existing fundamentalness conditions), both because it is shock specific and because it provides information about the "degree" of non-fundamentalness. This generalization is very

[3] About the converse implication see Franchi and Paruolo (2015).

important for applied work. As anticipated above, short systems are never fundamental (and the PMC is not even defined in this case). By contrast, we show below that we may have small deficiency, or even sufficiency, for a single shock of interest, when $q > n$. This implies a relevant fact which is not well known in the literature: small-scale VARs can in principle be successfully employed even when the number of shocks driving the economy is large.

As for square systems, let us stress that non-fundamental structural MA representations, far from being an oddity, are common in macroeconomic models. They may arise from slow diffusion of technical change (Lippi and Reichlin, 1994b), news shocks (Sims, 2012, Forni, Gambetti and Sala, 2014, Beaudry and Portier, 2015), fiscal foresight (Leeper, Walker and Young, 2013), noise shocks (Forni, Lippi, Gambetti and Sala (2013a, 2013b)). The seven-variable DSGE model of Section 3, for instance, produces a non-fundamental representation for our seven-variable VAR specification. As a consequence, the PMC does not hold. Despite this, the VAR is sufficient for the news shock (and approximately sufficient for most shocks).

Sims (2012) shows that a VAR may perform reasonably well even if fundamentalness does not hold. With his words, non-fundamentalness "should not be thought of as an "either/or" proposition —even if the model has a non-invertibility, the wedge between VAR innovations and economic shocks may be small, and structural VARs may nontheless perform reliably" (Sims, 2012, abstract). Both Beaudry and Portier (2015) and the present work provide further evidence about this fact. Our deficiency measure can be regarded as a formalization of the notion of "wedge between VAR innovations and economic shocks" discussed in Sims' paper.

2.3 Two examples

We consider two simple examples that illustrate the concept of sufficent information (partial fundamentalness) and approximate sufficient information.

Example 1: Partial fundamentalness in a square system

Let us assume that output deviates from its potential value because of a demand shock d_t inducing temporary fluctuations, and reacts negatively to the interest rate r_t, expressed in mean deviation, with a one-period delay. Precisely, the output gap y_t is given by

$$y_t = (1 + \alpha L)d_t - \beta r_{t-1},$$

where α and β are positive. The central bank aims at stabilizing output by responding to output gap deviations, so that the interest rate follows the rule

$$r_t = \gamma y_t + v_t,$$

where v_t is a discretionary monetary policy shock and $\gamma > 0$. The structural MA representation for the output gap and the interest rate is then

$$\begin{pmatrix} y_t \\ r_t \end{pmatrix} = \frac{1}{1+\gamma\beta L} \begin{pmatrix} 1+\alpha L & -\beta L \\ \gamma(1+\alpha L) & 1 \end{pmatrix} \begin{pmatrix} d_t \\ v_t \end{pmatrix}. \tag{8}$$

Here the determinant of the MA matrix is $(1+\alpha L)$, which vanishes for $L = -1/\alpha$, so that the representation is non-fundamental if $|\alpha| > 1$. From the policy rule we see that $v_t = r_t - \gamma y_t$, so that the monetary policy shock can be recovered from the present values of the variables included in the VAR, irrespective of α (of course, d_t cannot be found from the x's if $|\alpha| > 1$).

What happens when the above model is non-fundamental ($\alpha > 1$) and the econometrician tries to estimate the monetary policy shock and the related impulse-response functions? To answer this question we generated 1000 artificial data sets with 200 time observations from (8), with $\alpha = 3$, $\gamma = 0.4$, $\beta = 1$ and standard normal shocks. We then estimated for each data set a VAR with 4 lags and identified by imposing a standard Cholesky, lower triangular impact effect matrix, consistently with the model.

Figure 1 displays the true impulse response functions (red solid lines) along with the median (black dashed lines), the 5-th and the 95-th percentiles (grey area) of the distribution of the estimated impulse-response function to the demand shock d_t (first column) and the monetary policy shock v_t (second column). In the lower panels we report the distributions of the correlation coefficients between the estimated shocks and the true shocks.

The figure shows clearly that the impulse response functions are very poorly estimated for d_t, but very precisely for v_t. A similar result holds for the shocks themselves: the distribution of the correlation coefficients is very close to 1 for v_t and far from 1 for d_t.

Let us now have a look to the true and estimated variance decomposition. Table 1 shows the fraction of the forecast error variance of y_t and r_t accounted for by the monetary policy shock. The contribution of the monetary policy shock to total variance is severely underestimated on impact, slightly underestimated at horizon 1 and well estimated at longer horizons. We shall come back on forecast error variance estimation in Section 4.

Table 2 shows the values of δ_d^K, δ_v^K, $K = 1, 4, 1000$. The VAR is dramatically deficient for the demand shock, consistently with Figure 1, but exhibits perfect information for the second shock, the monetary policy shock. Note that x_t must be deficient for the demand shock, since the MA representation is nonfundamental.

Example 2: Partial approximate fundamentalness in a short system

Partial approximate sufficiency, far from being a statistical curiosity, is relevant in practice. As already noticed, most observed variables are likely affected by small macroeconomic shocks and/or measurement errors. Owing to these minor shocks, the applied researcher is usually faced with short systems, which are necessarily nonfundamental, but may be approximately sufficient for the shocks of interest.

As an example, consider the following news shock model, similar to the one used in Forni, Gambetti and Sala (2014). Total factor productivity, a_t, follows the slow diffusion process

$$a_t = a_{t-1} + \alpha \varepsilon_t + \varepsilon_{t-1} \tag{9}$$

where $0 \leq \alpha < 1$.

The representative consumer maximizes

$$E_t \sum_{t=0}^{\infty} \beta^t c_t, \tag{10}$$

where E_t denotes expectation at time t, c_t is consumption and β is a discount factor, subject to the constraint $c_t + \bar{p}_t n_{t+1} = (\bar{p}_t + a_t) n_t$, where $\bar{p}_t$ is the price of a share, n_t is the number of shares and $(\bar{p}_t + a_t) n_t$ is the total amount of resources available at time t. The equilibrium value for asset prices is given by:

$$\bar{p}_t = \sum_{j=1}^{\infty} \beta^j E_t a_{t+j}. \tag{11}$$

Using (9), we see that $E_t a_{t+k} = a_t + \varepsilon_t$ for all $k > 0$. Hence, $\bar{p}_t = (a_t + \varepsilon_t)\beta/(1-\beta)$ and $\Delta \bar{p}_t = b(1+\alpha)\varepsilon_t$, where $b = \beta/(1-\beta)$. Let us assume further that actual prices p_t are subject to a temporary deviation from the equilibrium, driven by the shock d_t, so that $p_t = \bar{p}_t + \gamma d_t$. In addition, let us add an orthogonal measurement error e_t on the technology variable a_t^*, observed by the econometrician. The structural MA representation of Δa_t^* and Δp_t is short, because we have three shocks and just two variables:

$$\begin{pmatrix} \Delta a_t^* \\ \Delta p_t \end{pmatrix} = \begin{pmatrix} \alpha + L & 0 & \theta(1-L) \\ b(1+\alpha) & \gamma(1-L) & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ d_t \\ e_t \end{pmatrix}. \tag{12}$$

We assume unit variance shocks, so we add a scaling factor θ to the impulse response function of e_t to control for the size of the measurement error. We set $\beta = 0.99$, $\alpha = 0.5$ and $\gamma = 20$. Moreover, we set $\theta = 0.5$, so that the measurement error is large (it explains more than 25% of the total variance of Δa_t^*).

Figure 2 shows the estimation results obtained by a Monte Carlo exercise with $T = 200$, i.i.d. unit variance Gaussian shocks and 1000 artificial data sets. The VAR is estimated with 4 lags and is identified by assuming that ε_t is the only one shock affecting a_t^* in the long run, consistently with the model. The estimates of the technology shock ε_t and the related impulse response functions are fairly good, even if a small distortion is visible. On the contrary, the temporary shock to stock prices d_t and the associated responses are poorly estimated. Correspondingly, the deficiency measure is 0.03 for ε_t and 0.97 for d_t (see Table 3).

2.4 DSGE models validation

To get the intuition of the usefulness of the measure, consider a situation where a researcher has two models with different predictions in terms of the response of a specific variable of interest. As an example, consider the technology-hours debate. The RBC model predicts that hours should increase while the New-Keynesian model predicts that hours should fall in response to a technology shock. Gali (1999) estimates a bivariate VAR in labor productivity and hours and identifies the technology shock as the only one shock driving labor productivity in the long run. He then checks whether hours respond positively, as implied by the RBC model, or negatively, as implied by sticky prices models. In order for the results obtained from the VAR analysis to be meaningful in discriminating between models, it is important to check that the deficiency measure associated to a bivariate VAR in in labor productivity and hours is close to zero in both models. Only in that case, the VAR evidence can support one of the two theories. If on the other side, the measure is large, the VAR analysis is inappropriate.

Validation should therefore be performed through the following steps.

1. Consider a calibrated/estimated DSGE model and its linear equilibrium representation as in (5)-(6) or (1). Select the shock of interest u_{it}.

2. Select a subset of variables of the model and a truncation lag K for the VAR.

3. Compute δ_i. If the measure is larger than a pre-specified threshold (e.g. 0.05 or 0.1), consider a different VAR specification (go back to step 2). Otherwise, go to the final step.

4. Verify whether the theoretical impulse response function lie within the VAR confidence bands. If they do, the model is validated. If they do not, there is something wrong either in the values of the parameters or in the model itself.

As observed above, in this paper we abstract from estimation issues. However, in practice the true VAR population parameters are unknown and must be estimated with a T-dimensional sample. If T is small, large-scale VAR specification may fail because of the estimation errors. Our recommendation is to use the most parsimonious specification among those having deficiency measure below the threshold level.

3 Empirics: Validating a theory of news shocks

3.1 The model

In this Section, we assess the theory of news shocks to technology. The model is a New-Keynesian DSGE, similar to the one used by Blanchard, Lorenzoni and L'Huillier (2013, BLL henceforth). The models features several frictions, such as internal habit formation in consumption, adjustment costs in investment, variable capital utilization, Calvo price and wage stickiness. The model also features seven exogenous sources of fluctuations, namely, a news shock and a surprise shock in technology, an investment-specific shock, a monetary policy shock, a shock to price markups, a shock to wage markups and a shock to government expenditures.

We assume that the log-linear exogenous process for technology follows

$$a_t = a_{t-1} + \varepsilon_{t-4} + (1-L)T_t \tag{13}$$

$$T_t = \rho T_{t-1} + v_t \tag{14}$$

where ε_t is a *news* shock that is known to agents at time t, but will be reflected in a_t at time $t+4$ and the part $(1-L)T_t$ is a temporary surprise technology shock.

Some parameters are calibrated using the values obtained by BLL (Table 4). The remaining parameters are estimated using Bayesian techniques. In the estimation, we use data on ???? We take the mode of the posterior distribution (Table 5).

3.2 A preliminary simulation

As a preliminary step of our analysis, we check for the validity of our procedure using a Monte Carlo simulation.

We first define three VAR specifications:

S1 : TFP, GDP, investment.

S2 : TFP, GDP, consumption, investment, hours, interest rate.

S3 : TFP, GDP, consumption, investment, hours, interest rate and inflation.

For each VAR specification, we computed the value of δ for all shocks. Table 6 shows the result. For specification S1 and S2 we cannot have fundamentalness, since there are less variables than shocks. S1 is highly deficient for almost all shocks, including the news shock; however, for the investment specific shock the nonfundamentalness bias is very low ($\delta = 0.05$), showing that a small-scale VAR, with just three variables, might perform reasonably well for this specific shock. Specification S2 is highly deficient for the price markup shock ($\delta = 0.72$); moreover, it does not satisfy the δ-criterion with threshold 0.05 for either the monetary policy shock or the wage markup shock. However, it satisfies the criterion for all other shocks, including the news shock $\delta = 0.00$. Specification S3 has as many variables as shocks. Hence we can verify whether the PMC is satisfied: it turns out that it is not (the largest eigenvalue is 1.01). Consistently with this result, it is seen from the table that δ is not exactly zero for at least two shocks. Hwever, S3 satisfies the δ-criterion for all shocks.

We then simulate data from the model and estimate the three VAR with different observables. For every specification, we obtain a linear combination of the VAR innovations, call it $M_i\epsilon_t$, that minimizes, with respect to the vector Q, the distance $\sum_{t=1}^{T}(Q\epsilon_t - u_{it})^2$, where u_{it} is the news shocks. Finally, we compare the VAR responses with model responses.[4]

If our measure correctly reveals the ability of the VAR in estimating the structural shocks and the associated impulse responses, we should find that the VAR and model impulse response functions are similar when the measure is close to zero and different when it is larger than zero.

Figures 3 to 5 plot the impulse response functions of the news shock estimated in a VAR with simulated data with specification S1, S2 and S3, respectively, as well as the DSGE model responses. The solid black line is the median and the dark and light gray areas represent the 68% and 90% confidence intervals, respectively, of the VAR impulse response functions obtained from generated data. The blue dashed line are the impulse response function of the DSGE model. In S2 and S3, consistently with the fact that the δ-criterion is zero for the news shock, the model responses (dashed blue lines) almost always lie within the VAR bands. Of course, the deviation is due to the finiteness of the sample size. Hence the VAR performs reasonably well in recovering the shocks. On the contrary, the model responses are outside the bands obtained with S1, which again is in line with the values obtained for δ.

3.3 Results

We validate now the transmission mechanisms of news shocks implied by the model. As discussed above, in S2 and S3 partial fundamentalness holds. This means that the two specifications convey enough information to recover the news shock and can be used to assess the theory. On the other hand, the δ-criterion in S1 is close to zero for almost all shocks.

Using real data we estimate the news shocks using the restriction that the news shock is the only shock driving total factor productivity in the long run. Figure 6 plots the results.

[4]Even if the VAR is informationally sufficient, a wrong identification scheme would distort the impulse response functions and make the comparison with model responses meaningless. Finding the linear combination that is closer to the structural shocks is a simple and effective way to make the identification procedure of the VAR irrelevant for the results.

The solid black line is the median, the dark and light gray areas represent the 68% and 90% probability intervals, respectively. The blue dashed line is the impulse response function of the DSGE model. The theory does remarkably well. The model impulse response functions lie within the band except for consumption, GDP and hours, three variables for which the model predicts a larger effect of the shock, see Sala (2015). All in all, the theory seems to be validated. MARIO: DOVRESTI AGGIUNGERE LA INTERPRETAZIONE DELLE DIFFERENZE PER CONSUMO E ORE

Similar results are obtained using S2 in Figure 8. This case is particularly interesting since it is commonly believed that VAR with a number of variables smaller than the number of shocks cannot be successfully used. As shown here, this is incorrect. There might be cases where a single or a subset of shocks can be obtained from a VAR whose variables have a short-system representation.

Finally, Figure 7 plots the impulse response functions obtained in S1. The effects of the identified news shock on GDP and investment estimated with the VAR are larger than those predicted by the model. By neglecting our measure, one would incorrectly conclude that the theory does not capture the dynamics of investment. The δ-criterion, which is 0.72 in this case, suggests that the comparison between the VAR and the DSGE model is inconclusive, since the VAR is not informationally sufficient.

4 Some structural VAR theory for nonfundamental and "short" models

4.1 Fundamentalness: definition and standard results

Let us begin by reviewing the definition of fundamentalness and a few related results.

Definition 1 (Fundamentalness). *We say that u_t is fundamental for x_t, sl and the MA representation $x_t = A(L)u_t$ is fundamental, if and only if $u_{it} \in H_t^x$, $i = 1, \ldots, q$, where $H_t^x = \overline{\text{span}}(x_{1,t-k}, \ldots, x_{n,t-k}, k = 0, \ldots, \infty)$.*

Now, consider the theoretical projection equation of x_t on its past history, i.e. equation (2). The Wold representation of x_t is

$$x_t = B(L)\epsilon_t, \tag{15}$$

where $B(0) = I_n$.

The following result is standard in time series theory.

Proposition 1. *u_t is fundamental for x_t if and only if there exist a nonsingular matrix Q such that $u_t = Q\epsilon_t$.*

It is apparent from the above condition that fundamentalness cannot hold if the system is short. In this case, a matrix Q satisfying Proposition 1 does not exist, since for any $q \times n$ matrix Q, with $q > n$, the entries of $Q\epsilon_t$ are linearly dependent, whereas the entries of u_t are mutually orthogonal.

By contrast, in the square case $n = q$ fundamentalness clearly holds if the impulse response function matrix $A(L)$ is invertible; for, in this case, we can write

$$A(L)^{-1}x_t = u_t,$$

so that the condition defining fundamentalness is satisfied. It is easily seen from (15) that in this case Proposition 1 holds with $Q = A(0)^{-1}$ and $A(L) = B(L)A(0)$.

In the particular case of $A(L)$ being a matrix of rational functions, fundamentalness of u_t for x_t is equivalent to the following condition (see e.g. Rozanov, 1967, Ch. 2).

Condition R. *The rank of $A(z)$ is q for all complex numbers z such that $|z| < 1$.*

When $A(L)$ is a square matrix the above condition reduces to the well known condition that the determinant of $A(z)$ has no roots smaller than one in modulus. Fundamentalness is therefore slightly different from invertibility, since invertibility rules out also roots with modulus equal to 1. Hence invertibility implies fundamentalness, whereas the converse is not true.[5]

4.2 VAR deficiency and sufficient information

For simplicity we shall assume here that the target of VAR estimation is the single shock of interest u_{it} (along with the corresponding impulse response functions). The generalization to any subvector v_t of u_t, including $s \leq q$ shocks, is straightforward.

Let us go back to the VAR representation of x_t, i.e. equation (2), and the projection equation (3). The following proposition says that the structural VAR strategy, i.e. approximating u_{it} by means of the VAR residuals, is optimal in the sense that it provides the best linear approximation, given the VAR information set.

Proposition 2 (Optimality of the structural VAR procedure). *The projection of u_{it} onto the entries of ϵ_t, i.e. $M\epsilon_t$, is equal to the projection of u_{it} onto H_t^x.*

Proof. From (3) it is seen that H_t^x is the direct sum of the two orthogonal spaces H_{t-1}^x and $\text{span}(\epsilon_{jt}, j = 1, \ldots, n)$. Hence $P(u_{it}|H_t^x) = P(u_{it}|\epsilon_{jt}, j = 1, \ldots, n) + P(u_{it}|H_{t-1}^x)$. Since u_{it}

[5]The unit root case is economically interesting in that, if $x_t = \Delta X_t$ and the determinant of $A(z)$ vanishes for $z = 1$, then the entries of X_t are cointegrated. Non-invertibility implies that x_t does not have a VAR representation and VAR estimates do not have good properties. However this problem can be solved by estimating an ECM or a VAR in the levels X_t.

is orthogonal to the past values of the x's, the latter projection is zero. Hence $P(u_{it}|H_t^x = P(u_{it}|\epsilon_{jt}, j = 1, \dots, n) = M\epsilon_t$. QED

Proposition 2 motivates the following definitions.

Definition 2 (VAR Deficiency and Sufficient information). *The informational deficiency of x_t (and the related VAR information set H_t^x) with respect to u_{it} is*

$$\delta_i = \text{var}[u_{it} - P(u_{it}|H_t^x)]/\sigma_{u_i}^2 = \sigma_{e_i}^2/\sigma_{u_i}^2.$$

We say that x_t is informationally sufficient for u_{it} if and only if $\delta_i = 0$, i.e. $u_{it} \in H_t^x$, or, equivalently, $u_{it} = M\epsilon_t$.

As an immediate consequence of Definitions 1 and 2, we have the following result.

Proposition 3. u_t is fundamental for x_t if and only if x_t is informationally sufficient for u_{it}, $i = 1, \dots, q$, or equivalently, $\delta_i = 0$ for all i.

4.3 Partial sufficiency: IRFs

Until now we have focused on the conditions under which the VAR is able to recover the shock u_{it}. However, the ultimate goal of the VAR validation procedure are the impulse response functions, rather than the shock itself. Hence a basic question in our framework is the following. Let the VAR be sufficient for u_{it} and the identification restrictions be correct. Are the impulse response functions obtained from the VAR equal to the theoretical ones? An additional, related question is: is the forecast error variance decomposition equal to its theoretical counterpart? Standard results in VAR identification theory guarantee a positive answer to both questions when the MA representation is (globally) fundamental. But what happens if this is not the case?

The structural VAR procedure consists in inverting the VAR representation to estimate the Wold representation (15) and choosing identification restricttions which deliver an "identification matrix", say Q. The structural shocks are then obtained as $v_t = Q\epsilon_t$ and the corresponding impulse response functions as $A^*(L) = B(L)Q^{-1}$.

Let us assume that H_t^x is sufficient for u_{it}, so that u_{it} can be recovered as the linear combination $M\epsilon_t$. We shall use the following definition.

Definition 4 (Correct identification). *An identification matrix is a nonsingular $n \times n$ matrix Q such that $Q\Sigma_\epsilon Q'$ is diagonal, i.e. the entries of $v_t = Q\epsilon_t$ are orthogonal. An identification matrix is correct for u_{it} if and only if $v_t = Q\epsilon_t$ is such that $v_{ht} = u_{it}$ for some $1 \leq h \leq n$, i.e., denoting with Q_h the h-th line of Q, $Q_h = M$.*

If Q is a correct identification matrix, we can write the impulse response function representation derived from the VAR as

$$x_t = A^*(L)v_t = A_h^*(L)v_{ht} + A_{-h}^*(L)z_t = A_h^*(L)u_{it} + A_{-h}^*(L)z_t, \tag{16}$$

where $A_h^*(L)$ is the h-th column of $A^*(L)$, $A_{-h}^*(L)$ is the $n \times n-1$ matrix obtained by eliminating the h-th column from $A^*(L)$ and $z_t = (v_{1t} \ \cdots v_{h-1,t} \ v_{h+1,t} \ \cdots v_{nt})'$. $A_h^*(L)$ is the vector of impulse response functions derived from the VAR —let us say the "empirical" impulse response functions, even if, of course, we are speaking of the population VAR (with infinite sample size).

Now let $A_i(L)$ be the i-th column of $A(L)$ and $A_{-i}(L)$ be the $n \times q - 1$ matrix obtained by eliminating the i-th column from $A(L)$. The structural MA representation can then be written as

$$x_t = A(L)u_t = A_i(L)u_{it} + A_{-i}(L)w_t, \tag{17}$$

where $w_t = (u_{1t} \ \cdots u_{i-1,t} \ u_{i+1,t} \ \cdots u_{qt})'$. $A_i(L)$ is the vector of the "true" impulse response functions.

Proposition 4. *Let x_t and the related VAR be informationally sufficient, and the identification matrix be correct for u_{it}. Then the empirical impulse response functions are equal to the true impulse response functions, i.e. $A_h^*(L) = A_i(L)$ for some h, $1 \leq h \leq n$.*

Proof. Let us first observe that the entries of v_t are orthogonal at all leads and lags, since $v_t = Q\epsilon_t$ is a vector white noise and Q is an identification matrix. It follows that u_{it} is orthogonal to the entries of z_t at all leads and lags. Moreover, by the assumptions of model (1), u_{it} is also orthogonal to u_{jt}, $j \neq i$, and therefore to the entries of w_t, at all leads and lags. From (16) and (17) we get $A_h^*(L)u_{it} + A_{-h}^*(L)z_t = A_i(L)u_{it} + A_{-i}(L)w_t$. Projecting both sides onto $u_{i,t-k}, k \geq 0$ we get $A_h^*(L)u_{it} = A_i(L)u_{it}$, which imples the result. QED

Let us observe that the equality result in Proposition 4 translates into a consistency result for real-world, finite-sample VARs, provided that the parameters of the population VAR are estimated consistently, and the truncation lag increases with the sample size, following a consistent information criterion.

4.4 Partial sufficiency: variance decomposition

Partial identification has also implications in term of variance decomposition. It is well known that $H_t^\epsilon \subseteq H_t^u$ and, if u_t is nonfundamental for x_t, $H_t^\epsilon \subset H_t^u$. As a consequence, in the nonfundamental case, the prediction error of v_t is larger than the one of u_t. Precisely, for any horizon $s \geq 0$, we have

$$\text{var}\ [P(x_{i,t+s}|H_{t-1}^u) - x_{i,t+s}] \leq \text{var}\ [P(x_{i,t+s}|H_{t-1}^v) - x_{i,t+s}], \tag{18}$$

and the inequality is strict at least for $s = 0$ if u_t is nonfundamental for x_t. Hence if x_t is informationally sufficient for u_{it}, but not for all shocks, then total forecast error variance is overestimated by the VAR model at short horizons. On the other hand, Proposition 4 implies that the impulse response functions of u_{it}, and therefore the variance of the forecast errors, is estimated consistently. Putting thins together, the fraction of total variance accounted for by u_{it}, derived from the VAR, is downward biased, since the numerator is unbiased, whereas the denominator is upward biased.[6]

An alternative variance decomposition, which is not affected by this bias, is obtained by using integrals of the spectral densities over suitable frequency bands (see e.g. Forni, Gambetti and Sala, 2014). Let $A_{i,j}(L)$ and $A^*_{h,j}(L)$, be the j-th elements of the matrices $A_i(L)$, $A^*_h(L)$, respectively. As is well known, the variance of the component of x_{jt} which is attributable to v_{ht} can be computed as $\sigma^2_{v_h} \int_0^\pi A_{-i,j}(e^{-i\theta})A_{-i,j}(e^{i\theta})d\theta/\pi$. If we are interested for instance in the variance of waves of business cycle periodicity, say between 8 and 32 quarters, the corresponding angular frequencies (with quarterly data) are $\theta_1 = \pi/4$ and $\theta_2 = \pi/16$ and the corresponding variance is $\sigma^2_{v_h} \int_{\theta_1}^{\theta 2} A_{-i,j}(e^{-i\theta})A_{-i,j}(e^{i\theta})d\theta/\pi$. On the other hand, the total "cyclical" variance of x_{jt} is given by $\int_{\theta_1}^{\theta 2} S_{x_j}(\theta)d\theta/\pi$, where $S_{x_j}(\theta)$ denotes the spectral density of x_{jt}. Hence the contribution of v_{ht} to the cyclical variance of x_{jt} is given by

$$\frac{\sigma^2_{v_h} \int_{\theta_1}^{\theta 2} A^*_{h,j}(e^{-i\theta})A^*_{h,j}(e^{i\theta})d\theta}{\int_{\theta_1}^{\theta 2} S_{x_j}(\theta)d\theta}.$$

Similarly, the contribution of u_{it} is given by

$$\frac{\sigma^2_{u_i} \int_{\theta_1}^{\theta 2} A_{i,j}(e^{-i\theta})A_{i,j}(e^{i\theta})d\theta}{\int_{\theta_1}^{\theta 2} S_{x_j}(\theta)d\theta}.$$

Under the assumptions of Proposition 4, the numerators are equal, so that the ratios are equal. Therefore this kind of variance decomposition analysis is preferable to the standard one in that it is not biased in the case of partial fundamentalness.

4.5 Finite-order VAR deficiency

In this subsection we consider a quite natural extension of the deficiency measure to the case of finite-order VARs. Let us denote the VAR(K) information set as $H^x_t(K) = \text{span}(x_{j,t-k}, j = 1, \dots, n, k = 0, \dots, K)$ and consider the orthogonal decompositions

$$x_t = P(x_t|H^x_t(K)) + \epsilon^K_t \quad (19)$$

$$u_{it} = M^K \epsilon^K_t + e^K_{it}. \quad (20)$$

[6]If $x_{it} = \Delta X_{it}$ a similar result holds for the decomposition of the forecast error variance of the level $X_{i,t+s}$. This explains the large estimation error, at horizon 0, reported in Table 1 for the variable r_t.

Proposition 2 still holds for the finite-order VAR.

Proposition 2′ *The projection of u_{it} onto the entries of ϵ_t^K, i.e. $M^K \epsilon_t^K$, is equal to the projection of u_{it} onto $H_t^x(K)$.*

The proof is the same as that of Proposition 2, with ϵ_t^K in place of ϵ_t and $H_\tau^x(K)$ in place of H_τ^x, $\tau = t, t-1$. The VAR(K) deficiency can then be defined as

$$\delta_i^K = \text{var}[u_{it} - P(u_{it}|H_t^x(K))]/\sigma_{u_i}^2 = \sigma_{e_i^K}^2/\sigma_{u_i}^2. \tag{21}$$

Correspondingly, we can say that $H_t^x(K)$ *is informationally sufficient for* u_{it} if and only if $\delta_i^K = 0$, i.e. $u_{it} \in H_t^x(K)$, or, equivalently, $u_{it} = M^K \epsilon_t^K$. As K increases, the spaces $H_t^x(K)$ are nested, so that the sequence $\delta_i(K)$ is non-increasing in K and $\delta_i \leq \delta_i(K)$ for any K. The difference $\delta_i(K) - \delta_i$ provides a precise information about the effect of lag truncation on estimation of the shock u_{it}.

Unfortunately, this is not true for the impulse response functions, since Proposition 4 does not hold for the finite-order VAR. It might be the case that the K-order VAR is sufficient for u_{it}, but the corresponding impulse response functions are biased.[7] Hence the difference $\delta_i(K) - \delta_i$ cannot be regarded as a measure of the additional bias due to lag truncation. Nevertheless, $\delta_i(K)$ deserves interest in validation exercises, in that it provides a lower bound for the overall bias due to nonfundamentalness and lag truncation.

4.6 Computing VAR deficiency

To compute VAR deficiency, a simple formula can be derived as follows. Let us write the projection equation of u_{it} onto $H_t^x(K)$ as

$$u_{it} = P(u_{it}|H_t^x(K)) + e_{it}^K = Fy_t + e_{it}^K,$$

[7] This is because the VAR residuals in ϵ_t^K might be serially correlated. By inverting the finite-order VAR, we get the representation

$$x_t = B^K(L)\epsilon_t^K,$$

and, by imposing the identification constraints, we get the "shocks" $v_t^K = Q\epsilon_t^K$ and the corresponding impulse response functions $A^K(L) = B^K(L)Q^{-1}$. To get unbiasedness of these response functions we need the assumption of Proposition 4, along with the additional condition that ϵ_t^K is a vector white noise.

Proposition 4′. *Let a VAR(K) be informationally sufficient, and the identification matrix be correct for u_{it}. Assume further that the VAR residual ϵ_t^K is a vector white noise. Then the empirical impulse response functions are equal to the true impulse response functions, i.e. $A_h^K(L) = A_i(L)$.*

We omit the proof, which is essentially the same as that of Proposition 4. Notice that, starting from the parameters of the economic model, we can check in principle whether serial correlation of ϵ_t^K is satisfied.

where $y_t = (x'_{t-1} \ \cdots \ x_{t-K})'$ and $F = E(u_{it}y'_t)\Sigma_y^{-1}$, Σ_y being the variance covariance matrix of y_t. From (21) and the above equation we get

$$\delta_i(K) = 1 - F\Sigma_y F'/\sigma^2_{u_i} = 1 - E(u_{it}y'_t)\Sigma_y^{-1}E(u_{it}y'_t)/\sigma^2_{u_i}.$$

Using (17) it is easily seen that $E(u_{it}x'_t) = A_i(0)'$ and $E(u_{it}x'_{t-k}) = 0$ for all $k > 0$, so that $E(u_{it}y'_t) = (A_i(0)' \ 0 \ \cdots \ 0)$.

Hence

$$\delta_i(K) = 1 - A_i(0)'GA_i(0)/\sigma^2_{u_i}, \tag{22}$$

where G is the $n \times n$ upper-left submatrix of Σ_y^{-1}, Σ_y being

$$\Sigma_y = \begin{pmatrix} \Gamma_0 & \Gamma_1 & \cdots & \Gamma_K \\ \Gamma_{-1} & \Gamma_0 & \cdots & \Gamma_{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{-K} & \Gamma_{-K+1} & \cdots & \Gamma_0 \end{pmatrix},$$

where $\Gamma_k = E(x_t x'_{t-k})$, $k = 0, \ldots, K$. The covariance matrices of x_t can easily be computed from the MA representation (1) by using the covariance generating function

$$A(L)\Sigma_u A(L^{-1})' = \sum_{k=-\infty}^{\infty} \Gamma_k L^k.$$

As for δ_i, it can simply be approximated with any desired precision by using a suitably large K.[8]

5 Conclusions

DSGE models are fascinating theoretical constructions. But structural VAR models are alive and well. They can be used for the empirical validation of macroeconomic models which possess

[8]An exact formula for δ_i is

$$\delta_i = 1 - \sigma^2_{u_i} A_i(0)'\Sigma_\epsilon^{-1} A_i(0).$$

This formula is obtained from (3) by observing that $M = E(u_{it}\epsilon_t{}')\Sigma_\epsilon^{-1}$ and noting that, by (2) and (1), $E(\epsilon_t u_{it}) = E(x_t u_{it}) = A_i(0)\sigma^2_{u_i}$. If the model can be written in the state-space form

$$s_t = As_{t-1} + Bu_t \tag{23}$$
$$x_t = Hs_t \tag{24}$$

the matrix $\Sigma_e psilon$ can be obtained from the Wold representation

$$x_t = \left\{I_n + H(I_m - AL)^{-1}KL\right\}\epsilon_t \tag{25}$$

where K is the steady-state Kalman gain: Σ_ϵ is giben by HPH', where P is the steady-state variance-covariance of the states.

an impulse-response representation, even if such representation is nonfundamental. They can be used successfully even if the number of variables is smaller than the number of structural shocks in the economic model.

For the validation exercise to be meaningful, a necessary condition is that the VAR conveys enough information to recover the shocks of interest and the related impulse response functions.

VAR deficiency for a given shock can be measured by δ_i, i.e. the fraction of unexplained variance of the linear projection of this shock onto the VAR information set. Hence a crucial step of the validation procedure is to verify whether δ_i is acceptably small.

For DSGE models including news or foresight shock, nonfundamentalness is endemic. Such models are often regarded as incompatible with VARs, in that a VAR representation in the structural shocks does not exist. Hence we illustrate our ideas by conducting a validation exercise with a news shock DSGE model. We show that our VAR specification can be used for model validation, despite nonfundamentalness. We find that the DSGE model performs reasonably well in fitting the impulse response functions derived from US data.

References

[1] Beaudry, P., Fève, P., Guay, A. and F. Portier, 2015. When is nonfundamentalness in VAR a real problem? An application to news shocks. NBER Working Paper No. 21466.

[2] Canova, F., 2002. Validating Monetary DSGE Models through VARs, CEPR Discussion Papers 3442, C.E.P.R. Discussion Papers.

[3] Canova, F., 2007. Methods for Applied Macroeconomic Research. Princeton: Princeton University Press.

[4] Chari, V.V., Kehoe, P.J. and E.R. McGrattan, 2008. Are structural VARs with long-run restrictions useful in developing business cycle theory? Journal of Monetary Economics 55, 1337-1352.

[5] Christiano, L., M. Eichenbaum, Vigfusson, R., 2004. The Response of Hours to a Technology Shock: Evidence Based on Direct Measures of Technology, Journal of the European Economic Association 2, 381-395.

[6] Christiano, L., M. Eichenbaum, Vigfusson, R., 2007. Assessing Structural VARs, NBER Chapters, in: NBER Macroeconomics Annual 2006, vol.21, pages 1-106, National Bureau of Economic Research, Inc.

[7] Fernandez-Villaverde J., J.F. Rubio-Ramirez, T.J. Sargent and M.W. Watson, 2007. ABCs (and Ds) of Understanding VARs, American Economic Review 97, 1021-1026.

[8] Forni, Gambetti, L., Sala, L., 2014. No News in Business Cycles. The Economic Journal 124, pp. 1168-1191.

[9] Forni, M., Gambetti, L., Lippi, M., Sala, L., 2013. Noisy News in Business Cycles. CEPR Discussion Papers series no. 9601.

[10] Forni, M., Gambetti, L., Lippi, M., Sala, L., 2016. Noise Bubbles. The Economic Journal, forthcoming.

[11] Forni, M., Giannone, D., Lippi, M., Reichlin, L., 2009. Opening the Black Box: Structural Factor Models with Large Cross-Sections. Econometric Theory 25, 1319-1347.

[12] Franchi, M. and Paruolo, P., 2015. Minimality of State Space Solutions of DSGE Models and Existence Conditions for Their VAR Representation, Computational Economics 46, 613-626.

[13] Giacomini, R., 2013. The Relationship Between DSGE and VAR Models, Advances in Econometrics, vol. 32: VAR Models in Macroeconomics New Developments and Applications: Essays in Honor of Christopher A. Sims, Emerald Group Publishing.

[14] Giannone, D., Reichlin, L., 2006. Does Information Help Recovering Structural Shocks from Past Observations? Journal of the European Economic Association 4, 455-465.

[15] Giannone, D., Reichlin, L., and Sala, L., 2006. VARs, common factors and the empirical validation of equilibrium business cycle models, Journal of Econometrics 132, 257-279.

[16] Hansen, L.P., Sargent, T.J., 1991. Two problems in interpreting vector autoregressions. In: Rational Expectations Econometrics, L.P. Hansen and T.J. Sargent, eds. Boulder: Westview, 77-119.

[17] Leeper, E.M., Walker, T.B., Yang, S.S., 2013. Fiscal Foresight and Information Flows. Econometrica 81, 1115-1145.

[18] Lippi, M., Reichlin, L., 1993. The Dynamic Effects of Aggregate Demand and Supply Disturbances: Comment. American Economic Review 83, 644-652.

[19] Lippi, M., Reichlin, L., 1994a. Diffusion of Technical Change and the Decomposition of Output into Trend and Cycle, Review of Economic Studies, Oxford University Press, vol. 61(1), 19-30.

[20] Lippi, M., Reichlin, L., 1994b. VAR analysis, non-fundamental representation, Blaschke matrices. Journal of Econometrics 63, 307-325.

[21] Mertens, K., Ravn, M.O., 2010. Measuring the Impact of Fiscal Policy in the Face of Anticipation: A Structural VAR Approach. Economic Journal 120, 393-413.

[22] Ravenna, F., 2007. Vector autoregressions and reduced form representations of DSGE models. Journal of Monetary Economics 54, 2048-2064.

[23] Rozanov, Yu., 1967. Stationary Random processes. San Francisco: Holden Day.

[24] Sala, L., 2015. Dsge Models in the Frequency Domains. Journal of Applied Econometrics 30, 219-240.

[25] Schmitt-Grohe, S., Uribe, M., 2012. What's News in Business Cycles. Econometrica

[26] Sims, E.R., 2012. News, Non-Invertibility, and Structural VARs, Working Papers 013, University of Notre Dame, Department of Economics.

[27] Yang, S.S., 2008. Quantifying tax effects under policy foresight. Journal of Monetary Economics 52, 1557-1568.

Appendix: The DSGE model

The model follows closely Blanchard, L'Huillier and Lorenzoni (2013). The preferences of the representative household are given by the utility function:

$$E_t\left[\sum_{t=0}^{\infty}\beta^t\left(\log\left(C_t - hC_{t-1}\right) - \frac{1}{1+\varsigma}\int_0^1 N_{jt}^{1+\varsigma}dj\right)\right],$$

C_t is consumption, the term hC_{t-1} captures internal habit formation, and N_{jt} is the supply of specialized labor of type j. The household budget constraint is

$$P_tC_t + P_tI_t + T_t + B_t + P_tC(U_t)\bar{K}_{t-1} = R_{t-1}B_{t-1} + Y_t + \int_0^1 W_{jt}N_{jt}dj + R_t^kK_t,$$

where P_t is the price level, T_t is a lump sum tax, B_t are holdings of one period bonds, R_t is the one period nominal interest rate, Y_t are aggregate profits, W_{jt} is the wage of specialized labor of type j, N_{jt}. R_t^k is the capital rental rate.

Households choose consumption, bond holdings, capital utilization, and investment each period so as to maximize their expected utility subject to the budget constraint and a standard no-Ponzi condition. Nominal bonds are in zero net supply, so market clearing in the bonds market requires $B_t = 0$.

The capital stock $\bar{K}_t$ is owned and rented by the representative household and the capital accumulation equation is

$$\bar{K}_t = (1-\delta)\bar{K}_{t-1} + D_t\left[1 - G(I_t/I_{t-1})\right]I_t,$$

where δ is the depreciation rate, D_t is a stochastic investment-specific technology parameter, and G is a quadratic adjustment cost in investment

$$G(I_t/I_{t-1}) = \chi(I_t/I_{t-1} - \Gamma)^2/2,$$

where Γ is the long-run gross growth rate of TFP. The model features variable capacity utilization: the capital services supplied by the capital stock $\bar{K}_{t-1}$ are $K_t = U_t\bar{K}_{t-1}$, where U_t is the degree of capital utilization and the cost of capacity utilization, in terms of current production, is $C(U_t)\bar{K}_{t-1}$, where $C(U_t) = U_t^{1+\zeta}/(1+\zeta)$.

The investment-specific technology parameter $d_t = \log D_t$ follows the stochastic process:

$$d_t = \rho_d d_{t-1} + \varepsilon_{dt}.$$

ε_{dt} and all the variables denoted with ε from now on are i.i.d. shocks.

Consumption and investment are in terms of a final good which is produced by competitive final good producers using the CES production function

$$Y_t = \left(\int_0^1 Y_{jt}^{\frac{1}{1+\mu_{pt}}} dj \right)^{1+\mu_{pt}}$$

which employs a continuum of intermediate inputs. Y_{jt} is the quantity of input j employed and μ_{pt} captures a time-varying elasticity of substitution across goods, where $\log(1+\mu_{pt}) = \log(1+\mu_p) + m_{pt}$ and m_{pt} follows the process $m_{pt} = \rho_p m_{pt-1} + \varepsilon_{pt} - \psi_p \varepsilon_{pt-1}$.

The production function for intermediate good j is

$$Y_{jt} = (K_{jt})^{\alpha} (A_t L_{jt})^{1-\alpha},$$

where K_{jt} and L_{jt} are, respectively, capital and labor services employed. The technology parameter a_t follows the process

$$a_t = a_{t-1} + \varepsilon_{t-4} + (1-L)T_t \tag{26}$$

$$T_t = \rho T_{t-1} + v_t, \tag{27}$$

where ε_t is a *news* shock that is known to agents at time t, but will be reflected in a_t at time $t+4$ and the part $(1-L)T_t$ is a persistent, but temporary, surprise technology shock.

BLL (2013) treat explicitly the constant term in TFP growth by letting $A_t = \Gamma^t e^{a_t}$, but calibrate $\Gamma = 1$.

Intermediate good prices are sticky with price adjustment as in Calvo (1983). Each period intermediate good firm j can freely set the nominal price P_{jt} with probability $1-\theta_p$ and with probability θ_p is forced to keep it equal to P_{jt-1}. These events are purely idiosyncratic, so θ_p is also the fraction of firms adjusting prices each period.

Labor services are supplied to intermediate good producers by competitive labor agencies that combine specialized labor of types in $[0,1]$ using the technology

$$N_t = \left[\int_0^1 N_{jt}^{\frac{1}{1+\mu_{wt}}} dj \right]^{1+\mu_{wt}},$$

where $\log(1+\mu_{wt}) = \log(1+\mu_w) + m_{wt}$ and m_{wt} follows the process $m_{wt} = \rho_w m_{wt-1} + \varepsilon_{wt} - \psi_w \varepsilon_{wt-1}$.

The presence of differentiated labor introduces monopolistic competition in wage setting as in Erceg, Henderson and Levin (2000). Specialized labor wages are also sticky and set by the household. For each type of labor j, the household can freely set the price W_{jt} with probability $1-\theta_w$ and has to keep it equal to W_{jt-1} with probability θ_w.

Market clearing in the final good market requires

$$C_t + I_t + C(U_t)\bar{K}_{t-1} + G_t = Y_t.$$

Market clearing in the market for labor services requires $\int L_{jt}dj = N_t$.

Government spending is set as a fraction of output and the ratio of government spending to output is $G_t/Y_t = \psi + g_t$, where g_t follows the stochastic process

$$g_t = \rho_g g_{t-1} + \varepsilon_{gt}.$$

Monetary policy follows the interest rate rule

$$r_t = \rho_r r_{t-1} + (1 - \rho_r)(\gamma_\pi \pi_t + \gamma_y \hat{y}_t) + q_t,$$

where $r_t = \log R_t - \log R$ and $\pi_t = \log P_t - \log P_{t-1} - \pi$, π is the inflation target, $\hat{y}_t$ is defined below and q_t follows the process

$$q_t = \rho q_{t-1} + \varepsilon_{qt}.$$

The model is solved and a log-linear approximation around a deterministic steady-state is computed.

Given that TFP is non-stationary, some variables need to be normalized to ensure stationarity. Here we define $\hat{c}_t$ as

$$\hat{c}_t = \log(C_t/A_t) - \log(C/A),$$

where C/A denotes the value of C_t/A_t in the deterministic version of the model in which A_t grows at the constant growth rate Γ. Analogous definitions apply to the quantities $\hat{y}_t, \hat{k}_t, \hat{\bar{k}}_t, \hat{\imath}_t$. The quantities N_t and U_t are already stationary, so $n_t = \log N_t - \log N$, and similarly for u_t. For nominal variables, it is necessary to take care of non-stationarity in the price level, so: $\hat{w}_t = \log(W_t/(A_t P_t)) - \log(W/(AP)), r_t^k = \log(R_t^k/P_t) - \log(R^k/P), m_t = \log(M_t/P_t) - \log(M/P), r_t = \log R_t - \log R, \pi_t = log(P_t/P_{t-1}) - \pi$.

Finally, for the Lagrange multipliers: $\hat{\lambda}_t = \log(\Lambda_t A_t) - \log(\Lambda A), \hat{\phi}_t = \log(\Phi_t A_t/P_t) - \log(\Phi A/P)$. Φ_t is the Lagrange multiplier on the capital accumulation constraint. The hat is only used for variables normalized by A_t.

The first order conditions can be log-linearized to yield

$$
\begin{aligned}
\hat{\lambda}_t &= \frac{h\beta\Gamma}{(\Gamma-h\beta)(\Gamma-h)}E_t\hat{c}_{t+1} - \frac{\Gamma^2+h^2\beta}{(\Gamma-h\beta)(\Gamma-h)}\hat{c}_t + \frac{h\Gamma}{(\Gamma-h\beta)(\Gamma-h)}\hat{c}_{t-1} + \\
&\quad + \frac{h\beta\Gamma}{(\Gamma-h\beta)(\Gamma-h)}E_t[\Delta a_{t+1}] - \frac{h\Gamma}{(\Gamma-h\beta)(\Gamma-h)}\Delta a_t \\
\hat{\lambda}_t &= r_t + E_t[\hat{\lambda}_{t+1} - \Delta a_{t+1} - \pi_{t+1}] \\
\hat{\phi}_t &= (1-\delta)\beta\Gamma^{-1}E_t[\hat{\phi}_{t+1} - \Delta a_{t+1}] + (1-(1-\delta)\beta\Gamma^{-1})E_t[\hat{\lambda}_{t+1} - \Delta a_{t+1} + r^k_{t+1}] \\
\hat{\lambda}_t &= \hat{\phi}_t + d_t - \chi\Gamma^2\left(\hat{\imath}_t - \hat{\imath}_{t-1} + \Delta a_t\right) + \beta\chi\Gamma^2 E_t\left(\hat{\imath}_{t+1} - \hat{\imath}_t + \Delta a_{t+1}\right) \\
r^k_t &= \zeta u_t \\
m_t &= \alpha r^k_t + (1-\alpha)\hat{w}_t \\
r^k_t &= \hat{w}_t - \hat{k}_t + n_t
\end{aligned}
$$

Log-linearizing the accumulation equation for capital and the equation for capacity utilization, yields

$$
\begin{aligned}
\hat{k}_t &= u_t + \hat{\bar{k}}_{t-1} - \Delta a_t \\
\hat{\bar{k}}_t &= (1-\delta)\Gamma^{-1}\left(\hat{\bar{k}}_t - \Delta a_t\right) + \left(1-(1-\delta)\Gamma^{-1}\right)d_t + \hat{\imath}_t.
\end{aligned}
$$

Approximating and aggregating the intermediate goods production function over producers and using the final good production function yields

$$
\hat{y}_t = \alpha\hat{k}_t + (1-\alpha)n_t
$$

Market clearing in the final good market yields

$$
(1-\psi)\hat{y}_t = \frac{C}{Y}\hat{c}_t + \frac{I}{Y}\hat{\imath}_t + \frac{R^kK}{PY}u_t + g_t
$$

C/Y, I/Y and $R^kK/(PY)$ are all equilibrium ratios in the deterministic version of the model in which A_t grows at the constant rate Γ.

Aggregating individual optimality conditions for price setters yields the Phillips curve

$$
\pi_t = \beta E_t\pi_{t+1} + \kappa m_t + \kappa m_{pt}
$$

where $\kappa = (1-\theta_p\beta)(1-\theta_p)/\theta_p$.

Finally, aggregating individual optimality conditions for wage setters yields

$$
\begin{aligned}
\hat{w}_t &= \frac{1}{1+\beta}\hat{w}_{t-1} + \frac{\beta}{1+\beta}E_t\hat{w}_{t+1} - \frac{1}{1+\beta}(\pi_t + \Delta a_t) + \frac{\beta}{1+\beta}E_t(\pi_{t+1} + \Delta a_{t+1}) - \\
&\quad -\kappa_w\left(\hat{w}_t - \zeta n_t + \hat{\lambda}_t + \kappa_w m_{wt}\right)
\end{aligned}
$$

where $\kappa_w = \frac{(1-\theta_w\beta)(1-\theta_w)}{\theta_w(1+\beta)\left(1+\zeta\left(1+\frac{1}{\mu_w}\right)\right)}$.

Tables and Figures

Horizon	0	1	4	16
y_t, median estimate	0.00	0.10	0.12	0.12
y_t, true	0.00	0.11	0.12	0.12
r_t, median estimate	0.41	0.45	0.45	0.45
r_t, true	0.86	0.48	0.45	0.45

Table 1: Fraction of forecast error variance accounted for by the monetary policy shock in the empirical simulation of Example 1.

Shocks of interest	$\delta(1)$	$\delta(4)$	$\delta(1000)$
Demand shock, d_t	0.8904	0.8889	0.8889
Monetary shock v_t	0.0000	0.0000	0.0000

Table 2: The measure of informational deficiency δ for Example 1.

Shocks of interest	$\delta(1)$	$\delta(4)$	$\delta(1000)$
Shock ε_t	0.0347	0.0344	0.0342
Shock d_t	0.9732	0.9687	0.9653
Shock e_t	0.4891	0.2558	0.0899

Table 3: The measure of informational deficiency δ for Example 2.

Calibrated parameters	
ζ (elasticity of k utilization)	2.07
χ (I adj. cost)	5.5
h (habit persistence)	0.53
ς (inverse Frish elast.)	3.98
θ_w (W stickiness)	0.87
θ_p (P stickiness)	0.88
γ_π (π in Taylor rule)	1.003
γ_y (Y gap in Taylor rule)	0.0044
μ_p (SS P markup)	0.3
μ_w (SS W markup)	0.05
α (coeff. in prod. function)	0.19
Γ (TFP growth)	1
ψ (G/Y)	0.22
δ (K depreciation)	0.025
β (discount factor)	0.99

Table 4: Calibrated parameters. We use the posterior mean values estimated by BLL.

Estimated parameters		
Parameter	Prior	Mode
ρ_r (i smoothing)	$Beta(0.5, 0.2)$	0.66
ρ (temp. technology)	$\mathcal{N}(0.0\ ,\ 0.4)$	0.95
ρ_q (monetary)	$\mathcal{N}(0.0\ ,\ 0.5)$	0.17
ρ_d (I specific)	$\mathcal{N}(0.0, 0.5)$	0.74
ρ_p (P markup)	$\mathcal{N}(0.0, 0.5)$	0.88
ρ_w (W markup)	$\mathcal{N}(0.0, 0.5)$	0.71
ρ_g (G)	$\mathcal{N}(0.0, 0.5)$	0.99
ψ_p (MA in P mkup)	$Beta(0.5, 0.2)$	0.57
ψ_w (MA in W mkup)	$Beta(0.5, 0.2)$	0.68
σ_ε (permanent tech.)	$I\Gamma(0.5, 1.0)$	0.96
σ_v (temporary tech.)	$I\Gamma(1.0, 1.0)$	1.31
σ_q (monetary)	$I\Gamma(0.15, 1.0)$	0.24
σ_d (I specific)	$I\Gamma(5.0, 1.5)$	4.65
σ_p (p markup)	$I\Gamma(0.15, 1.0)$	0.14
σ_w (w markup)	$I\Gamma(0.15, 1.0)$	0.37
σ_g (gov exp.)	$I\Gamma(0.5, 1.0)$	0.44
Posterior value at median		

Table 5: Parameter estimates: mode.

shock		S1	S2	S3
news	ε_t	**0.72**	**0.00**	**0.00**
temp. tech	v_t	0.28	0.00	0.00
price mkup	p_t	0.97	0.73	0.02
wage mkup.	w_t	0.99	0.13	0.00
gov't exp.	g_t	0.45	0.00	0.00
inv. spec.	d_t	0.05	0.01	0.01
mon. policy	q_t	0.92	0.22	0.00

Table 6: The measure of informational deficiency δ for the VAR specification estimated in Section 3.

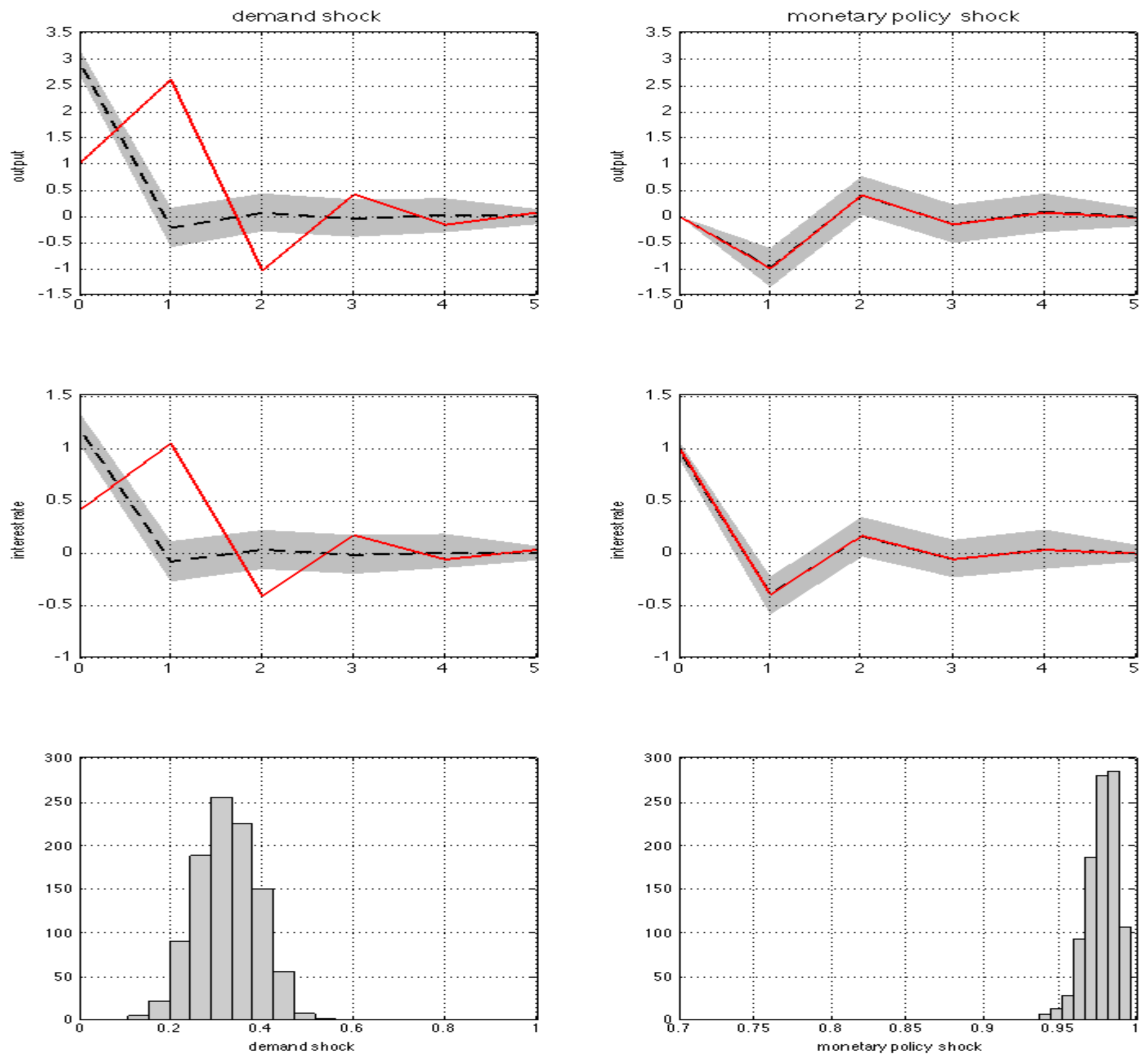

Figure 1: Impulse response functions of model (8) with $\alpha = 3$, $\gamma = 0.4$, $\beta = 1$ and standard normal shocks. Red solid lines: true impulse response functions. Black dashed lines: median of estimated impulse response functions. Grey area contains the 90% confidence interval computed from the Monte Carlo simulations. Lower panels report the distributions of the correlation coefficients between the estimated shocks and the true shocks (demand shock, d_t: left column; monetary policy shock, v_t: right column).

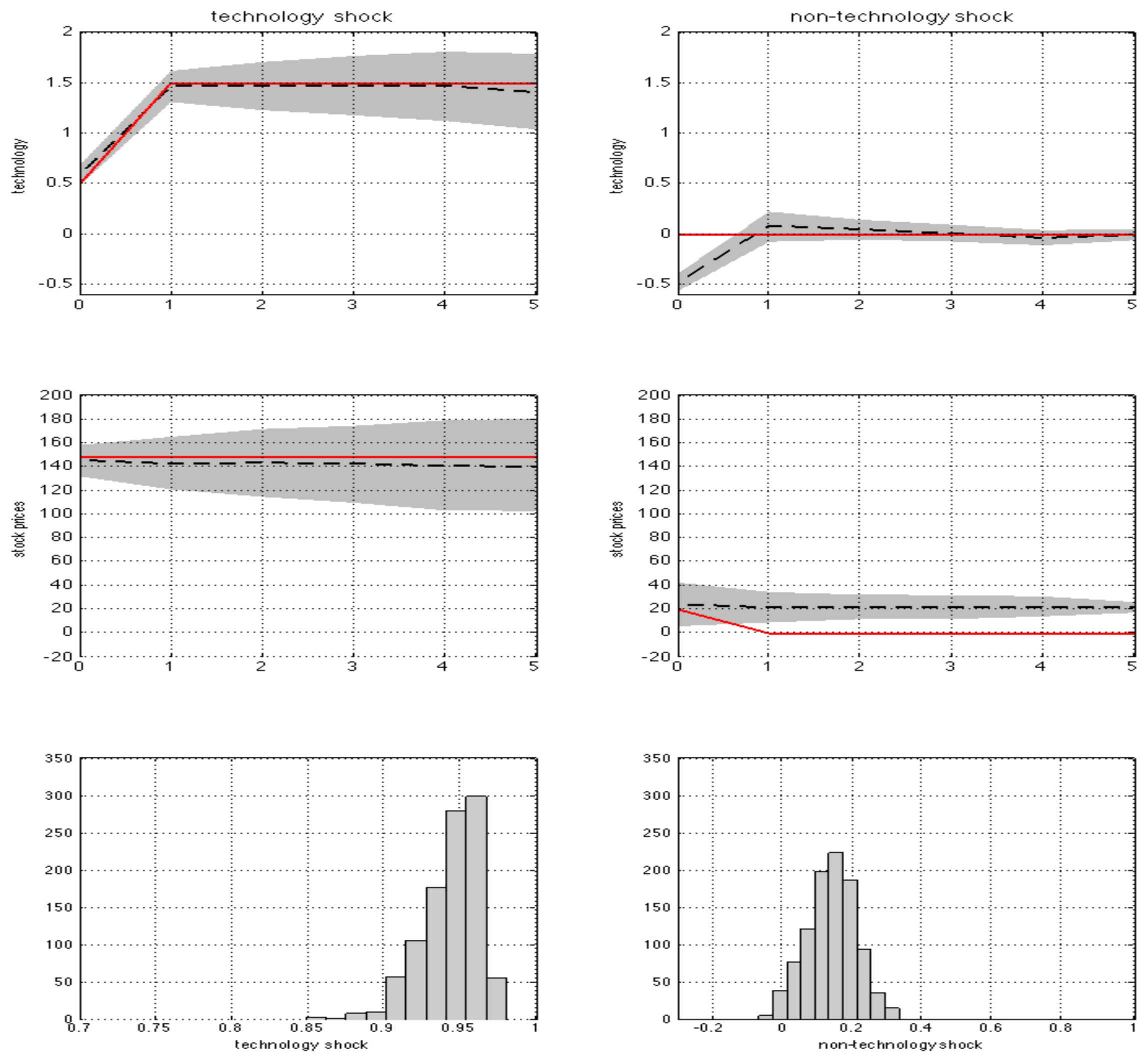

Figure 2: Impulse response functions of model (12) with $\beta = 0.99$, $\alpha = 0.5$, $\gamma = 20$, $\theta = 0.5$ and standard normal shocks. Red solid lines: true impulse response functions. Black dashed lines: median of estimated impulse response functions. Grey area contains the 90% confidence interval computed from the Monte Carlo simulations. Lower panels report the distributions of the correlation coefficients between the estimated shocks and the true shocks (ε_t, technology shock: left column; d_t, temporary shock to prices: right column).

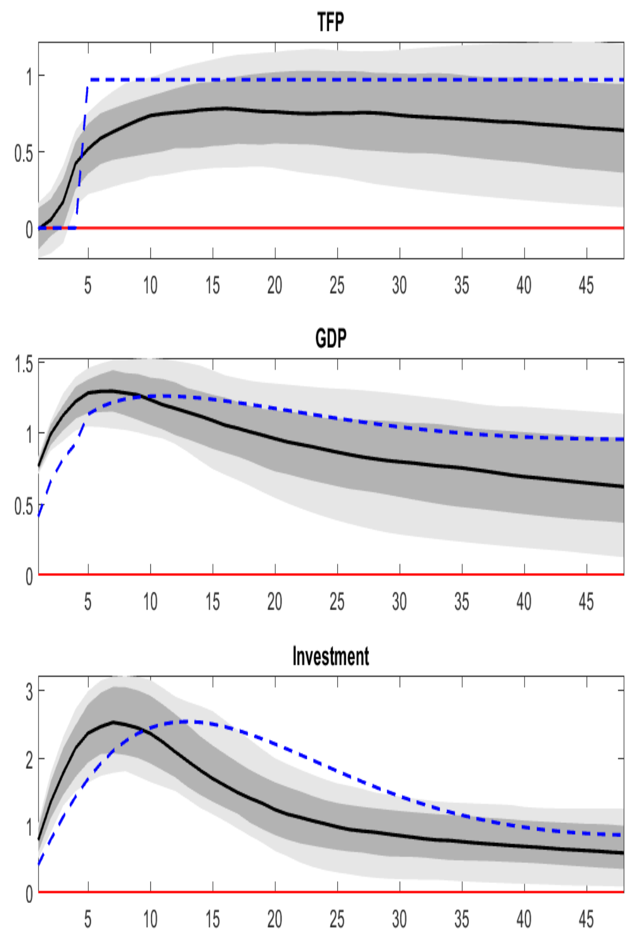

Figure 3: Impulse responses - S1 VAR. Minimizing the distance from the actual shocks.

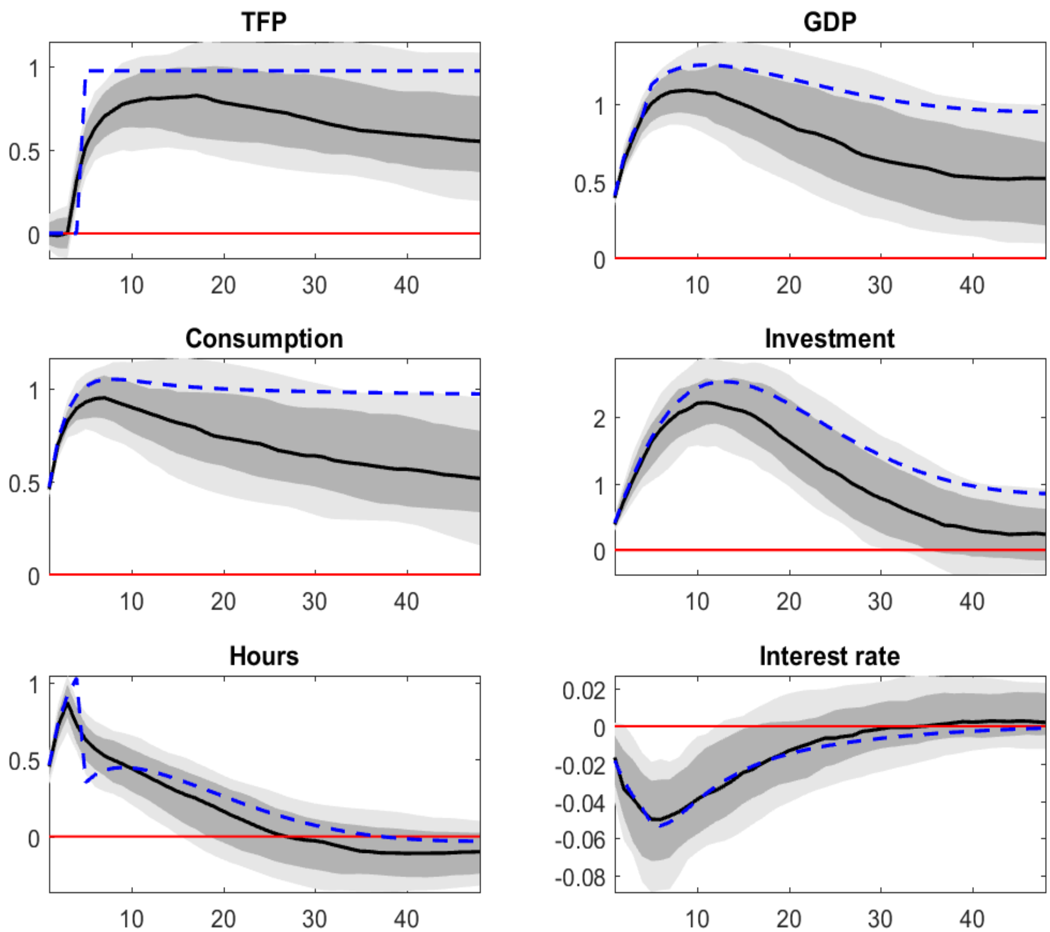

Figure 4: Impulse responses - S2 VAR. Minimizing the distance from the actual shocks.

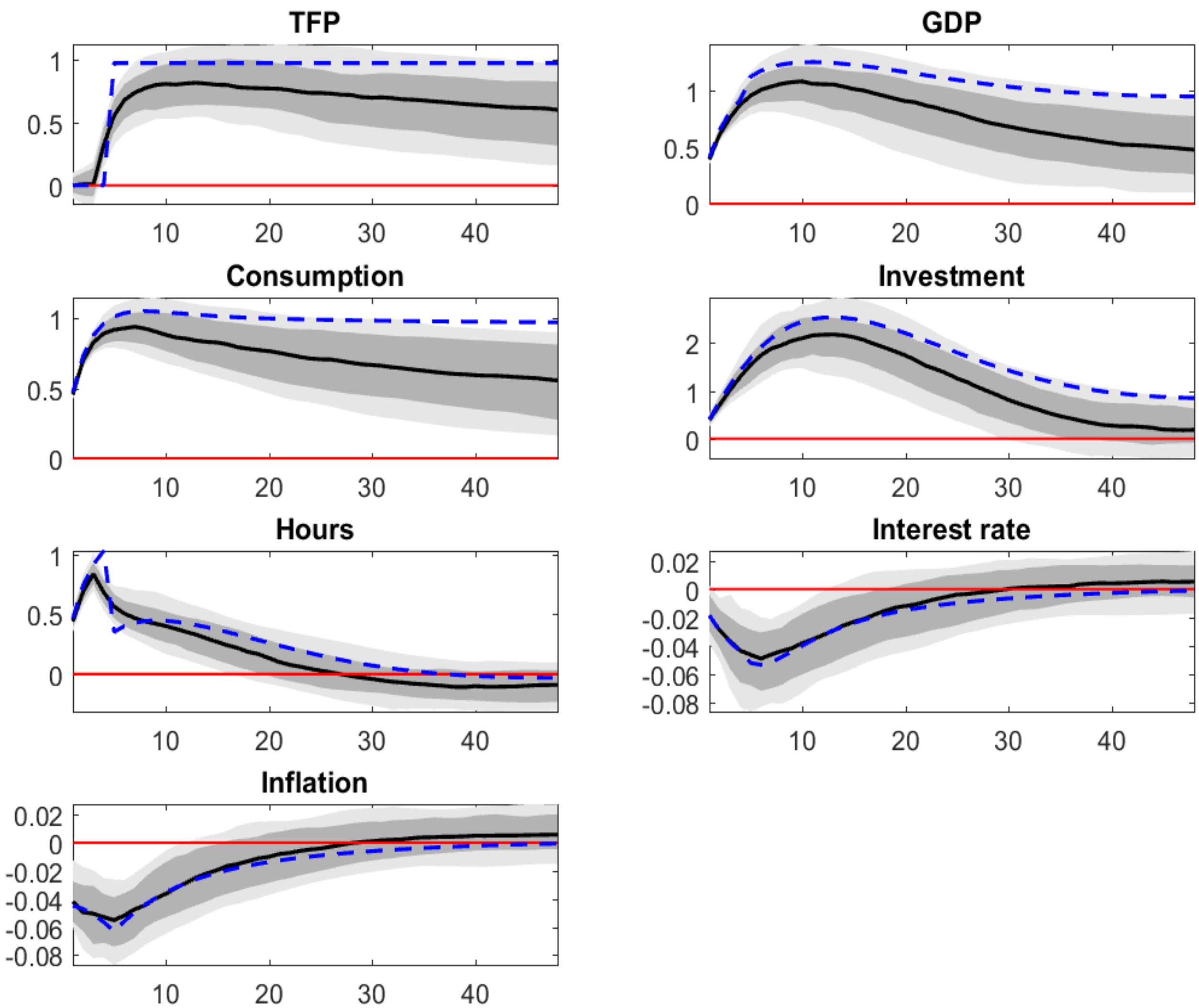

Figure 5: Impulse responses - S3 VAR. Minimizing the distance from the actual shocks.

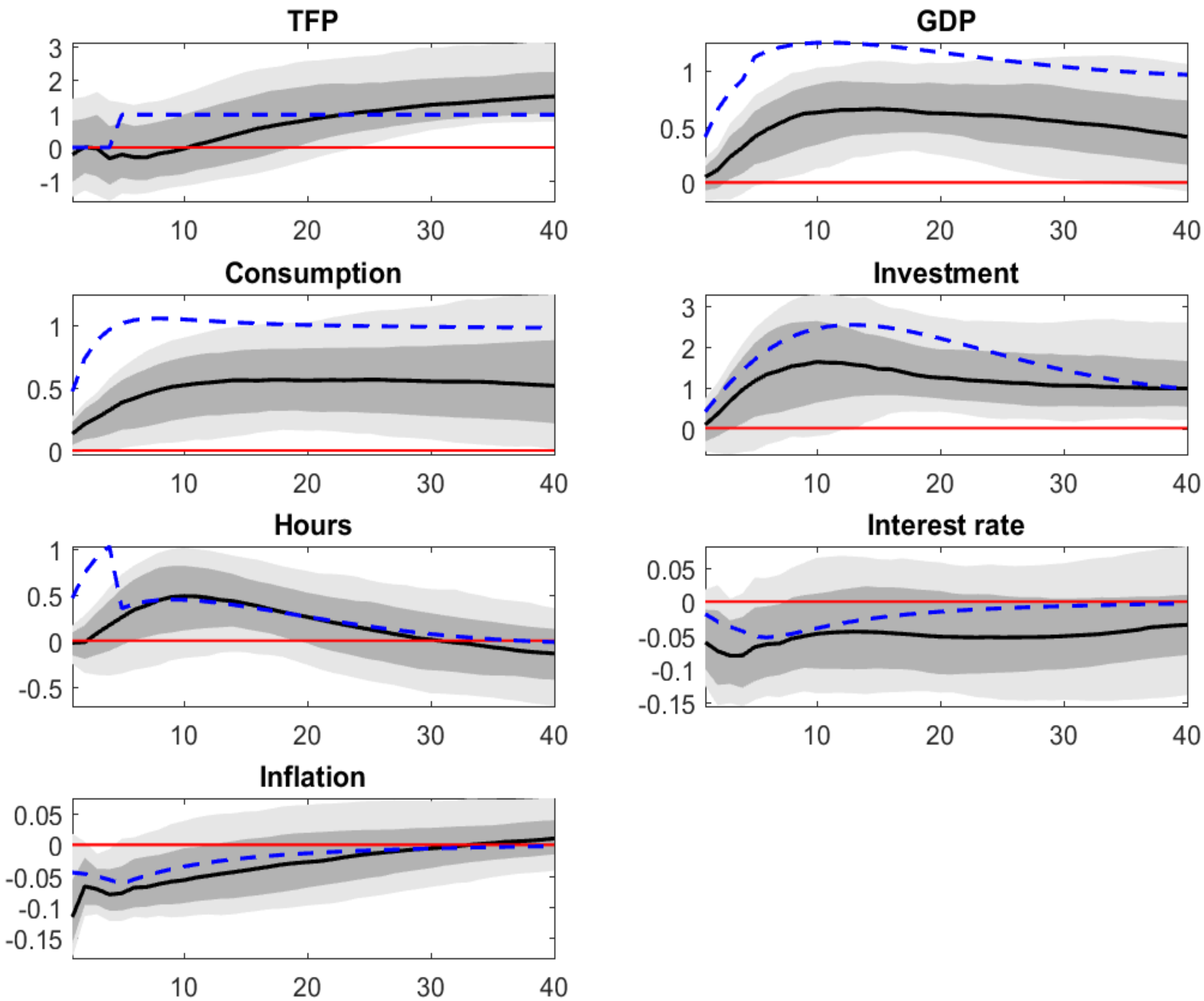

Figure 6: Impulse responses - S3 VAR on actual data. The news shock is the only shock with a long run effect of TFP.

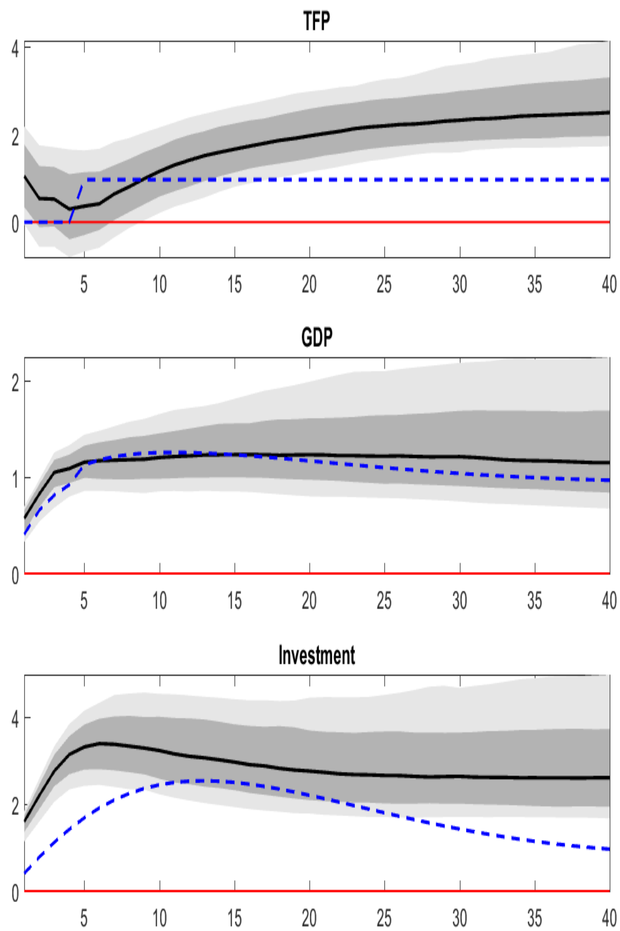

Figure 7: Impulse responses - S1 VAR on actual data. The news shock is the only shock with a long run effect of TFP.

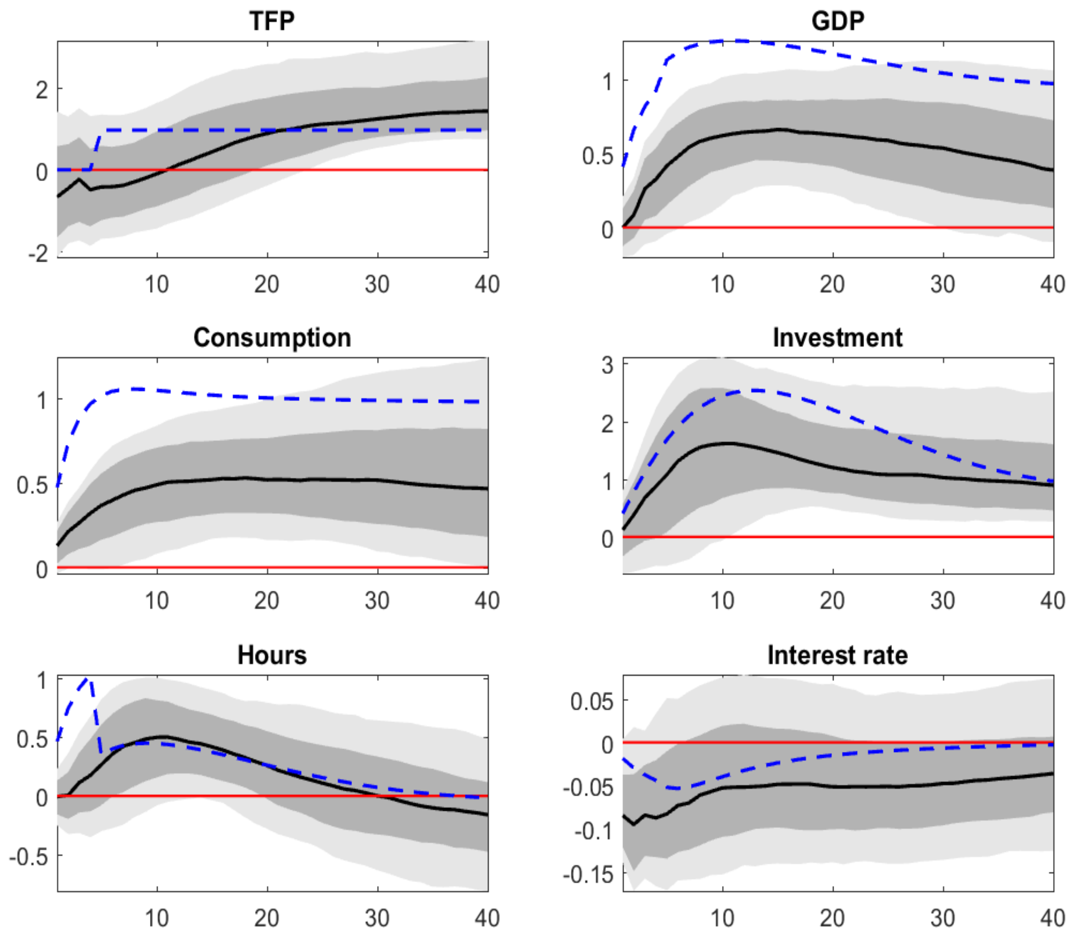

Figure 8: Impulse responses - S2 VAR on actual data. The news shock is the only shock with a long run effect of TFP.